UNIVERSIDADE ESTADUAL DE CAMPINAS
Instituto de Geociências

VINICIUS MURARO DA SILVA

FUTURES STUDIES AND FORESIGHT FOR SCIENCE, TECHNOLOGY AND INNOVATION: TRENDS OF USING BIG DATA AND MACHINE LEARNING

ESTUDOS DE FUTURO E FORESIGHT PARA CIÊNCIA, TECNOLOGIA E INOVAÇÃO: TENDÊNCIAS NO USO DE BIG DATA E MACHINE LEARNING

CAMPINAS
2021

VINICIUS MURARO DA SILVA

FUTURES STUDIES AND FORESIGHT FOR SCIENCE, TECHNOLOGY AND INNOVATION: TRENDS OF USING BIG DATA AND MACHINE LEARNING

ESTUDOS DE FUTURO E FORESIGHT PARA CIÊNCIA, TECNOLOGIA E INOVAÇÃO: TENDÊNCIAS NO USO DE BIG DATA E MACHINE LEARNING

THESIS PRESENTED TO THE INSTITUTE OF GEOSCIENCES OF THE UNIVERSITY OF CAMPINAS TO OBTAIN THE DEGREE OF DOCTOR OF SCIENTIFIC AND TECHNOLOGICAL POLICIES

TESE APRESENTADA AO INSTITUTO DE GEOCIÊNCIAS DA UNIVERSIDADE ESTADUAL DE CAMPINAS COMO REQUISITO PARA OBTENÇÃO DO TÍTULO DE DOUTOR EM POLÍTICA CIENTÍFICA E TECNOLÓGICA.

ORIENTADOR: PROF. DR. SÉRGIO LUIZ MONTEIRO SALLES FILHO

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO ALUNO VINICIUS MURARO DA SILVA E ORIENTADA PELO PROF. DR. SÉRGIO LUIZ MONTEIRO SALLES FILHO.

CAMPINAS

2021

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Geociências
Marta dos Santos - CRB 8/5892

Si38f

Silva, Vinícius Muraro da, 1989-
Futures studies and foresight for science, technology and innovation : trends of using big data and machine learning / Vinícius Muraro da Silva. – Campinas, SP : [s.n.], 2021.

Orientador: Sérgio Luiz Monteiro Salles Filho.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Geociências.

1. Previsão. 2. Incerteza. 3. Prospectiva. 4. Big data. 5. Aprendizado de Máquina. I. Salles Filho, Sérgio Luiz Monteiro, 1959-. II. Universidade Estadual de Campinas. Instituto de Geociências. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Estudos de futuro e foresight para ciência, tecnologia e inovação : tendências do uso de big data e machine learning
Palavras-chave em inglês:
Forecasting
Uncertainty
Foresight
Big data
Machine learning
Área de concentração: Política Científica e Tecnológica
Titulação: Doutor em Política Científica e Tecnológica
Banca examinadora:
Sérgio Luiz Monteiro Salles Filho [Orientador]
Ian Douglas Miles
Bernardo Pereira Cabral
Nicholas Spyridon Vonortas
Sérgio Robles Reis de Queiroz
Data de defesa: 15-03-2021
Programa de Pós-Graduação: Política Científica e Tecnológica

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0001-8531-0793
- Currículo Lattes do autor: http://lattes.cnpq.br/4564890992428149

UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE GEOCIÊNCIAS

AUTOR: VINICIUS MURARO DA SILVA

FUTURES STUDIES AND FORESIGHT FOR SCIENCE, TECHNOLOGY AND INNOVATION: TRENDS OF USING BIG DATA AND MACHINE LEARNING

ESTUDOS DE FUTURO E FORESIGHT PARA CIÊNCIA, TECNOLOGIA E INOVAÇÃO: TENDÊNCIAS NO USO DE BIG DATA E MACHINE LEARNING

ORIENTADOR: Prof. Dr. Sérgio Luiz Monteiro Salles Filho

Aprovado em: 15 / 03 / 2021

EXAMINADORES:

Prof. Dr. Sérgio Luiz Monteiro Salles Filho - Presidente

Prof. Dr. Nicholas Spyridon Vonortas

Prof. Dr. Bernardo Pereira Cabral

Prof. Dr. Ian Douglas Miles

Prof. Dr. Sérgio Robles Reis de Queiroz

A Ata de Defesa assinada pelos membros da Comissão Examinadora consta no processo de vida acadêmica do aluno.

Campinas, 15 de março de 2021.

ACKNOWLEDGEMENTS

Some people and institutions' support were essential to develop this project, which I will briefly mention.

I would like to thank the supervisor of this thesis, Prof. Dr. Sérgio Salles Filho, for all the support and discussion in developing this work, giving me learning opportunities over the past five years, and introducing myself to the academic world.

To the evaluation committee: Prof. Dr. Ian Miles, Prof. Dr. Nicholas Vonortas, Prof. Dr. Bernardo Cabral, Prof. Dr. Sérgio Queiroz, Prof. Dr. José Maria da Silveira, Prof. Dr. Roberto Marcondes Jr, Dr. Fábio Batista Mota and Dr. Jackson Maia for accepting my invitation to evaluate this work.

To InSySPo, a São Paulo Excellence Chair (SPEC) program funded by FAPESP and led by Prof. Dr. Nicholas Vonortas, for supporting my research and giving me several international experiences.

To GEOPI, a research group led by Prof. Dr. Sérgio Salles-Filho, Prof. Dr. Ana Maria Carneiro, and Prof. Dr. Adriana Bin, for giving me practical experiences in evaluating science, technology, and innovation, through projects and courses. Also, I would like to mention my friends from GEOPI, Sonia Tilkian, Luciana Lenhari, Camila Zeitoum and Luciane Graziele Ferrero.

To my colleagues and friends from Science and Technology Policy Department (DPCT): Rafaela Andrade, Victo Silva, Anna Navarro, Ana Carolina Spatti, Tatiana Bermudez, Cristina Monaco, Beatriz Ribeiro, and Paola Schaeffer for supporting each other in the academic pathway and collaborating in publications.

To my husband, family, and friends: Mauricio Cintra Lima, Angela Maria Muraro, Eli Ferreira da Silva, João Vitor Muraro, Raphaela Gândara, Prof. Dr. Leonice Domingos, Prof. Dr. Ronaldo Cintra Lima, Felipe Monteiro, Jean Mozart, Francine Domingos, Stephanie Rodrigues, and Jair Mendes for supporting and encouraging me in several moments in the past years, for listening and discussing my (sometimes crazy) ideas and for being part of my life.

This study was financed in part by the São Paulo Research Foundation (FAPESP), grant #2018/05144-4. This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) - Finance Code 001.

ABSTRACT

Futures studies have shown an accelerated growth since the post-World War II period, in which governments and private companies have been attentive to the importance of "forecasting" new trends, mainly technological ones, for their security as an institution. Such studies have gained a new panorama from the proliferation of data on massive scales and the increasing processing capacity, leading to new approaches, mainly in data-driven studies. Big Data and Machine Learning (BDML) has become powerful tools to extract and analyze data for future-oriented activities. The central question about using BDML tools is to understand the specific impacts of these mechanisms on futures studies' conceptual and methodological approaches. This work intends to respond to these questions by analyzing academic publications about futures studies supported by BDML and the opinions of 479 futures studies experts. The proposed methodology aims to comprehend how these tools are employed, the future benefits and limitations of BDML in foresight. The bibliometric results point to a reduced but increasing number of prospective studies supported by BDML published in the past decades. In general, these studies employ BDML techniques such as text and data mining in at least one part of the foresight process. Futures studies experts' opinions suggested that 1) analytical competencies are essential to deal with the complexity of the digital revolution, and 2) robust data analysis and automated tools support the transfer of study results to policy- and strategy-making. However, 3) the lack of data reliability and manipulation can play an uncertain role in this environment. The thesis concludes that BDML impact future-oriented activities in three dimensions: 1) Data reliance, 2) Data-Method integration, and 3) Decision-making. Data manipulation may increase the perception of substantive uncertainty in futures studies. However, integrating BDML techniques in foresight methodologies strongly decreases procedural uncertainty and will support effective decision-making. The limitation of this work is mainly two. First, non-academic futures studies publications were not collected in the bibliometric analysis. Second, the expert's population and sample characteristics were not compared due to a limitation of population data in survey analysis.

Keywords: Uncertainty; Futures studies; Foresight; Big Data; Machine Learning.

RESUMO

Estudos de futuro têm mostrado um crescimento acelerado desde o período pós-Segunda Guerra Mundial, em que governos e empresas privadas se atentaram à importância de "prever" novas tendências, principalmente tecnológicas, para sua segurança institucional. Tais estudos ganharam um novo panorama a partir da proliferação de dados em escalas massivas e da capacidade de processamento crescente, levando a novas abordagens principalmente em estudos baseados em dados. Big Data e Machine Learning (BDML) se tornaram ferramentas poderosas para extrair e analisar dados para atividades prospectivas. A questão central sobre o uso de ferramentas de BDML é entender os impactos específicos desses mecanismos nas abordagens conceituais e metodológicas de estudos futuros. Este trabalho pretende responder a essas questões através da análise de publicações acadêmicas sobre estudos de futuros apoiados por BDML e a aplicação de uma survey com 479 especialistas em foresight. A metodologia proposta visa compreender como essas ferramentas são empregadas e os futuros benefícios e limitações de BDML em foresight. Os resultados bibliométricos apontam para um número reduzido, mas crescente, de estudos futuros apoiados por BDML publicados nas últimas décadas. Em geral, esses estudos empregam técnicas de BDML, como mineração de texto e dados, em ao menos uma parte do processo de previsão. As opiniões dos especialistas em estudos futuros sugerem que 1) as competências analíticas são essenciais para lidar com a complexidade trazida pela revolução digital e 2) a robusta análise de dados e ferramentas automatizadas apoiam a transferência dos resultados dos estudos para o desenvolvimento de políticas e estratégias. No entanto, 3) a falta de confiabilidade e possível manipulação dos dados pode desempenhar um papel incerto neste ambiente. O trabalho conclui que BDML impacta as atividades orientadas para o futuro em três dimensões: 1) Confiança de dados, 2) Integração de dados e métodos e, 3) Tomada de decisão. A manipulação de dados pode aumentar a percepção de incerteza substantiva em estudos futuros. No entanto, a integração de técnicas de BDML em metodologias de previsão diminui fortemente a incerteza processual e apoiará a tomada de decisão eficaz. As limitações deste trabalho são principalmente duas. Primeiro, publicações de estudos de futuro não acadêmicos não foram coletadas e analisadas. Segundo, as características da amostra e da população de especialistas consultados não foram comparadas devido a uma limitação dos dados da população na análise da survey.

Palavras-Chave: Incerteza; Estudos de futuro; Foresight; Big Data; Machine Learning.

LIST OF FIGURES

Figure 1 - Related uncertainty and structure of problems over time 20
Figure 2 - Taxonomy of uncertainties 21
Figure 3 - Epistemic and aleatory components of objective uncertainty 22
Figure 4 - Scope of uncertainty: past and future 23
Figure 5 - Level of Potential Surprise 24
Figure 6 - Four levels of residual uncertainty 25
Figure 7 - Risk and uncertainty 26
Figure 8 - Stacey Matrix 29
Figure 9 - Taxonomy of uncertainty and decision 31
Figure 10 - Generic foresight process 35
Figure 11 - Foresight phases 36
Figure 12 - Continuous foresight process 37
Figure 13 - Foresight Diamond 45
Figure 14 - Industrial revolutions over time 53
Figure 15 - Big Data Analytics process 57
Figure 16 - Traditional programming vs. ML programming. 62
Figure 17 - Generic linear regression 64
Figure 18 - Generic decision tree graph 65
Figure 19 - Generic support vector machine application 66
Figure 20 - Generic neural network 68
Figure 21 - Data manipulation in data pipeline 74
Figure 22 - Roadmapping supported by text mining 83
Figure 23 - Scenario supported by text mining 83
Figure 24 - Process of search and use of Vantage Point Software. 89
Figure 25 - Model of data collection results for FSTI, BDML and FSTI+BDML (1968-2019) 93
Figure 26 - Publications by year in FSTI, BDML and FSTI+BDML 98
Figure 27 - Publication per country in FSTI+BDML papers 99
Figure 28 - Countries' co-occurrence network in FSTI+BDML papers 101
Figure 29 - Publications per journal in FSTI+BDML papers 102
Figure 30 - Top affiliations by classification in FSTI+BDML papers 104
Figure 31 - Affiliation's co-occurrence network in FSTI+BDML papers 105
Figure 32 - Top funding organization by classification in FSTI+BDML papers 106
Figure 33 - Funding organizations' co-occurrence network in FSTI+BDML papers 107
Figure 34 - Bibliographic coupling network in FSTI+BDML papers 109
Figure 35 - Publications per methodological approach in FSTI+BDML papers 113
Figure 36 - Publications per BDML technique in FSTI+BDML papers 113
Figure 37 - Countries per type of methodological approach in FSTI+BDML papers 114
Figure 38 - Countries per BDML techniques in FSTI+BDML papers 115

Figure 39 - Type of affiliation per type of methodological approach in FSTI+BDML papers 116
Figure 40 - BDML techniques per type of affiliation in FSTI+BDML papers 117
Figure 41 - Methodological approach's co-occurrence network in FSTI+BDML papers 118
Figure 42 - Occurrence of author's keywords per year in FSTI+BDML papers 119
Figure 43 - Wordcloud of author's keywords in FSTI+BDML papers 121
Figure 44 – Generic foresight process 126
Figure 45 - Experts demographic profile 134
Figure 46 - Experience and implications of BDML tools 135
Figure 47 - Distribution of answers in EP, DE and IF dimensions 137
Figure 48 - Scatterplot of projections according to probability, impact and desirability for 2025 138
Figure 49 – Group of projections based on EP and IF 150
Figure 50 - BDML effects on uncertainty in foresight activities 162

LIST OF TABLES

Table 1 - Types of uncertainty 32
Table 2 - Qualitative methods 41
Table 3 - Quantitative methods 43
Table 4 - Semi-quantitative methods 43
Table 5 - Foresight methodologies and frequency of use 46
Table 6 - Methodological steps for bibliometric analysis 90
Table 7 - Keywords for Foresight STI and BDML 91
Table 8 - Boolean queries and document results 92
Table 9 - Unique documents on each thematic data corpus 93
Table 10 - Methodologies classification 96
Table 11 - Productive and influential countries in FSTI+BDML papers 100
Table 12 - Productive and influential journals in FST+BDML papers 103
Table 13 - Productive and influential authors in FSTI+BDML papers 108
Table 14 - Author's degree, betweeness centrality and cluster in FSTI+BDML papers 110
Table 15 - Authors' clusters keywords, methods, countries and affiliations in FSTI+BDML papers 112
Table 16 - Countries methodological approaches in FSTI+BDML papers 114
Table 17 - Methodological approach per type of affiliation in FSTI+BDML papers 116
Table 18 - Rank of author's keywords per year in FSTI+BDML papers 120
Table 19 - Methodological steps for survey analysis 125
Table 20 - Description of 12 projections about influence of Big Data and Machine Learning 126
Table 21 - Median and mean values of survey projections 136
Table 22 - Distribution of answers for P1 – Support objective definition 139
Table 23 - Distribution of answers for P2 – Guide methodological choice 140
Table 24 - Distribution of answers for P3 – Require analytical skills 140
Table 25 – Distribution of answers for P4 – Support data collection 141
Table 26 - Distribution of answers for P5 – Improve data quality 141
Table 27 - Distribution of answers for P6 – Enhance data analysis 142
Table 28 - Distribution of answers for P7 – Drive qualitative methods 143
Table 29 - Distribution of answers for P8 – Increase data manipulation 143
Table 30 - Distribution of answers for P9 – Support results transfer 144
Table 31 - Distribution of answers for P10 – Increase decision accuracy 144
Table 32 - Distribution of answers for P11 – Support study evaluation 145
Table 33 - Distribution of answers for P12 – Support objectives' reach 146
Table 34 - Selected comments per foresight step 146
Table 36 - Uncertainty impacts in BDML-based foresight 165

SUMMARY

INTRODUCTION 14
Thesis structure 16
PART 1 - FUNDAMENTAL CONCEPTS 18
Chapter 1 - The Role of Uncertainty in Futures Studies 18
1.1 Uncertainty and related concepts 18
1.1.1 Dimensions of Uncertainty 19
1.1.1 Risk 26
1.1.2 Ambiguity 27
1.1.3 Intuition 27
1.1.4 Ergodicity and Complex Systems 28
1.1.5 Bounded Rationality 30
1.1.6 Types of Decision 30
1.2 Futures studies 33
1.2.1 Historical introduction to futures studies 33
1.2.2 Futures studies types and process 35
1.2.3 Futures studies methodologies 41
1.2.4 New approaches and limitations to futures studies 46
1.2.5 Competitive gain on the use of foresight 48
1.3 Conclusion 50
Chapter 2 - Evolution and Application of Big Data and Machine Learning 52
2.1 The fourth industrial revolution 52
2.1.1 Technologies 4.0 53
2.2 What is Big Data 55
2.2.1 The concept of Big Data 56
2.2.2 Big Data Analytics 57
2.3 What is Machine Learning 60
2.3.1 Natural and Machine Learning 60
2.3.2 The concept of Machine Learning 61
2.3.3 Supervised and unsupervised learning 63
2.3.4 Machine Learning Techniques 64
2.4 Applications and Limitations of Big Data and Machine Learning 68

2.4.1 Artificial Intelligence 68
2.4.2 Fake news and Deep fakes 70
2.4.3 Limitations of the technologies 73
2.5 Big Data and Machine Learning tools for futures studies 77
2.5.1 Statistics and futures studies 77
2.5.2 Text mining and Big Data Analytics 78
2.5.3 Machine Learning in futures studies 79
2.5.4 Deep Learning in Bibliometrics 80
2.5.5 Cluster Analysis 80
2.5.6 Social Network Analysis (SNA) 81
2.5.7 Integration with foresight methodologies 82
2.6 Conclusion 84
PART 2 - THE FRONTIERS OF BIG DATA AND MACHINE LEARNING IN FUTURES STUDIES 86
Chapter 3 - Panorama of Futures Studies Supported by Big Data and Machine Learning 86
3.1 Bibliometrics and social network analysis 86
3.2 Methodological steps: bibliometric analysis 90
3.2.1 Keywords definition for bibliometrics 90
3.2.2 Boolean query string definition for bibliometrics 91
3.2.3 Data fusion 93
3.2.4 Data cleaning 94
3.2.5 Organizations classification 95
3.2.6 Methodologies and BDML techniques classification 95
3.2.7 Data visualization 96
3.3 Results and discussion 96
3.3.1 Overview 97
3.3.2 Country analysis 98
3.3.3 Journal analysis 101
3.3.4 Affiliation analysis 103
3.3.5 Funding organization analysis 105
3.3.6 Author analysis 108
3.3.7 Methods analysis 113
3.3.8 Keyword analysis 118

3.4 Conclusions 121
Chapter 4 - The Future of Futures Studies 124
4.1 Methodological approach: survey analysis 124
4.1.1 Development of projections 125
4.1.2 Online survey creation 132
4.1.3 Defining target respondents: foresight experts 133
4.1.4 Send, follow and monitor survey answering 133
4.1.5 Analyze results and data visualization 134
4.2 Survey results 134
4.2.1 Overview 135
4.2.2 Pre-foresight projections 139
4.2.3 Recruitment projections 140
4.2.4 Generation projections 142
4.2.5 Action projections 144
4.2.6 Renewal projections 145
4.2.7 Qualitative comments on the survey 146
4.3 Discussion: future of futures studies 149
4.3.1 Foresight Team, Results Transfer and Data Uncertainty 150
4.3.2 Foresight Methods and Data Analytics 152
4.3.3 Decision Making and Foresight Evaluation 153
4.3.4 Futures Studies Objective 154
4.4 Conclusion 155
FINAL REMARKS 158
Uncertainty and BDML-based futures studies 161
Limitations of the study 165
BIBLIOGRAPHY 167
ANEXXES 181
ANNEX 1 - Free and Informed Consent Term 181
ANNEX 2 - Survey Questionnaire 183

INTRODUCTION

Since the post-World War II period, governments and companies have been attentive to the importance of "forecasting" technological trends for their security as an institution (MILES ET AL., 2008). Over the decades, futures studies' evolution creates a powerful tool to support more appropriate science and technology policy-making and corporate strategy-making, adapted to complex and uncertain environments at the regional or national level.

In recent years the term foresight has been widely used as a description for a series of studies that support decision-making at both governmental and corporate levels. It generates a shared vision of the future among agents and resources in their technological development efforts (MILES ET AL., 2008, 2016)

Among the many objectives that a futures study aims, Miles et al. (2008) define five general rationales for foresight analysis:

1) Direct or prioritize investments in Science, Technology, and Innovation;

2) Create new contacts and links between the actors around a shared direction;

3) Expand the breadth of knowledge and visions for the future;

4) Bring new players to the strategic debate;

5) Improve policy development and strategy formation in areas where science and innovation play a significant role.

The concept of uncertainty is an essential element in futures studies and has been studied for a long time (TANNERT ET AL., 2007). Uncertainty is a situation closely related to imperfect information or knowledge. It manifests itself in several research fields as psychology, physics, philosophy, economics, sociology, and engineering. Dosi & Egidi (1991) categorize uncertainty based on its sources: 1) substantive uncertainty is related to the lack of information necessary to make a decision, and 2) procedural uncertainty is associated with the limitation on cognitive and computational capabilities of agents. In other words, the availability of data, the cognitive capacity, and available time are limitations to rational thinking in decision-making, such as Simon (2000) describes as Bounded Rationality.

How to reach a foresight method that minimizes uncertainty effects in future-oriented activities is one of the main challenges faced by futures studies, due to the future's uncertain and complex environment. However, futures studies may successfully meet their objectives based on substantial analyzed data, correct methodologies mix, and interaction among the actors.

Futures studies have gained a new panorama with data proliferation on massive scales and increased processing capacity. Big Data refers to a massive data set in terms of acquiring, managing, and processing information. Loukides (2010) defines Big Data as a set of data whose size becomes a problem, and the usual collection, storage, management, and analysis tools do not fit correctly. Big Data is explained by 3 V's - Volume, Variety, and Velocity (LOUKIDES, 2010; MCGUIRE ET AL., 2012; MGI, 2011; OECD, 2013, 2014). Machine Learning is briefly explained as the field of study that algorithms may improve their tasks' efficiency according to their experience (SAMUEL, 1959). The learning mechanism is categorized as supervised and unsupervised. Supervised learning is a machine-learning algorithm that learns the "correct answer" mapping inputs and outputs through examples inserted in the system. On the other hand, unsupervised algorithms have no labels and look for structure on sample data, grouping them on clusters of similar rules, discovering hidden patterns. The use of information technology tools in foresight studies can increasingly support the practitioner through improved communication and data consistency. Thereby, it turns the decision-making process more assertively (GRACHT ET AL., 2015).

However, Big Data and Machine Learning limitations are translated into four main points: low data quality, data reliability, data interpretability, and ethical issues (REIMSBACH-KOUNATZE, 2015). First, data quality can be measured by its relevance, accuracy, and credibility (OECD, 2011). Second, fake news, data manipulation, data biases, and lack of data security are responsible for reducing Big Data reliance. Third, decisions have to be made pragmatically, it means not only by looking at data analysis without a full understanding of the specific context. Inaccurate understanding of causal relations among variables may induce misplaced decisions. Fourth, data-collection has to deal with several ethical and privacy issues.

The central question about using Big Data and Machine Learning tools in foresight studies is to understand the impact of these mechanisms on bringing a new perspective of uncertainty for practitioners and decision-makers. The subjective uncertainty in its two dimensions (substantive and procedural) is affected when capacities obtained by Big Data and Machine Learning are present. Big Data may virtually affect uncertainty's substantive dimension. On the other hand, Machine

Learning deals with a powerful processing capacity, affecting the procedural dimension. Big Data and Machine Learning will not eliminate substantive and procedural uncertainty at all. Still, it is expected that they may reduce the influence of these dimensions of uncertainty on futures studies' outcomes.

Some questions remain when discussing the role of Big Data and Machine Learning for future-oriented activities. There are few studies about accuracy improvement in using Big Data and Machine Learning in foresight outcomes or even data-based evidence of better results in complex environments. Given the evidence that Big Data and Machine Learning have obvious effects on the availability and processing of large amounts of data, to what extent these effects will affect the capabilities, the methodological approaches, and the tools employed for futures studies? Which roles Big Data and Machine Learning may play in the permanent endeavor of dealing with uncertainty?

Thesis structure

To answer and explore the question above, the overall objective of this thesis is to identify and analyze trends of futures studies in science, technology, and innovation (STI) from the perspective of Big Data and Machine Learning, to understand how they are changing conceptual and methodological approaches for futures studies. For specific objectives:

1. Analyze the status and the potential of Big Data and Machine Learning tools applied to prospective studies in STI.
2. Discuss how Big Data and Machine Learning may affect the present understanding of uncertainty in futures studies.
3. Identify impacts of Big Data and Machine Learning on the methodological approaches for futures studies STI.

This thesis is divided into two parts and the final remarks.

<u>Part I – Fundamental concepts:</u> The first two chapters of this work are dedicated to depicting the conceptual elements necessary to fully understand Big Data and Machine Learning's role on uncertainty and futures studies. The methodology to achieve the objectives of part 1 is a literature review.

<u>Chapter 1 - The role of uncertainty in futures studies</u>**:** Some elementary concepts are indispensable to understand the dynamics of futures studies thoroughly. The notion of uncertainty and the interpretation of its different dimensions are vital to

comprehend Big Data and Machine Learning's impact on futures studies. Other elements like risk, ambiguity, intuition, complexity, and theory decision will also be explored to cover futures studies' theoretical aspects. It seeks to discuss and analyze these concepts' evolution over time and explore the relationship between them.

Chapter 2 - Evolution and application of Big Data and Machine Learning**:** This chapter's main objective is to introduce the concepts of Big Data, Machine Learning and understand how these approaches are being used in different contexts to support prospective activities. Actual applications and limitations of these approaches will also be explored.

Part II – Frontiers of Big Data and Machine Learning in futures studies**:** The next two chapters of the thesis are dedicated to performing practical experiments and analyses of Big Data and Machine Learning tools to support foresight activities.

Chapter 3 - Panorama of futures studies supported by Big Data and Machine Learning**:** This chapter's main intention is to obtain an overview of the futures studies supported by Big Data and Machine Learning in the past decades. The purpose of this panorama is to explore the employment of different foresight methodologies interacting with Big Data, Machine Learning, and their tools. In this chapter, a bibliometric process analyzes scientific publications about futures studies supported by Big Data and Machine Learning.

Chapter 4 – The future of futures studies**:** The fourth chapter of this thesis is dedicated to assessing the future impact of Big Data and Machine Learning in foresight. A survey was performed with 479 foresight specialists to understand their perceptions about 12 near-future projections based on Machine Learning and Big Data's impacts in futures studies practice.

Final Remarks**:** Finally, the last session discusses the bibliometric analysis and survey results based on the conceptual framework of uncertainty presented in the first part of this thesis. It also explains the findings and limitations of this research.

PART 1 - FUNDAMENTAL CONCEPTS

The first two chapters of this work are dedicated to present the fundamental conceptual elements to understand Big Data and Machine Learning's role in uncertainty and futures studies.

Chapter 1 - The Role of Uncertainty in Futures Studies

Some elemental concepts are fundamental to understand the dynamics of futures studies thoroughly. The notion of uncertainty and the interpretation of its different dimensions are vital points to comprehend Big Data and Machine Learning's actual impact on futures studies. Chapter 1 will explore other elements like risk, ambiguity, intuition, complexity, and theory decision to cover prospective studies' theoretical aspects. Thus, this chapter seeks to discuss and analyze these concepts' evolution over time, exploring the relationship between them and their impacts on related fields. Through a literature review, section 1.1 examines approaches of uncertainty and other associated concepts that underlie futures studies. Sections 1.2 depict the aspects related to futures studies such as history, processes, methodologies, and future trends are presented.

1.1 Uncertainty and related concepts

Many authors study the principle of uncertainty since the beginning of the twentieth century, mainly discussing the future of economics, technology, society behavior. Decoupling the concept of risk and uncertainty opens the possibility of further studies on the future and its construction through human decisions. The idea of uncertainty has been studied for a long time, since Socrates and Plato. They doubt whether scientific knowledge can express reality because of our experience's limit, preventing us from predicting the future and making decisions with certainty (TANNERT ET AL., 2007).

At the beginning of the nineteenth century, Pierre Simon Laplace, based on scientific theories and successful mathematical calculations, concluded that the universe is entirely determined. Such an approach is based on the idea that if it is possible to know the perfect and exact state of a phenomenon at a given moment, it is also possible to predict the future state of this same phenomenon, which also includes predicting human behavior. Approximately 100 years later, the indeterminacy principle in quantum physics was fundamental to reject Laplace's hypothesis. The indeterminacy principle, based on probability, embraces the idea of the indeterminate world. It is impossible to determine precisely a phenomenon in the future since it is not possible to calculate, with sufficient precision, the current state of the universe (HAWKING, 1988). The phenomenon of unpredictability and randomness support philosophical and scientific discussions about uncertainty which the level of complexity defines a limit where the laws of classical physics - or deterministic relationships - no longer work (MAGRUK, 2017).

1.1.1 Dimensions of Uncertainty

Authors in many research fields use to define uncertainty in very distinct ways. In general characterization, uncertainty is a situation closely related to imperfect or unknown information and occurs at the limits of knowledge. In other words, uncertainty permeates several moments, not only in academic fields but also in daily life (WAKEHAM, 2015). It has a significant influence on our daily decisions, from choosing whether to take an umbrella on a cloudy day and invest or not in stocks on the stock exchange. Each decision carries its risks and benefits that can have consequences at various levels.

The progress of social, economic, and technological systems leads the future to an even more uncertain pathway due to the increase of the variables that runs the world. To create wealth and increase the quality of life in this constantly changing world is necessary to manage the changes and monitor uncertainty to make better decisions. The nature of uncertainty is time-dependent. The time horizon increases uncertainty at the proportional level as it increases the inability to make decisions based on known patterns. Events' structure decreases over time due to the increasing number of components and less information about their characteristics. Less-structured problems led to an increase in uncertainty (Figure 1) (SARITAS & ONER, 2004).

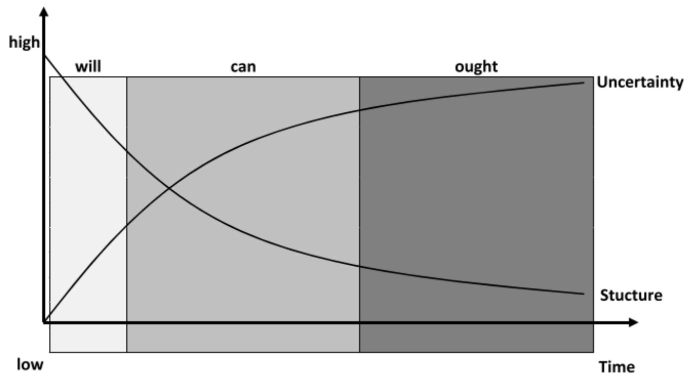


Figure 1 - Related uncertainty and structure of problems over time

Source: Saritas & Oner (2004)

Tannert et al. (2007) propose a taxonomy of uncertainty (Figure 2) with a subjective and an objective dimension. The author considers uncertainty a phenomenological, cognitive, and emotional experience. When new experiences or new evidence challenge personal beliefs, some may experience uncertainty in an individual aspect, in a subjective dimension. The lack of moral rules in taking a decision generates a situation of moral uncertainty. In this case, decision-makers need to look at more general moral rules and deduce guidance for the situation. The rule uncertainty situation is caused by uncertainty in the moral rules itself, relying on decision based on intuition.

On the other side, the objective dimension of uncertainty assumes the world path is based on knowledge, but it is only possible to learn it at some level. It is not necessarily a feeling (to distinguish from subjective dimension), but a feature of life and time. On this dimension, epistemological uncertainty is caused by gaps in knowledge that research can pursue, and it is required to avoid risks or dangers. Stochastic characteristics of a situation, which can involve complex systems, results in ontological uncertainty, which means the non-linear behavior associated with these situations makes it impossible to reduce these uncertainties.

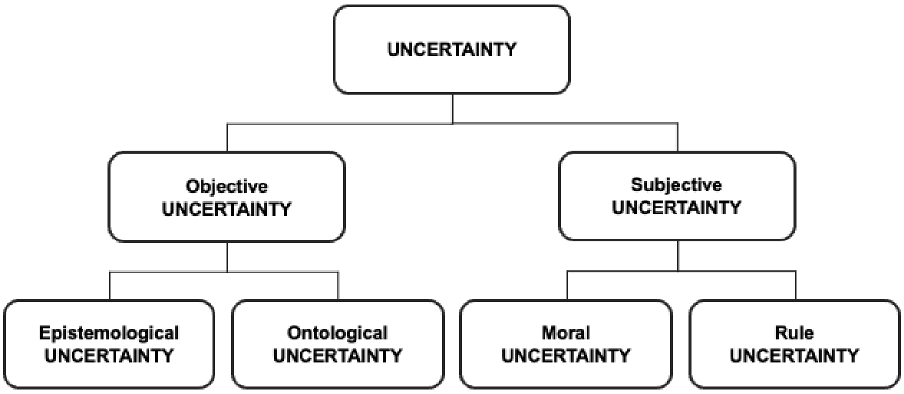


Figure 2 - Taxonomy of uncertainties

Source: based on Tannert et al. (2007)

In general, one can deal with both uncertainty dimensions simultaneously, experiencing subjective uncertainty and dealing with the reality of knowledge gaps as objective uncertainty. Researchers emphasize different dimensions of uncertainty depending on their field or research subject. They also employ other related concepts that are commonly followed by uncertainty discussions, such as risk, ambiguity, intuition, and complexity.

Figure 3 presents the relationship between epistemic and ontological (also called aleatory) in overall uncertainty as a function of the knowledge base. The lack of knowledge and the randomness causes variability in epistemic and aleatory components of uncertainty. As the knowledge base advance over time, through scientific research and testing, epistemic uncertainty reduces while the aleatory component still plays its role in overall uncertainty, even in a perfect knowledge base. (VARDE & PECHT, 2018)

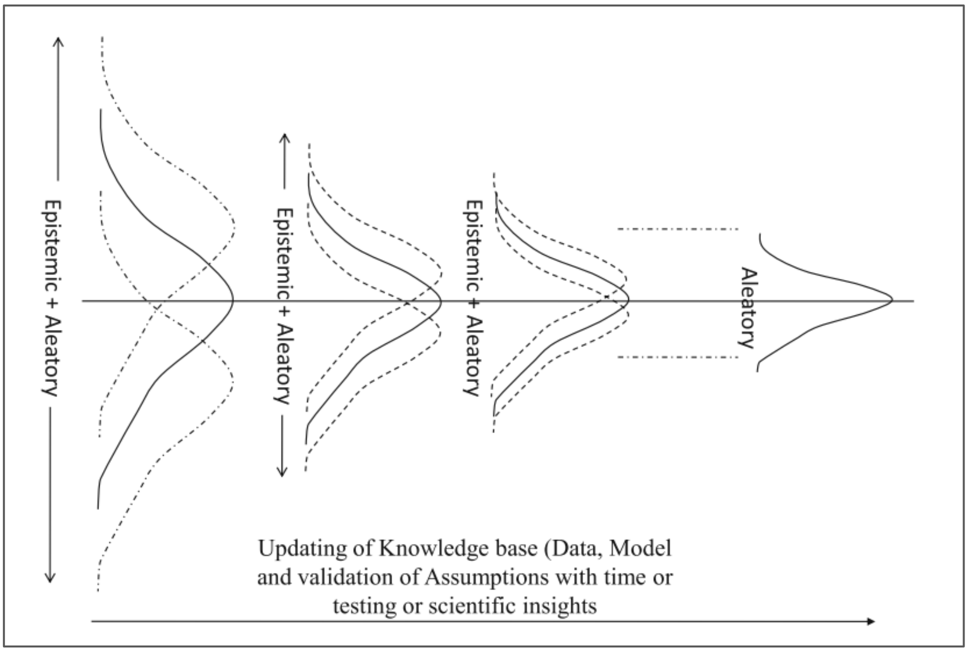


Figure 3 - Epistemic and aleatory components of objective uncertainty

Source: Varde & Pecht (2018)

Magruk (2017) proposes a new representation of uncertainty over time in the form of a light cone with the framework of the relativity theory (Figure 4). The absolute future event E is inside the cone of the future, which contains all events that can affect what happens in P. The cone's past-region is the collection of all events that impacted P, for which the information could get to E. If it's possible to obtain all the information in all points at the past-region cone on deterministic systems, it would be possible to predict what will occur in E before it happens. On indeterministic systems, the information of events on the past and the present can only distribute probabilities of possible states in the future. The space "Elsewhere" includes all events that cannot affect or be affected by P.

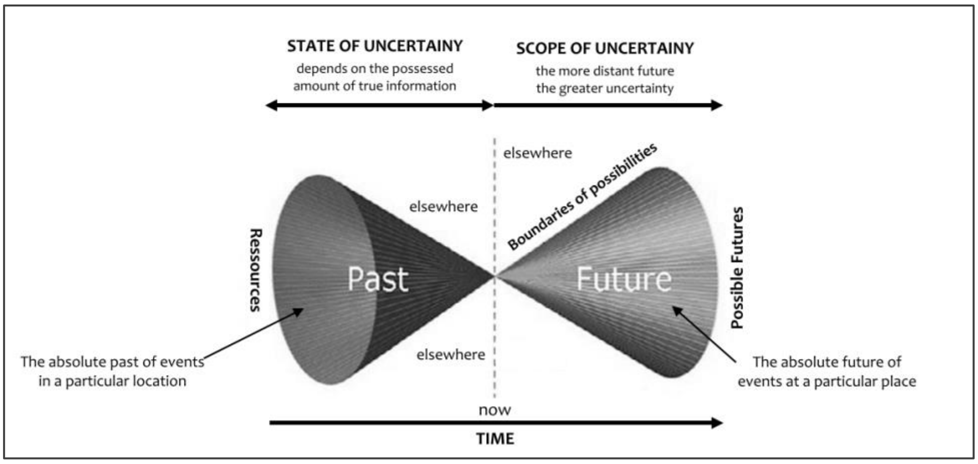


Figure 4 - Scope of uncertainty: past and future

Source: Magruk (2017)

Economics has been the social science that explores the most the concept of uncertainty. In order to model action and behavior, uncertainty imposes a theoretical obstacle on the economic research field. Thus, domesticate and tame uncertainty is the objective of the economics study (QUIGGIN, 2005).

Keynes (1921) and Knight (1921) are two economists who, in 1921, published books dealing with the problem of uncertainty. Keynes discusses the logical relationships between two events based on a subjective probabilistic assessment of uncertainty and accuracy. Some relations between two rival events can be numerically defined, in probabilities, defining risk. However, when numerical relations cannot be established, the authors consider it as fundamental uncertainty.

Shackle (1955), based on Knight and Keynes' uncertainty traditions, also establishes his studies on uncertainty as a theory of decision making. Shackle also sees uncertainty as different from risk. Uncertainty for Shackle is subjective and can be analyzed from two different views: as a distributional or non-distributional variable. As a distributional variable, uncertainty can be seen as a set of probabilities (i.e., risk in Knight's view), when one has a complete list of rival hypotheses, which means no residual hypothesis. As a non-distributional variable, uncertainty can be seen as levels of possibility and surprise. An event with the perfect possibility of happening will offer a null surprise to the observer waiting for that event. Otherwise, an event that has a perfect impossibility to occur offers a maximum surprise. This analogy makes clear the concept of potential surprise. In this way, it is allowed that more than one rival hypothesis has the same level of potential surprise, this being the measure of the possibility of an event.

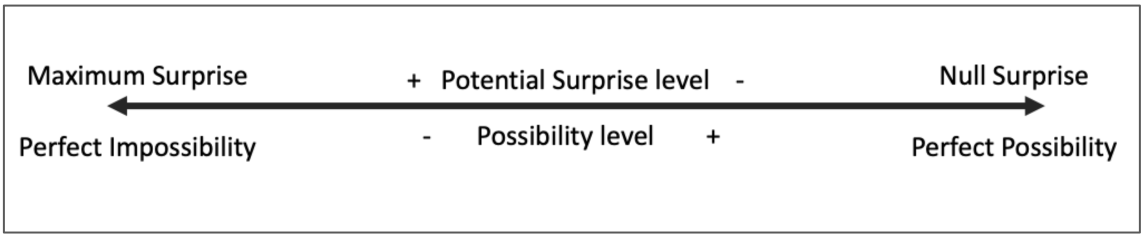


Figure 5 - Level of Potential Surprise

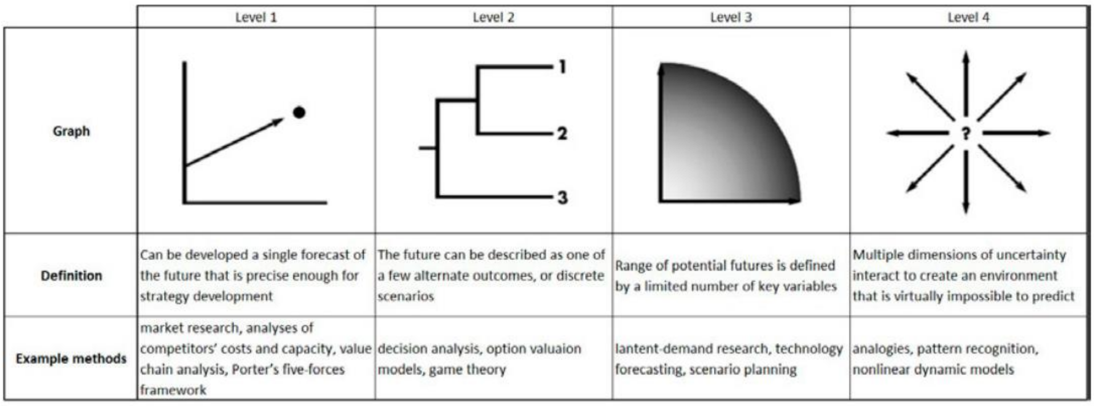
Source: Author's elaboration, based on Shackle (1969)

Some authors describe uncertainty based on typologies related to the context of decision making. Dosi & Egidi (1991) propose a typology based on two main sources of uncertainty: 1) substantive uncertainty, based on the lack of information necessary for decision-making with particular results, and 2) procedural uncertainty, based on the limitation in the cognitive and computational capabilities of agents in pursuing their goals, given the available information. Such concepts are aligned with Simon's concepts of substantive and procedural rationality (SIMON, 1959).

Dosi & Egidi (1991) further explore distinctions of substantive uncertainty. The weak substantive uncertainty refers to cases where uncertainty derives simply from the lack of information about the occurrence of a given event within a list of finite and known events. This concept is analogous to risk (KNIGHT, 1921) and explored later in this chapter. On the other hand, strong substantive uncertainty is due to the lack of information about the rival events themselves, which means the rival events are unknown (even unlimited), or it is impossible to define the distribution probability of the events.

Strong substantive uncertainty can also be understood in two ways: ambiguity and fundamental uncertainty. Ambiguity is classified by Dequech (2000) as the less-strong type of strong uncertainty. In ambiguous situations, uncertainty is related to the specification of the probabilistic distribution between events, and this uncertainty is due to a lack of information. When this probabilistic distribution among all possible events is unambiguous, the situation involves risk but no ambiguity. On the other hand, fundamental uncertainty is the situation of significant indeterminacy of the future. In a condition of fundamental uncertainty, it is impossible to know all the information about future events at the time of the decision, much less to define a probabilistic distribution. It is characterized by the possibility of creativity and not predetermined structural changes. Dequech (2000) highlights the ontological approach to fundamental uncertainty.

From a perspective of possible futures and strategic foresight, Courtney (2001) defines residual uncertainty as "the uncertainty left after the best possible analysis to separate the known from the unknowable" (COURTNEY, 2001, p.4). Residual uncertainty can assume one of four levels: 1) A Clear-Enough Future, 2) Alternate Futures, 3) A Range of Futures, and 4) True Ambiguity (Figure 6).



	Level 1	Level 2	Level 3	Level 4
Graph		1 2 3		?
Definition	Can be developed a single forecast of the future that is precise enough for strategy development	The future can be described as one of a few alternate outcomes, or discrete scenarios	Range of potential futures is defined by a limited number of key variables	Multiple dimensions of uncertainty interact to create an environment that is virtually impossible to predict
Example methods	market research, analyses of competitors' costs and capacity, value chain analysis, Porter's five-forces framework	decision analysis, option valuaion models, game theory	lantent-demand research, technology forecasting, scenario planning	analogies, pattern recognition, nonlinear dynamic models

Figure 6 - Four levels of residual uncertainty

Source: Courtney (2001)

In a situation of level 1 uncertainty, a clear and well-defined future can be predicted with the necessary accuracy for developing a strategy. Examples of methods used in this situation are market research, analysis of competitors' costs and capabilities, and value chain analysis.

In a situation of level 2 uncertainty, the future can be interpreted as a finite series of discrete and alternative scenarios. Strategies for level 2 uncertainty are decision analysis, option valuation models, and game theory.

Level 3 uncertainty is defined as a range of potential futures based on a limited number of key variables. Strategies for level 3 uncertainty are technology forecasting, scenario planning, and latent-demand research.

At level 4 uncertainty, multiple dimensions of uncertainty interact and create an impossible prediction environment. In this situation, analogies, pattern recognition, and dynamic models.

1.1.1 Risk

Uncertainty and risk are concepts frequently paired in several fields such as economics and social sciences. Several concepts of risk are discussed and theorized in different and specific sociocultural backgrounds. Garland (2003) considers the concepts of risk so heterogeneous that they have no connection to each other but the assumption of distinction between reality and possibility.

Knight (1921) was the first author to exalt the conceptual difference between risk and uncertainty in economics. In Knight's view, the risk is considered probabilistically measurable, while uncertainty is not. The risk can be measured from logical and statistical probability, which are empirical generalizations from data from previous events. The estimate is not strict logical reasoning but an intuitive judgment to measure the real uncertainty.

Magruk (2017) proposes a graphical model of risk and uncertainty (Figure 7). The risk is presented as a situation that the complete set of alternative futures (A) is presented with their respective probabilities (p) of occurrence. On the other hand, in an uncertain situation, possible futures are divided into known futures (A) and unknown futures (U). The probability of each of these futures occurring is also unknown.

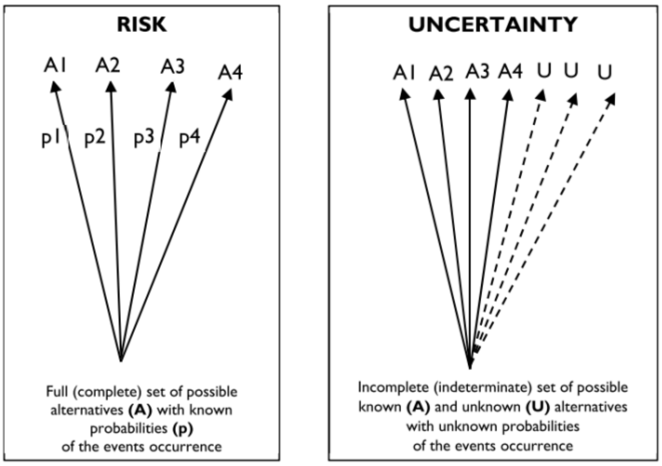


Figure 7 - Risk and uncertainty

Source: Magruk (2017)

In common sense, the risk is frequently related to potentially negative outcomes of an action or decision. Individuals tend to avoid risks or mitigate them. On the subjective dimension, an individual dealing in an uncertain environment will consider the presumable impacts of some decision: the benefits and the risks (potential adverse outcomes) of the action to be taken. This perception of risk may induce the feeling of uncertainty. In the objective sense, risk refers to the broader set of unknown deployments that may affect decisions and actions. Also, in this case, risk can be accounted for, like probabilities (ZINN, 2009).

1.1.2 Ambiguity

Ambiguity implies uncertainty about the "probabilities of possible outcomes rather than the outcome itself" (CAMERER & WEBER, 1992; DEQUECH, 2000).

Dequech (2000) highlights the difference between fundamental uncertainty and ambiguity to facilitate the discussion between economists of different schools of thought, or lines of research. Ambiguity, in his terms, refers to a situation that there is uncertainty about probabilities of occurring rival events due to the lack of information.

Ambiguity can be exemplified through Ellsberg's problem: there is an urn with 90 balls, in which it is known that there are 30 red balls and 60 black or yellow balls, and the proportion of black and yellow balls is unknown. One ball is to be drawn at random, and people offer two pairs of bets, twice. They have to choose between a bet on a red ball or a bet on a black ball in the first bet. Then, they have to choose between a bet on a “red or yellow” ball or on a “black or yellow” ball in the second. Often people bet on a red ball in the first attempt and on a “black or yellow” on the second attempt, contradicting the rationality proposed by subjective expected utility theory. People prefer to bet on a red on the first try and on a non-red on the second try, avoiding ambiguity, which means avoiding the lack of information about the probability of black or yellow balls. It's important to say that this information about the proportion of black and yellow balls exists, but it is hidden at the moment of the decision. If people have this complete information, ambiguity could be eliminated (ELLSBERG, 1961).

1.1.3 Intuition

Intuition embraces several definitions from different domains. Many authors agree that intuition is different from analytical reasoning. The general definition that emerges from the literature is that intuition is “knowing without being able to explain how we know” (VAUGHAN, 1979).

For Simon (1987) decisions made by intuition are non-rational decisions. Simon also describes conscious analytic decisions as rational decisions (or logical decisions) and emotion-based decisions as irrational decisions. These definitions are essential to categorize and understand the role of intuition and emotions in decision-making for several fields.

Intuition is based on experience, tacit knowledge, and pattern recognition. The importance of sensorial and emotional elements in the definition of intuition is significant. Vaughan (1979) separates intuition into four discrete levels of awareness: physical, emotional, mental, and spiritual. The physical level is associated with body perceptions and can be exemplified by how individuals physiologically respond to events consciousness. At the emotional level, intuition comes through feelings. It can include “sensitivity to other people’s feelings, a vague sense that one is supposed to do something and instances of immediate liking or disliking with no apparent reason” (SHIRLEY & LANGAN-FOX, 1996). Mental intuition is also called “inner vision” and that implies getting some accurate conclusions based on insufficient information. Finally, spiritual intuition is associated with mystical experience and is not related to sensations, feelings, or thoughts.

1.1.4 Ergodicity and Complex Systems

Davidson (2006) adds to the conceptual framework of uncertainty, the concept of ergodicity and non-ergodicity. In ergodic systems future events are a mere static consequence of past events, the basis of neoclassical economic theory. A future with a creative and uncertain reality is present in non-ergodic systems. Time is irreversible and the reality created from human decisions, which are subject to unpredictability. Thus, uncertainty (mainly ontological uncertainty) is defined in terms of the existence of non-ergodic processes, where the observation of past events does not reproduce future events, making the future not calculable.

Complexity theory is an emerging field in science and has involved many areas of knowledge, such as mathematics, physics, life sciences, economics, organizational studies, and computing. The theory is based in two premises. First, the modern science does not express the world's absolute reality but an ordered, linear and simplified existence. Second, the integration of transdisciplinary knowledge is needed to solve contemporary problems. In general, complexity theory is based on studies of complex and adaptive systems (YING & SUI PHENG, 2014).

Complex systems are systems based on many interacting parts, where the collective behavior is greater than the sum of the individual actions. Examples include the financial market, the economy, the brain function, the immune system, insect colonies, the internet, and human societies (NEWMAN, 2011).

Complexity is a critical theory for shifting from a Cartesian / Newtonian view of project management to a more complex idea (COOKE-DAVIES ET AL., 2007). The characteristics of complex systems are based on nonlinearity, self-organization, and emerging properties. In this way, scientists of complexity are instructed to understand the nonlinearity of the relationships between system agents, their spontaneous (re-) organization, and the collective characteristics that arise (or emerge) from the agents' interaction. (YING & SUI PHENG, 2014).

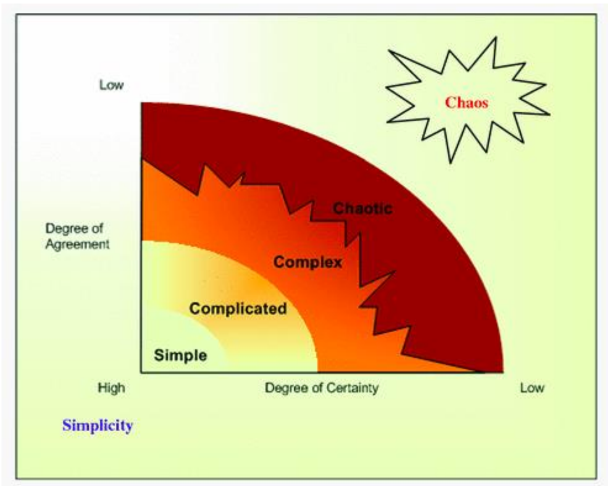


Figure 8 - Stacey Matrix

Source: Ying & Sui Pheng (2014) based on Stacey (2001)

In a simplified way, Stacey (2001) proposes a representation of the complexity level in a decision based on different situations of uncertainty and agreement. Figure 8 shows the different levels of complexity, from simple (certainty and agreement degree are high) to chaos (certainty and agreement degree are low). On the X-axis, the level of uncertainty of the analyzed model is shown. The Y-axis measures the level of agreement on a subject or decision between a group or organization. Stacey matrix provides a guide for decision making by understanding the complexity level of the situation.

1.1.5 Bounded Rationality

Bounded rationality defines the type of rationality that supports decisions in complex environments in which mental or processing capacity are limited, under the framework of evolutionary economics theories (BARNEY ET AL., 1987). Bounded rationality is constructed through the following steps (SIMON, 2000):

1) People pursue multiple possible conflicting objectives.
2) Alternatives to pursue the objectives are not given, so the decision-maker generate the alternatives.
3) The limits of mental (or processing) capacity are not enough to deal with a complex environment. Therefore, decision-makers do not consider all the alternatives.
4) The limits of mental (or processing) capacity also prevent decision-makers from considering the consequences of the alternatives.
5) Decision-makers choose a "satisficing" rather than an optimized alternative due to the lack of information and cognitive capacity.

1.1.6 Types of Decision

The decision means a cut between past and future, an action that causes a change between the situation that existed before and what will exist afterward (SHACKLE, 1969). Tannert et al. (2007) explore decision-making based on his taxonomy of uncertainty, as shown in Figure 9.

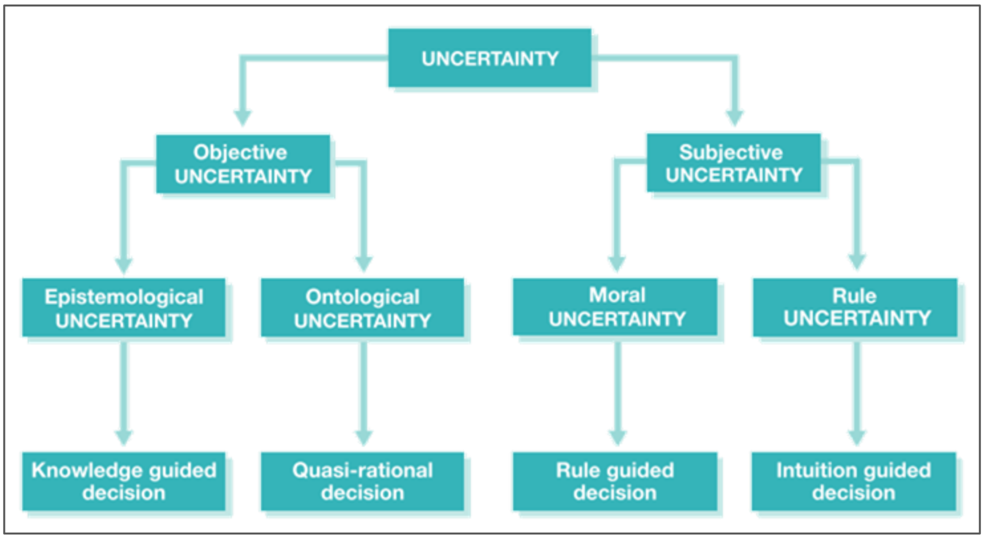


Figure 9 - Taxonomy of uncertainty and decision

Source: Tannert et al. (2007)

Decision-makers may rely on available knowledge and information to make rational "knowledge-guided" decisions in epistemological uncertainty situation. Research and advances in knowledge may close the epistemological uncertainty gap.

In ontological uncertainty situations, non-linear behavior prevents decision-makers from making rational decisions because it is impossible to solve randomness by research or determining reasoning. The random nature of ontological uncertainty supports "quasi-rational" decisions due to some guidance provided by past experiences and the unpredictable effects.

The lack of moral rules makes decision-maker relies on general rules to deduce a guidance for the situation, such as service oaths or behavior guidelines. Tannert et al. (2007) call this situation a "rule-guided" decision, and Varde & Pecht (2018) call a conscience-driven decision.

There are no explicit or implicit moral rules to support the decision in rule uncertainty situations. Decision-makers rely on intuition and in a subconscious level of deduction. It can be attributed to fuzzy or imprecise knowledge (TANNERT ET AL., 2007).

Shackle (1969) also brings some important concepts to understand decision-making under a situation of uncertainty. First, the term situation is established from a temporal view, where a single situation exists only for a moment, with its own characteristics. A story is made from a sequence of situations that succeed each other in time. Furthermore, human history is not determined, which means human decisions themselves modify history. To better explain this idea, Shackle express decision-making characteristics in contexts with different reality assumptions, as seen in the Table 1.

Table 1 - Types of uncertainty

Assumptions of reality	Type of Decision
History is pre-determined	Illusory decision
History is fully known (perfect foresight, no uncertainty)	Empty decision
History doesn't have an order (non-existent foresight, non-limited uncertainty)	Powerless decision

Source: Author's elaboration, based on Shackle (1969) and Metcalf et. al (2014)

In a fictional world where history is already predetermined, there is a destiny already defined for the future, and decisions become illusory. Regardless of the cut that the decision causes in past and future situations, nothing changes in the situation.

When future situations are fully known, a perfect foresight, the possible decisions to be made are characterized as empty decisions. That is, from the moment the detailed impact of each action is known, the decision is already evident, and there is no "choice". In this case, the decision is already known ex-ante, and uncertainty in this context does not exist.

When there is a diametrically opposed situation, when there is no cause-and-consequence order throughout history, and there is no foresight about the decision to be made, the decision is impotent. Without any knowledge about the impact of its actions and without knowing what to expect from our previous knowledge of history, the decision loses its power of direction of the future, uncertainty is unlimited.

However, the spectrum of decisions in which our study begins follows the definition of Shackle: "Decision, we have claimed, is choice, but not choice in the face of perfect foreknowledge, not choice in the face of complete ignorance. The decision, therefore, is chosen in the face of bounded uncertainty" (SHACKLE, 1969, p.5), which means this work will focus on non-illusory, non-empty, non-powerless decisions.

Thus, the decision (non-illusory, non-empty and non-powerless) is a choice in a situation of limited uncertainty.

1.2 Futures studies

This session presents the main object of this work: futures studies. Thinking about the future is inherent to human society. So, this session explores the historical development of thinking about the future up to the modern foresight techniques used today. Then, the types of futures studies and different generic processes for their development are presented. Futures studies methodologies are presented in section 1.2.3. The following sections introduce new tools, competitive gain, and the role of foresight in the digital world.

1.2.1 Historical introduction to futures studies

The search for understanding the future accompanies human history over time. Prophets, oracles, seers, and sorcerers have sought, since ancient times, to understand future developments of society through various forms, such as observation of stars and nature (CUHLS & JOHNSTON, 2008). Deterministic prediction and mystical orientation mark this phase as the first paradigm of futures studies (KUOSA, 2011).

Futures studies have been formalized as a decision-support tool in the post-World War II period (GEORGHIOU ET AL., 2008). Governments and private companies have been attentive to the importance of "forecasting" new technological trends for their security as an institution. The end of the Second World War taught the world the value of good planning, strategies, and management of complex issues. The Stanford Research Institute and Rand Corporation, founded in 1946 and 1948, were critical in popularizing and developing long-term planning, addressing political, social, and technological issues for military and industrial purposes (MILES, 2010).

In the following decades, futures studies focused on the development of engineering with military applications during the Cold War. This period marks the first conceptual basis for futures studies, based on futures' probabilistic analyses and past data extrapolation (also called forecasting).

Some significant associations emerged in the 1970s, such as the World Future Society (WFS) in Washington, the USA, and the World Futures Studies Federation (WFSF) in Paris, France. Two impactful future-oriented works were published in the same period: Limits of Growth by the Club of Rome (MEADOWS ET

AL., 1973) and "Modelo Mundial Latino Americano", by the Bariloche Foundation (HERRERA, 1976). The "Modelo Mundial Latino Americano" was developed to refute the "Limits of Growth" vision of the future, proposing a model based on alternative assumptions. In "Limits of Growth", the authors explore limits imposed on growth based on the physical environment, such as natural resources and consumption. They also defend population control. On the other hand, the "Modelo Mundial Latino Americano" discusses the importance of political and social problems, extolling the world's unequal distribution of power. Both works made projections about the future of humanity and employed quantitative and qualitative methodologies and analysis (ALBORNOZ, 2008).

Still in the 70’s, The Office of Technology Assessment (OTA), the Forecasting and Assessment in the field of Science and Technology (FAST), and the National Institute of Science and Technology Policy (NISTEP) were established in the USA, Europe, and Japan. They aimed to analyze scientific and technology changes globally and propose long-term initiatives for their governments (CAGNIN, 2014; SCHENATTO ET AL., 2011).

In the mid-1980s, future-oriented exercises began to be performed more frequently by national governments and companies. Like Shell and GE, companies also began to use scenarios to assess the future linked to their business and support strategic decisions (ROHRBECK ET AL., 2015).

Approaches based on forecasting and a deterministic view of the world begin to lose strength and give way to more qualitative studies. In this context, La Prospective appears in France, with a qualitative and plural future approach (GODET, 1982). The book Foresight in Science, by the authors John Irvine and Ben Martin (of the renowned Science Policy Research Unit - SPRU), also marks the term "Foresight" as a tool for understanding the forces that shape the long-term future and inform policy formulation, planning, and decision-making (MARTIN, 2010)

Foresight became then a popular tool for science, technology, and innovation (MILES ET AL., 2017). These studies allowed decision-makers to generate policies and technological strategies that align institutional objectives with the perceived technological trends in the world and, in a way, to influence the direction of these trends. The practice of foresight, unlike forecast, does not focus on guessing or predicting the future but on generating and creating guidelines for the expected future to become a reality by aligning and sharing the same future visions with the actors involved in this environment. In this way, it is possible to affirm that foresight exercises go beyond of predicting the future. It explores plausible futures, drive strategy planning and strength network (LOVERIDGE, 2009).

1.2.2 Futures studies types and process

Several frameworks and processes of futures studies were developed from the popularization of its use. Voros (2003) proposes a generic foresight framework based on his projects on implementing foresight in a public-sector university in Australia. He proposes a four-phase process that consists of: 1) Gathering inputs (data); 2) Foresight work; 3) Output; 4) Strategy (Figure 10).

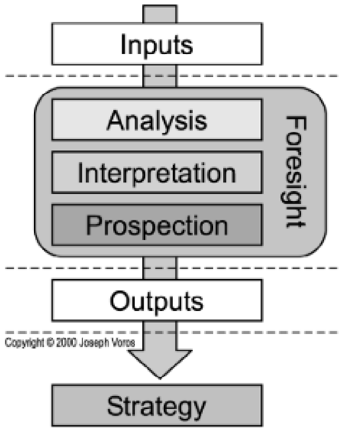


Figure 10 - Generic foresight process

Source: Voros (2003)

The first phase in Voros' process consists of gathering information and scanning for strategic intelligence. This information can be collected by several different methods as Delphi or environmental scanning (CHOO, 2002). This phase's main objective is to collect all data that will be used for analysis, interpretation, and prospection. The second phase comprises three steps. The "analysis" step is considered a preliminary stage to in-depth work, seeking the first data scene and their relationships. "Interpretation" look for develop a deeper structure and insights from data analysis, seeks patterns and causal relationships between variables, events, and trends. The "Prospection" step looks for the future through the analysis and interpretation of data to create forward views. The foresight outputs can comprise the range of options generated by the foresight activity (tangible output) and the changes in thinking, insights, and opinions that the whole process can bring (intangible outputs). At this point, the foresight has generated a real expanded perception of strategic options available. Finally, the strategy is a decision-making phase, in which decision-makers direct strategic actions and make decisions based on foresight outputs.

More focused on the execution and implementation of technology foresight in companies, Reger (2001) proposes a seven-step process (Figure 11). According to the author, the seventh step is not directly understood as part of technology foresight. However, the implementation of the project supported by the study is added to the model due to the importance of foresight studies in decision-making.

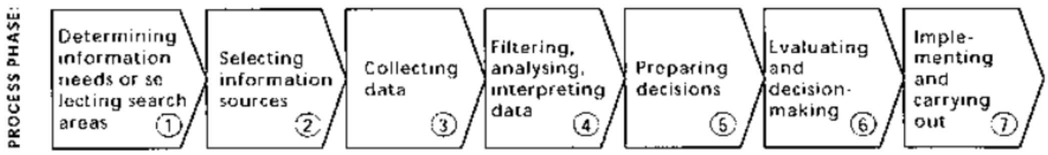


Figure 11 - Foresight phases

Source: Reger (2001)

Determining information need and selecting the search area: in this phase, the objectives, core issues, or search areas are generated. The author highlights two alternative approaches in this stage: "inside-out" or "outside-in". From an "inside-out" perspective, searches and observations will be carried out within the company's technological domain. In the "outside-in" perspective, searches go beyond the company's technical domain and have a vast technological scope.

Selecting information sources: Formal and informal information sources are selected. Formal sources is represented by documents, internal or external companies' boundaries. Informal knowledge is not-written data, often orally transmitted. The selection of methods and tools is also carried out at this stage of the process.

Collecting data: data related to science and technology (articles and patents, for example) are often collected in technology foresight processes because of their utility in presenting future research and technology trends. Data on competitors, suppliers, customers, and universities, as well as interviews and primary data collected from experts in the field, are essential to achieve the proposed objectives.

Filtering, analyzing, and interpreting information: analyzing and interpreting the collected data is crucial for decision making. An interactive and discursive process between project teams, research groups, and specialists is essential so that the correct causal relationships between data are drawn and become relevant future-oriented information.

Preparing decisions: The outputs of the analyses are used to develop action proposals according to the objectives and the area in which the study was designed. The author focuses on proposals with technological impacts, such as new R&D projects, new strategic innovation projects, innovation fields, technology purchase or sale decisions.

Evaluating and decision-making: This step aims to evaluate the proposals developed in the previous step and start the implementation of the selected projects and recommendations. The assessment consists of calculating the costs, risks, and impacts before making the final decision. Decision-makers include the company's strategic committee, program managers, and the sponsor of the study.

Implementing and carrying out: the last step presented corresponds to the implementation, monitoring, and execution of the projects defined in the previous step.

Popper (2008) and Miles (2002) define foresight as a process that involves intense reflection, network consultations, discussions and leads to a joint future vision. They present a foresight approach as a continuous process composed of five steps: Pre-Foresight, Recruitment, Generation, Action, and Renewal.

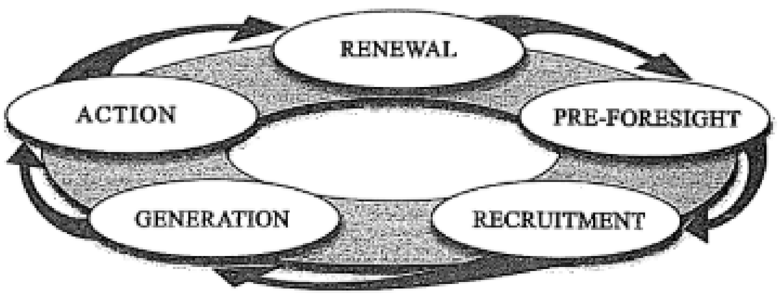


Figure 12 - Continuous foresight process

Source: Popper (2008) and Miles (2002)

On Pre-Foresight, the activities are limited to define rationales and objectives, assembling the project team, and design a methodology. The participation of the study sponsor in the definition of goals is essential for informing future decisions. Once the objectives have been defined, the foresight team is assembled, and the methodological framework is decided internally by the sponsor and foresight team.

Methodological decisions are affected by expertise, political, technological support, infrastructure, and deadlines. As the last task of this phase, the team defines the communication tools. Some methods, such as literature review or bibliometrics, can be used at this stage to identify the objectives.

The core team is responsible, in the Recruitment phase, for recruiting new members (i.e., facilitators, experts, and other stakeholders) and collect necessary data for generating future insights. In general, the recruitment stage is constant throughout the process, in a higher or lower intensity. Methodologies such as stakeholder analysis, brainstorming, bibliometric and patent analysis can be used in this step to identify research groups and relevant data sources.

Generation is considered the heart of the foresight process. This stage of the process focuses on the prospective effort of the exercise. In general, at this stage, existing knowledge is analyzed and synthesized, tacit knowledge is codified, new knowledge is generated, and shared visions of the future are created. The generation phase can be divided into three stages: Exploration, Analysis, and Anticipation. The three steps include understanding the study domain's context, identifying the main issues and drivers of influence, and anticipating possible and desirable futures.

In the Action phase, the foresight process is up to its primary objective, informing decision-making. In some cases, if intermediate outputs are not new or do not stimulate the sponsor, this may result in a lack of action, which means the foresight outputs may not support policy-, strategy-, or decision-making. Efforts to ensure that foresight informs decisions involve: 1) prioritization and decision-making through methods such as polling and multi-criteria analysis, and 2) innovation and change through technology roadmap, narrative scenarios, and backcasting.

The Renewal phase consists of monitoring and evaluating processes to verify if foresight has achieved its goals and prepare the foresight’s new cycle. The development of useful indicators and the systematic tracking of events and results are the main challenges in this phase.

Different approaches to futures studies require different processes for their execution and implementation, according to the study's objective. Below we present the terms most frequently used for different approaches to futures studies.

- Forecasting

The first formal futures studies were based on the idea that it was possible to predict the future through rational analysis supported by available data. According to Martin (2010), forecasting is based on probabilistic statements about the future, with relatively high confidence. In general, forecasting uses mathematical tools to analyze historical series and extrapolate it to the future (SCHENATTO ET AL., 2011). In some situations, it is possible to see the term "forecasting" is also used by practitioners to refer to qualitative studies, such as scenario analysis.

- La Prospective

The approach known as "La Prospective" emerged in France in 1970 and was led by Godet (1982). According to the author, this framework was born to oppose the mostly quantitative approaches (forecasting), bringing exploratory and normative futures approach to prospective exercises. Thus, "La Prospective" analyzes the actions of the present to collaboratively create possible and desirable futures, seeking anticipation. The term "La Prospective" is frequently used as a synonym of strategic foresight.

- Technology Assessment

The concept of technology assessment was developed by the Office of Technology Assessment (OTA) of the USA. It refers to studies evaluating emerging technologies' potential. Technology assessment aims to anticipate the likely benefits and adverse impacts of implementing technology. It is also part of the technology assessment scope to establish cost-benefit analysis according to replaceable technologies and follow technological trajectories (BLAIR, 1994; SCHENATTO ET AL., 2011). Such an approach is intensely used for regulatory, economic, social purposes, and assessing technologies (mainly healthcare).

- Future oriented Technology Analysis - FTA

Porter et al. (2004) present the concept of Future-oriented Technology Analysis (FTA) as a broad set of techniques focused on analyzing future technology, such as technological intelligence, forecasting, roadmapping, assessment, and foresight. FTA is primarily applied for elaboration science, technology, and innovation policies, through the judgment of characteristics of emerging technologies, developments in technological trajectories, and potential impacts of technology in the future (EEROLA & MILES, 2011).

- Horizon Scanning

Horizon Scanning is a methodology for systematic analysis of potential future problems, threats, opportunities, and likely developments. Horizon Scanning explores new and unexpected issues, as well as persistent problems, trends, and weak signals (MILES & SARITAS, 2012). Horizon Scanning can be seen as an instrument with two functions: 1) alert and assist policymakers in anticipating emerging issues and 2) support creativity to reassembly or create new emerging issues through the analysis and integration of the scanning data. Thus, Horizon Scanning supports the formulation of new policies and the evaluation of already implemented policies in light of new emerging issues (AMANATIDOU ET AL., 2012)

- Foresight

For Saritas & Burmaoglu (2015), the foresight differentiates itself from other futures studies by 1) foresight does not have the claim to be a prediction, its main emphasis is on exploring multiple and possible futures, with less deterministic and more prospective bias, 2) foresight is based on methods that are typically tested and repeated to deal with uncertainty and complexity through quantitative and qualitative approaches, developing policy visions and strategies, and 3) foresight is a collaborative and participatory process and seeks diversity, involving diverse points of view and generating a common and shared vision among stakeholders.

In recent years the term foresight has been widely used as a description for a series of studies that support decision making at both governmental and corporate levels, as well as generating a common direction among the main agents and resources in their development efforts (MILES ET AL., 2008, 2016).

Miles et al. (2008) define five general rationales for foresight studies:

1. Direct or prioritize investments in science, technology, and innovation;
2. Create new contacts and links between the actors around a shared vision;
3. Expand the breadth of knowledge and ideas for the future;
4. Bring new players to the strategic debate;
5. Improve policy development and strategy formation in areas where science and innovation play a significant role.

Many other variants of foresight have been developed in recent years, such as Strategic Foresight (HAARHAUS & LIENING, 2020; IDEN ET AL., 2017), Fully-Fledged Foresight (MILES, 2010), Ethical Foresight (FLORIDI & STRAIT, 2020), and Foresight for Science, Technology and Innovation - ForSTI (MILES ET AL. 2016).

Corporate Foresight is defined by (ROHRBECK & KUM, 2018) as a series of foresight practices that allow firms to obtain a good position in future markets and create future perspectives shared among the management team.

1.2.3 Futures studies methodologies

The methodology is a fundamental element of foresight, by guiding the actions for exploring new ideas, thinking about the future and supporting decision-making. The fundamental goal of a foresight methodology is to transform it into a systematic activity, with clear inputs, processes and outputs (SARITAS, 2013).

Popper (2008) identifies 33 most used methods in foresight and classifies them according to their type: qualitative, quantitative and semi-quantitative.

Qualitative methods provide meaning to events and perceptions. Such interpretations tend to be based on subjectivity or creativity that is often difficult to corroborate (e.g. opinions, brainstorming sessions, interviews). Table 2 presents a short description of the 19 qualitative methods listed in alphabetical order.

Table 2 - Qualitative methods

Backcasting is an approach that involves working back from an imagined future, to establish what path might take us there from the present. One version of backcasting involves simulation modelling – indeed, this method is much employed with planning models. More commonly, backcasting is used in aspirational scenario workshops.
Brainstorming is a creative and interactive method used in face-to-face and online group working sessions to generate new ideas around a specific area of interest. Aiming at removing inhibitions and breaking out of narrow and routine discussions, it allows people to think more freely and move into new areas of thought, and to propose new solutions to problems.
Citizen Panels are groups of citizens (members of a polity and/or residents of a particular geographic area) dedicated to providing views on relevant issues, often for a regional or national government. The panel is more than a conventional opinion survey, since its members are encouraged to deepen their understanding of the issues involved.
Conferences/Workshops are events or meetings lasting from a few hours to a few days, in which there is typically a mix of talks, presentations, and discussions and debates on a particular subject. The events may be more or less highly structured and “scripted”: participants may be assigned specific detailed tasks or left very much to their own devices.
Essays/Scenario Writing involves the production of accounts of “plausible” future events based on a creative combination of data, facts and hypotheses. This activity requires insightful and intuitive thinking about possible futures, normally based on a systematic analysis of the present.
Expert Panels are groups of people dedicated to analyzing and combining their knowledge concerning a given area of interest. They can be local, regional, national or international. Panels are typically organized to bring together “legitimate” expertise, but can also attempt to include creative, imaginative and visionary perspectives.
Genius Forecasting is an activity carried out by respected individuals that requires both expertise and creativity in relatively similar proportions. It involves the preparation of forecasts based on insights of a brilliant specialist, scientist or authority in a given area.
Interviews are often described as “structured conversations” and are a fundamental tool of social research. In foresight they are often used as formal consultation instruments, intended to gather knowledge that is distributed across the range of interviewees. This may be tacit knowledge that has not

been put into words, or more documented knowledge that is more easily located by discussions with experts and stakeholders than by literature review.
Literature Review (LR) represents a key part of scanning processes. Good reviews generally use a discursive writing style and are structured around themes and related theories. Occasionally the review may seek to explicate the views and future visions of different authors.
Morphological Analysis is closely related to relevance trees, and to the soft-systems approach since it helps both complex problem-solving and management of change; it may be used in planning or scenario development. It maps promising solutions to a given problem and determines possible futures accordingly: the classic applications have involved systematically working through the entire range of conceivable technological solutions for a particular goal (such as attaining a manned mission to the moon).
Relevance Trees and Logic Charts are methods in which the topic of research is approached in a hierarchical way. Each begins with a general description of the subject and continues with a disaggregated exploration of its different components and elements, examining particularly the interdependencies between them.
Role Play/Acting requires reflection, imaginary interaction and creativity. The method tries to answer questions such as: If I were person X, how would I deal with problem Y? Or, if we were country X, what would be our position with regards to issue Y?
Scanning (often termed "environmental scanning" or "horizon scanning") involves observation, examination, monitoring and systematic description of the technological, socio-cultural, political, ecological and/or economic contexts of the actor in question – a country, industry, firm, organization, etc.
Scenarios refers to a wide range of approaches involving the construction and use of scenarios – more or less systematic and internally consistent visions of plausible future states of affairs. Generally, scenarios involve several features of the object of study, not just one or two parameters. They may be produced by means of deskwork, workshops or the use of tools such as computer modelling.
Science Fictioning (SF) is an activity that deals with stories assuming that possible events which have not yet materialized have taken place, usually at some point in the future, and elaborates on the consequences of this. Because it involves fictional narrative – and much commercial science fiction is driven more by the need to have adventure or surprise – the method is not very commonly linked to serious governmental or business policymaking. However, it is quite common for scenarios to be illustrated in reports by brief vignettes which use SF-like techniques to illustrate one or other point of the imagined future world.
Simulation Gaming is one of the oldest forecasting and planning techniques, in that war gaming has long been used by military strategists. It is a form of role-playing in which an extensive "script" outlines the context of action and the actors involved. There have long been technological aids used here, such as model battlefields, and now computer simulations.
Surveys, like interviews, are a fundamental tool of social research, and are widely used in foresight. A questionnaire is distributed or made available online, and responses drawn from what is usually hoped to be a large pool of respondents. High participation rates generally require attractive and clear design of the survey instrument.
SWOT Analysis is a method which first identifies factors internal to the organization or geopolitical unit in question (resources, capabilities, etc.) and classifies them in terms of Strengths and Weaknesses. It similarly examines and classifies external factors (broader socio-economic and environmental changes, for example, or the behavior of opponents, competitors, markets, neighboring regions, etc.) and presents them in terms of Opportunities and Threats.
Wild Cards & Weak Signals (Wi-We) are types of analysis that are usually carried out by small groups of highly skilled people capable of combining expertise, examining data and creative thinking. Wild Cards are situations/events with perceived low probability of occurrence but potentially high impact if they were to occur. Weak Signals are unclear observables warning us about the probability of future events (including Wild Cards). They implore us to consider alternative interpretations of an issue's evolution to gauge its potential impact.

Source: Popper (2008, pp. 44-88)

Quantitative methods measure variables and apply statistical analyses, using or generating reliable and valid data (e.g. socio-economic indicators, papers and patents). Many quantitative methods are used in foresight to provide an evidence base for futures thinking, or supply forecasting tools like trend extrapolation. Table 3 provides a short description of the 6 quantitative methods listed in alphabetical order.

Table 3 - Quantitative methods

Benchmarking is a method commonly used for marketing and business strategy planning and has recently become more popular in governmental and inter-governmental strategic decision-making processes. The main question here is what others are doing in comparison to what you are doing. **Bibliometrics** is a method based on quantitative and statistical analysis of publications. This may involve simply charting the number of publications emerging in an area, perhaps focusing on the outputs from different countries in different fields and how they are evolving over time. Impact analyses examine citations to assess, for example, the most influential pieces of work in specific areas. **Indicators/Time Series Analysis (TSA)** involve the identification of figures to measure changes over time. Indicators are generally built from statistical data with the purpose of describing, monitoring and measuring the evolution and the current state of relevant issues. **Modelling** generally refers to the use of computer-based models that relate together the values achieved by particular variables. Very simple models may be based on statistical relations between two or three variables only – even extrapolation is an elementary form of modelling (in which time is one variable). More complex models may use hundreds, thousands, or even more variables; econometric models are routinely used in economic policymaking, for example, and are "calibrated" from economic statistics and statistical analyses of their interrelations. **Patent Analysis** often resembles bibliometrics but uses patents rather than publications as its starting point. Quantitative analysis utilizes statistical methods to look at the number of patent registrations, assuming that increasing or decreasing registrations would (apparently) indicate, for example, low or high potential for technology developments in a specific area. More qualitative analyses may focus more on the contents of the patents. **Trend Extrapolation/Impact Analysis** are among the longest-established tools of forecasting. They provide a rough idea of how past and present developments may look like in the future – assuming, to some extent, that the future is a kind of continuation of the past. Recently, the concept of Megatrends has become popular to refer to macro level phenomena which include various (sometimes conflicting) sub-phenomena (e.g. globalization, ageing, climate change). On the other hand, Impact Analysis aims to identify potential impacts that major trends or events would have on systems, regions, policies, people, etc.

Source: Popper (2008, pp. 44-88)

Semi-quantitative methods apply mathematical principles to manipulate data derived from subjectivity, rational judgements, probabilities and viewpoints of experts and commentators (i.e. weighting opinions or probabilities). Table 4 provides a short description of the 8 quantitative methods listed in alphabetical order.

Table 4 - Semi-quantitative methods

Cross-impact/Structural Analysis (SA) attempts to work systematically through the relations between a set of variables, rather than examining each one as if it is relatively independent of the others. SA requires that a set of key variables are determined in order to understand the system that is of concern. Usually, expert judgement is used to examine the influence of each variable within a given system, in terms of the reciprocal influences of each variable on each other – thus a matrix is produced whose cells represent the effect of a variable on each other. **Delphi** is a well-established technique that involves repeated polling of the same individuals, feeding back (sometimes) anonymized responses from earlier rounds of polling, with the idea that this will allow for

better judgements to be made without undue influence from forceful or high-status advocates. The technique was developed so as to circumvent "follow the leader" tendencies of face-to-face exchanges, and other problems such as the reluctance to discard previously stated opinions. **Key/Critical Technologies** methods involve the elaboration of a list of key technologies for a specific industrial sector, country or region. A technology is said to be "key" if it contributes to wealth creation or if it helps to increase quality of life of citizens; is critical to corporate competitiveness; or is an underpinning technology that influences many other technologies. However, the method is implemented (expert panels or surveys, for instance) it implies some prioritization process (such as voting, multi-criteria and/or cross-impact analysis). **Multi-Criteria Analysis** is a prioritization and decision-support technique specially developed for complex situations and problems, where there are multiple criteria in which to weigh up the effect of a particular intervention. The method works by asking participants to assess the importance of various evaluative criteria, and the impact of a series of options, policies or strategies in each of the criteria. **Polling/Voting** refers to the use of voting or survey methods to gain an assessment of the strength of views about a particular topic among a set of participants. These may be members of a workshop, for example, who make a show of hands, place post-it stickers on one or other category on wall posters, enter views into a computer system, etc., to indicate how probable, uncertain, or important they consider events to be, which actions are priorities and how feasible alternatives are, and so on. **Quantitative Scenarios/SMIC** take various forms. One version involves quantification of the contingencies that bring about the scenario. Sometimes probabilistic analysis is established via expert opinion in order to build a system which evaluates the likelihood of occurrence of certain events. **Roadmapping** is a method which outlines the future of a field of technology, generating a timeline for development of various interrelated technologies and (sometimes) including factors like regulatory and market structures. It is a technique widely used by high-tech industries, where it serves both as a tool for communication, exchange, and development of shared visions, and as a way of communicating expectations about the future to other parties (e.g. sponsors). **Stakeholder Analysis/MACTOR** are strategic planning techniques which take into account the interests and strengths of different stakeholders, in order to identify key objectives in a system and recognize potential alliances, conflicts and strategies. These methods are quite common in business and political affairs. In futures work, there are techniques such as MACTOR that take this further, systematically considering whether stakeholders are in favor of or against particular objectives and representing the situation in terms of matrices that can be formally analyzed.

Source: Popper (2008, pp. 44-88)

Based on the source of knowledge of the methodology, Popper (2008) also classifies the methods in two different axes: from creativity- to evidence- and from expertise- to interaction-based methods. Thus, the Foresight Diamond is presented in Figure 13, which shows the 33 methods classified by type (qualitative, quantitative, or semi-quantitative) and the source of knowledge (expertise, interaction, creativity, and evidence).

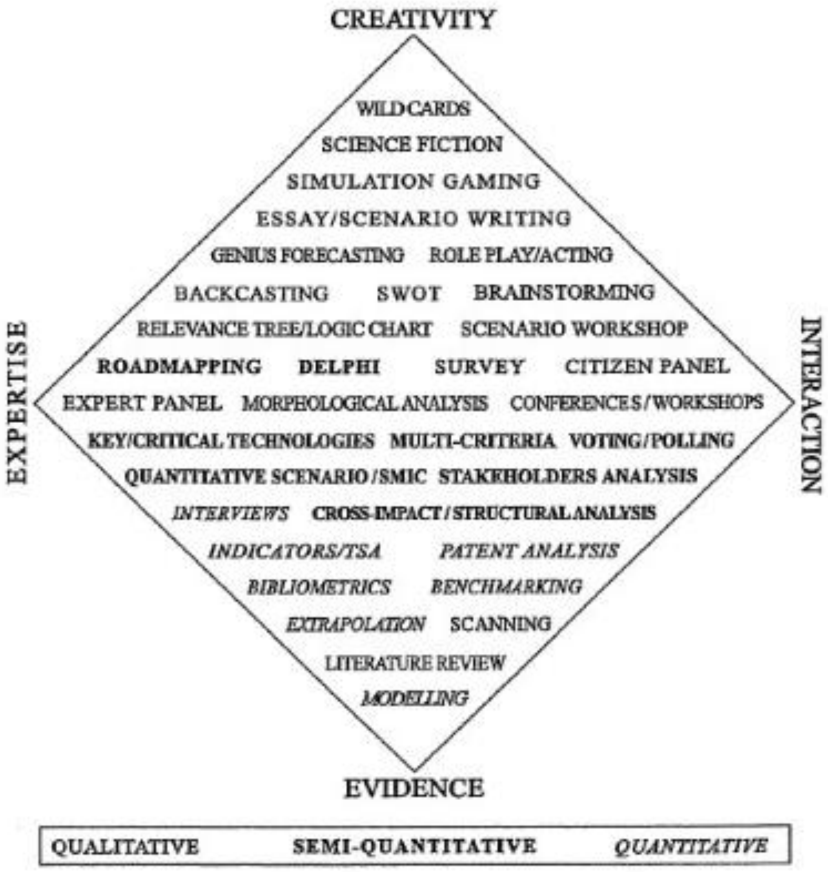


Figure 13 - Foresight Diamond

Source: Popper (2008, pp. 44-88)

Creativity-based methods involve imaginative thinking in two distinctive forms: 1) inventiveness of very skilled individuals (e.g., genius forecasting) and 2) inspiration from group discussions (e.g., brainstorming).

Expertise-based methods rely on individuals' or groups of experts' knowledges. They are frequently used to support top-down decisions.

Interaction-based methods are based on the interaction of experts and non-experts' stakeholders, prioritizing "democratic" ideals and legitimacy. They often support bottom-up, participatory, and inclusive decisions.

Evidence-based methods rely on to explain or forecast a particular phenomenon through reliable means of analysis.

The methodological choice for a foresight study is not a trivial decision and has a significant impact on the study process and results. Saritas & Burmaoglu (2015) define a set of criteria used to make this decision: proof of concept, resources, deadlines, level of participation, engagement, combination with other methods, study's objectives, data, and competence. Popper (2008) adds intuition, insight, impulsivity, inexperience, or irresponsibility as influence factor on practical methodological choice, pointing to the importance of choosing an appropriate mix of methods.

Some methodologies are often more used than others, as shown in Table 5, based on a study (EUROPEAN COMMISSION, 2009) involving more than 2000 foresight exercises from 2004 to 2008:

Table 5 - Foresight methodologies and frequency of use

Methodology	Frequency of use
Literature Review	Most frequently used
Expert Panels	
Scenarios	
Trend analysis	Commonly used
Interviews	
Delphi/Surveys	
Key Technologies	
Scanning	
Roadmapping	Less frequently used
Modeling/Simulation	
Bibliometrics	
Morphological analysis	
Gaming	
Multicriteria Analysis	

Source: European Commission (2009)

1.2.4 New approaches and limitations to futures studies

Practitioners are developing new approaches to incorporate new elements into prospective studies, such as Big Data and Machine Learning. Digital resources have revolutionized the way we produce knowledge in universities, governments, and companies. The available data and the increased processing capacity open a new panorama of analysis and decision making in complex environments.

The advantages of using Big Data in futures studies includes collecting large volumes of high-quality data from various sources without large investment (APREDA ET AL., 2016). Real-time data availability allows studies to reflect the real and instant scenario of the area by increasing the forecasts' accuracy. These tools also allow the substitution of traditional data sources, such as surveys, to prepare reliable futures studies with improved data analysis support (GINSBERG ET AL., 2009).

Some examples of this use can be mentioned. In their work, Apreda et al. (2016) adopt a methodology used in engineering, called Functional Analysis, in the context of technological foresight. This new methodology correlates the user's perspective with the products' technical specifications. It investigates product functionalities and technologies by a bibliometric analysis (papers and patents) to identify each technological trajectory. In this way, the authors offer a methodology accessible to small and medium companies to popularize this type of study and apply a digital tool to analyze bibliographic (big) data.

The International Center for Tropical Agriculture is an organization that aims to strengthen agricultural technologies, innovation, and knowledge that can help small farmers. In partnership with Colombian institutions, weather and harvest data from the last decades in Colombia were collected. The analysis of such data helped farmers in the Colombian region of Cordoba to predict climate change and explain the limitations of crop growth due to solar radiation and sensitivity to warm nights. From the analysis, a low-cost solution was found so that the farmers had fewer losses of production: to sow the harvests in the correct periods. Such a foresight study helped 170 farmers avoid the loss of $ 3.6 million and improve rice plantations' productivity from 1 to 3 tons per hectare. Through big data techniques and data analysis, such as statistical models and machine learning, it was possible to achieve such outcome. (CARIBONI, 2014; CCAFS, 2015; HILBERT ET AL., 2016).

IBM Smarter Cities Challenge is an IBM pilot project for real-time monitoring of some cities worldwide, providing relevant data, analysis, and information to municipal operational centers. Through a public-private partnership, the city of Rio de Janeiro has created a system for analyzing data from various sources, such as traffic, public transport, utility services, emergency services, and wheatear information, as well as unstructured data provided by citizens themselves as social media data. All this information is centralized in a command center. It enables policymakers and administrators to gain access to a better understanding of the city's dynamics and to find hidden correlations to create assertive policies, to predict crimes and natural disasters (HILBERT ET AL., 2016)

Thus, new processes are created from the development and application of Big Data and Machine Learning tools to foresight. Still, it is also possible to note that many methods are reshaped to introduce massive data to support prospective studies or technological decisions. This is made by adopting such tools in some part of the already consolidated methodology or by combining methods that use massive data in their processes (YUFEI ET AL., 2016).

Despite the great benefits and opportunities that can be achieved using information and communication technologies (ICT) tools for foresight, especially for developing countries and small businesses, some limitations must be highlighted. The use of data that reflects the past is inherent in any study that uses data to predict the future, but with Big Data's help, such a feature stands out. The reason for considering this point as a limitation is that reality is in constant change and at the mercy of countless variables, more variables than it is possible to analyze, even with the most technological computational tools already made. Thus, the tools available today can only detect the patterns that have occurred in the past. The successful projection of these same patterns in the future depends on whether the past and the future follow the same logical trajectory without significant discontinuities.

Among the significant limitations of using massive data is the unreliability that manipulating such data can bring. Data can be used to show a causal relationship between two "facts", but also massive data can create a false causal relationship and consequently generate false conclusions.

Fake news is briefly defined as news articles or data "which are intentionally and verifiably false and could mislead readers" (ALLCOTT & GENTZKOW, 2017). It was one of the most prominent issues during the 2016 US elections. Also, it showed how an element of increasing uncertainty in the context of a society based on digital information. As they have recently become known, deep fakes are deep learning applications popularized through the manipulation of human images and sound to generate fake content. It can combine anyone's faces, bodies, and voices with pre-existing videos and raise uncertainty and data reliability.

1.2.5 Competitive gain on the use of foresight

The beginning of the use of foresight as a tool for future planning and preparation goes back to the late 1940s, having its golden age in the 1950s mainly with the work of "La Prospective" School of Gaston Berger in France and Herman Kahn of the Rand Corporation in the US (ROHRBECK ET AL., 2015). Despite the long tradition of using such tools, studies of such practices’ impact on firm performance are still relatively scarce, mainly due to the difficulty of establishing a direct causal connection between a foresight study and long-term performance impact. The causality of success is often more directly associated with other factors such as macroeconomics (ROHRBECK & KUM, 2018).

Ashton et al. (1994) state that monitoring science and technology through prospective studies brings significant gains for the firm by keeping it abreast of technological developments at the national and international level, as well as avoiding duplicate research, promoting rapid response to competitors' developments, and increases overall business competitiveness.

Anderson (1997) also explores the possible competitive advantages that foresight studies can bring to a government or company in his article "Technology Foresight for a Competitive Advantage". He explores the UK government Foresight Program and highlights the program's importance primarily in identifying priorities for science, technology, and innovation infrastructure and strengthening the ties and commitment of actors in a shared vision for the future.

Developing and implementing a futures study is usually quite costly, and therefore the benefits of such a study should outweigh the embedded costs (THOM, 2012). Measuring the direct impacts of foresight is difficult but it is essential to improve innovation management and decision-making processes (BALACHANDRA & FRIAR, 1997). Some metrics proposed in the literature based on corporate foresight are:

- Monetary risk estimation compared to the cost of foresight for cost reduction through foresight (SLAUGHTER, 1996)

- Calculation of return on investment in competitive intelligence based on foresight measurable results (DAVIDSON, 2001)

- Increasing shareholder value as a result of a good opportunity and threat strategy (THOM, 2012)

Rohrbeck & Kum (2018) also, analyze competitive gains of foresight through a longitudinal analysis of 42 firms. Data on future preparedness and firm performance were collected in 2008 and 2015 and compared to understand how future preparedness influences company results and performance after eight years. Future preparedness is calculated by comparing two measures: firm foresight maturity level and firm foresight-need level. To calculate the maturity model, Rohrbeck (2010), collects data about information, networks, people, methods, culture, and corporate foresight in the company. To calculate foresight-need level, he collected data about the complexity and volatility of the environment. Both measures are compared and result in categories of future preparedness. In the first category, "vigilant", the level of maturity and need for corporate foresight is equal. In the category "neurotic", the maturity is more significant than the need level of corporate foresight. In "vulnerable", the maturity is smaller than the need level of corporate foresight. Finally, in the category "in danger", maturity is much smaller than the need level of corporate foresight. The firm's performance was recorded as profitability (EBITDA) and Market Capitalization Growth, which resulted in two categories: outperformers, average, and underperformers.

Through the analysis of the data, a positive relationship was found between future preparation in 2008 and the company's performance in 2015. Among those classified as "outperformers" in 2015, 63% were classified as vigilant in 2008. There was also an increase in profitability. In the sample, 40% of the vigilant firms increased their profitability, 55% maintained the same level, and only one firm decreased their position. Among those companies with deficiencies in future preparedness, only 9% climbed the profitability performance rating, 67% maintained the same level, and 24% decreased their position. Thus, the author concludes that there is evidence of a strong relationship between the positive impact of corporate forecasting and companies' performance.

1.3 Conclusion

The first chapter of this thesis intended to present the first part of the theoretical framework that supports this work. This chapter introduced concepts of uncertainty and its role in futures studies through a literature review. Some uncertainty-related concepts are also discussed, such as risk, ambiguity, intuition, ergodicity, and complexity. Futures studies' definitions, types, methods, and competitive gains are presented in section 1.2.

Uncertainty is a relevant topic in several research fields (WAKEHAM, 2015). Objective uncertainty assumes uncertainty as a feature of life and time, caused by gaps in knowledge. Epistemological (related to the available information) and ontological uncertainty (related to the world's stochastic aspect) are dimensions of objective uncertainty. Subjective uncertainty is a phenomenological, cognitive, and emotional aspect to be experienced by the individual. Moral (related to the lack of moral rules) and rule uncertainty (related to the lack of rule itself) compose subjective uncertainty (TANNERT ET AL., 2007). Economists, such as Keynes (1921), Knight (1921), and Shackle (1969) tend to understand quantified uncertainty as risk and unquantified as fundamental uncertainty. Dosi & Egidi (1991) distinguish substantive and procedural uncertainty based on the lack of information for the former and the limitation of cognitive capacity for the latter.

Uncertainty is a crucial concept to comprehend decision theory. Shackle (1969), based on his potential surprise theory, distinguishes illusory, empty, and powerless decisions. A decision is illusory in a determined world; a decision is empty in an uncertainty-free world; a decision is powerless in a world without discernible order. Non-powerless, non-empty, and non-illusory decisions are modeled by bounded ignorance and potential surprise (METCALFE ET AL., 2014).

These elements are essential to understand the role of futures studies in an innovation-driven economy. Since the post-World War II period, futures studies have gained relevance in forecasting new technological trends for governments and institutions. Several futures study structures have been developed to support planning, strategy, and decision-making. Forecasting, prospective, technology assessment, future-oriented technology analysis, horizon scanning, and foresight are types of futures studies established to address different future-oriented goals, such as bringing new players to strategic debate, identify emerging technologies, and future social needs. Popper (2008) classifies foresight approaches in qualitative, quantitative, and semi-quantitative methods.

The speed and complexity of technological evolution bring new and intense challenges for institutions such as companies and governments. Identifying the next technological steps and future perspectives of human behavior in this new scenario have made many of the traditional methods of prospecting rethink strategies and tools to enhance the reliability in exploring future scenarios and thus providing a competent view of the world. Big Data and Machine Learning-based techniques are integrated into futures studies methodologies to analyze and process massive data to reduce substantial and procedural uncertainty within the exercise.

Chapter 2 - Evolution and Application of Big Data and Machine Learning

This chapter will introduce Big Data and Machine Learning in order to explain their impact in futures studies, by supporting data acquisition, analysis, and processing. This chapter's main objective is to present concepts, applications, and limitation of these technologies in different contexts. A literature review of articles, books, and academic papers is carried out to achieve the objectives of this chapter.

2.1 The fourth industrial revolution

The integration of developing countries in the global economy, the decline of poverty and the expansion of middle class are elements that characterize the world since the 1990's. The new economic panorama has changed with China's rise, demanding new raw materials, new products, and dynamizing the economy in developing countries. However, the economic progress is still uneven, and differences in infrastructure, technology, production process, and salaries are even higher in different parts of the world. In this context, digital technologies' development and diffusion make a change, mainly for developing economies. This new economic model, also called Industry 4.0, has a potential impact on jobs, wealth creation, value chain and can conduct a new sustainable and inclusive way of industrialization, with higher energy efficiency and lower environmental impact (PRIMI & TOSELLI, 2020)

The term Industry 4.0 was first proposed during the Hannover Fair in 2011 and it was officially announced in 2013 as a German strategic initiative to revolutionize the German manufacturing sector (XU ET AL., 2018). According to Lukac (2016), the first industrial revolution started at the end of the 18th century and introduced mechanical production plants based on water and steam power. The second industrial revolution started at the beginning of the 20th century and was based on electrical energy. The third industrial revolution began in the 1970's, based on the automation of production, electronics, and internet technology. Now, the fourth industrial revolution or Industry 4.0 is represented by the automation technologies in manufacturing systems and, according to Industrial Internet Consortium, Fact Sheet[1] (2013), is "the

[1] Consortium II. Fact Sheet, 2013. Available from: http://www.iiconsortium.org/ docs/IIC _ FACT _ SHEET.pdf.

integration of complex physical machinery and devices with networked sensors and software, used to predict, control and plan for better business and societal outcomes."

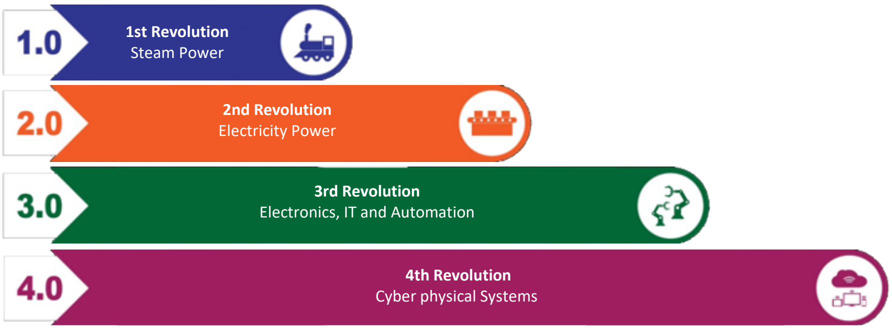


Figure 14 - Industrial revolutions over time

Source: Author's elaboration, based on Xu et al. (2018)

The recent developments in information and communication technologies (ICT) have fueled the emergence of Industry 4.0 over the last few years and allowed companies to digitalize and integrated their production process. Various ICT based technologies can be used to implement Industry 4.0 as cyber-physical systems (CPS) production, Cloud-based manufacturing, and Internet of Things (IoT) (HERMANN ET AL., 2016; KAGERMANN ET AL., 2013; LASI ET AL., 2014; MOEUF ET AL., 2018).

2.1.1 Technologies 4.0

Internet of Things (IoT) first emerged as identifiable interoperable connected objects using radio-frequency identification (RFID) technology, capable of tracking objects attached to RFID tags in real-time (ASHTON, 2009; XU ET AL., 2014). Later, other technologies were applied to IoT applications as sensors, actuators, Global Positioning System (GPS), and mobile devices connected via Wi-Fi, Bluetooth, cellular network, and near field communication (NFC). IoT is generally defined as a global network infrastructure where physical and virtual “things” or objects have identities, attributes, and virtual personalities, using intelligent interfaces integrated into information networks (KRANENBURG ET AL., 2011). Many industries apply IoT technologies, such as transportation, package delivery, healthcare, materials management, retailing and defense, for tracking and monitoring industrial processes. Industry 4.0 uses IoT techniques by applying data science and analytical models to analyze data from multiple machines, systems, and intelligent sensors for real-time processing, optimizing, controlling and monitoring manufacturing processes.

Cloud Computing is a computational technology that offers high performance and low cost by using cloud computing centers for storage and computation, facilitating manufacturing and production. Cloud Computing provides resource sharing, dynamic allocation, flexible extension, and other advantages. Cloud-based manufacturing contributes significantly to Industry 4.0 by enabling modularization and service-orientation on manufacturing, using a network of resources in a highly distributed way. Related concepts have emerged in this scene, such as Manufacturing-as-a-Service (MaaS) and Cloud design, applying Cloud Computing technologies to the manufacturing sector to co-create and personalize products. Cloud Computing also influences the decision-making process since many data and information are uploaded and stored in cloud servers. This technology can support complex decision-making activities.

Cyber-physical Systems (CPS) are the core foundation of Industry 4.0. CPS are engineered systems that integrate computational algorithms and physical components, presenting a higher integration and coordination level among these elements. Microcontrollers acting over sensors and actuators support CPS, allowing information and data exchange among computer terminals, wireless applications, clouds, houses, and factories. CPS-based production systems are the critical element to build smart factories, providing real-time, resources, and costs advantages compared to traditional production systems (GTAI, 2014).

The impact of ICT development on industry significantly affects industrial processes and production (XU ET AL., 2018). According to Kaynak (2005), industrial integration is catalyzed by the new ICT era in an on-going process since the Third Industrial Revolution. He states that the first half of the 20th century was referred to as hardware dominated, which means the speedy and accuracy increase of industrial machinery was mostly due to mechanical parts improvements. The second half was referred to as software dominated, which means the control systems and the advances in speedy and accuracy were based on software development such as Computer-Aided Design and Computer-Aided Manufacturing. Electronic control and signal processing replaced mechanical controls and switches in the era of industrial electronics. As a result of this, in the last few decades, the fusion of different technologies eroded the boundaries between industrial sectors and academic disciplines, between products and services, between producers and users. The transition of traditional industrial ecosystems to Industry 4.0 will require more than only ICT implementation, but new business models and the development of intra- and inter-organizational levels (COLOMBO ET AL., 2015).

However, some technical challenges needed to be addressed in implementing Industry 4.0. Existing ICT structures are not entirely ready to support the digital transformation for Industry 4.0 (DELLOITE, 2015; LIAO ET AL., 2017; OLIVEIRA & ÁLVARES, 2016). IoT scalability also becomes an issue when more and more physical and virtual objects are connected to the manufacturing network, which may cause a lack of efficiency in data transformation and communication. In addition to technical challenges, the standardization of this new global language of production, which Industry 4.0 means to be (GTAI, 2014), and industrial integration will require high international cooperation and system-level perspective (LIN ET AL., 2013; WANG ET AL., 2018). Some efforts have already been made in this context, such as developing the Reference Architecture Model for Industry 4.0 (RAMI 4.0). RAMI 4.0 is a standard for Industry 4.0 introduced by the German Electrical and Electronic Manufacturers' Association and The Industrial Internet Reference Architecture (IIRA), a standard-based open architecture introduced by the Industrial Internet Consortium (ROJKO, 2017). Security and privacy issues are also significant challenges since Industry 4.0 will manage a massive amount of personal, sensitive, and private data that, if not protected, could be hacked by cyber attackers.

Still, on the current limitations of Industry 4.0, data science and analytics can also become an issue because a massive amount of real-time data will be automatically produced. Big Data management, including data mining, data classification, and data storage, is essential to provide valued information from the collected data, mainly for decision-making. New data science and data analytics techniques should be developed, as machine and deep learning algorithms (LU, 2017; MIŠKUF & ZOLOTOVÁ, 2016).

2.2 What is Big Data

This section will describe the main characteristics that define Big Data and its most common data analytics methods.

2.2.1 The concept of Big Data

The term 'big data' has uncertain origins. Diebold (2013) states that the name probably originated in a lunch-time conversation at Silicon Graphics Inc. (SGI) in the mid-1990s, highlighting the figure of John Mashey. However, in 2012 Harris Interactive, on behalf of the company SAP, performed an online survey with 154 C-level global executives about their understanding of new concept of big data. They had collected different interpretations of the idea, focusing on what it is and what it does (GANDOMI & HAIDER, 2015).

Big Data refers to using a massive set of data in terms of acquiring, managing, and processing information. Loukides (2010) defines Big Data as a set of data whose size becomes a problem, and the usual collection, storage, management, and analysis tools do not fit correctly. Nowadays, it is widely accepted in the literature the definition of 3 V's of Big Data: Volume, Variety, and Velocity (LOUKIDES, 2010; MCGUIRE ET AL., 2012; MGI, 2011; OCDE, 2014; OECD, 2013)

The first characteristic that defines Big Data is volume. Such feature refers to the magnitude of the data reported in terabytes (10^{12} bytes or 1000 gigabytes) and petabytes (1000 terabytes). Although these data volumes are already extensive, technological development is still increasing the storage capacity, allowing even bigger datasets to be collected, transferred, and analyzed. For example, Facebook processes up to one million photographs per second, and earlier estimates indicate that it stores up to 260 billion photos using over 20 petabytes (BEAVER ET AL., 2019).

The second 'V', variety is the characteristic of the structural heterogeneity of data. In other words, big data applications employ several types of data, such as structured, semi-structured, and unstructured data. Structured data is the tabular data found in a spreadsheet or relational bases that use a predefined and fixed organizational schema. Structured constitutes only 5% of all existing data (CUKIER, 2010). Data as text, image, audio, and video formats are considered unstructured because they do not have the structural data organization necessary for the analysis. In between structured and unstructured, semi-structured data refers to data with some organization but does not follow the strict standards or a formal structure found in structured data. Typically, semi-structured data contains tags or other forms of markups to separate textual and semantic content (GANDOMI & HAIDER, 2015).

Finally, velocity refers to the rate at which data can be generated, transferred, processed, and analyzed. Devices as smartphones and sensors have led to a remarkable data creation rate and an increasing need for real-time analysis. The faster data can be collected and analyzed, the more commercial value it has. For example, Wal-Mart processes up to one million transactions per hour. This enormous information is used for a real-time and personalized offer for its customers, based on geospatial location, demographics, and personal preferences (CUKIER, 2010).

Other definitions of Big Data still include plus 3 V's: Veracity, Variability (and complexity), and Value. IBM stated veracity's characteristic, which is related to the reliability of the information retrieved and how to create value from inaccurate or uncertain data. SAS introduced the aspect of variability (and complexity) that refers to the variate of data flow rates and the complex myriad of sources that data are collected. Oracle coined the characteristic of value that refers to the low-value density of data. In other words, data in its original form has a relatively low value, but a higher value can be obtained when large volumes are analyzed.

There does not exist a universal benchmark for these characteristics of big data. These attributes' limits depend on the institution's size, location, and sector that appropriate the data.

2.2.2 Big Data Analytics

All the value that can be reached through big data comes from the decision-making based on the analyzed data. Labrinidis & Jagadish (2012) proposes an overall process of extracting useful information and new insights from big data (Figure 15). This process involves two main steps: data management and analytics. In the first one, the aim is to acquire, store, and prepare data for analysis. The analytics step refers to tools and techniques to analyze and retrieve intelligence from big data.

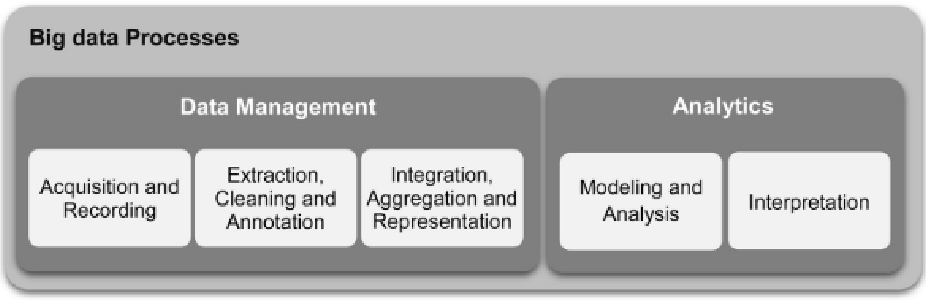


Figure 15 - Big Data Analytics process

Source: Gandomi & Haider (2015)

Big Data Analytics involves several types of analysis: text analytics, audio analytics, video analytics, social media analytics and predictive analysis.

Text Analytics refers to techniques that extract information from any type of textual data like social media, blogs, internet pages, online forums, survey responses, news, academic papers, patents, or any textual data. In general, text analytics support individuals, academics, companies, and governments to create a semantic summary from a large volume of text data, mainly for evidence-based decision-making. Examples of text analytics techniques are information extraction, text summarization, question answering, and sentiment analysis.

Information extraction techniques aim to extract structured data from unstructured text. Information extraction algorithms use two sub-tasks to organize data collected from unstructured text: 1) entity recognition tasks find the names and classify them in predefined categories, and 2) relation extraction tasks find the relationship between entities based on a semantic construction (JIANG, 2013).

Text summarization produces a concise summary of a document or a set of documents. Some text summarization algorithms use Natural Language Processing (NLP) to analyze the location and frequency of words and expressions on the text and create a summary that can be a subset of the main sentences on the original text, summing up the ideas.

Question answering provides answers to natural language questions using text analytics techniques, natural language processing, and machine learning. The healthcare, marketing, finance, and education sectors currently implement these systems in their products and services.

Sentiment analysis, also known as opinion mining, provides an analysis of texts expressing opinions, and it is beneficial for marketing, finance, political and social sciences. Companies are increasingly collecting their customers' opinions about their products and services, making sentiment analysis one of the most active research areas in text analytics (LIU, 2012). Sentiment analysis can be divided into three perspectives: document-level, sentence-level, and aspect-based. In document-level techniques, research is focused on the whole document's perception, if it expresses a positive, negative, or another specific sentiment. At the sentence-level, the analysis focuses on obtaining the position expressed in a single sentence, differentiating subjective from objective sentences. In aspect-based techniques, the algorithm recognizes all opinions and sentiments in a document and identifies which aspects of the entity it refers to. Thus, in-depth information about opinions referring to different features of a specific product, for example, can be analyzed.

Audio analytics techniques are used to capture and analyze unstructured data from audio formats. When applied to spoken language, it also can be called speech analysis. The healthcare sector uses audio analytics to support the diagnosis and treatment of some conditions, such as depression, autism disorders, schizophrenia, and cancer, by analyzing patients' communication patterns (HIRSCHBERG ET AL., 2010). Call centers also analyze a large volume of audio data to provide better customer experience, evaluate agent performance, increase sales, monitor compliance, gain insights from the clients and support Interactive voice response platforms to handle call services. Speech analysis performs in two different approaches: large-vocabulary continuous speech recognition (LVCSR) and phonetic-based. LVCRS, also known as the transcript-based approach, consists of a transcription of the speech content with algorithms that match sounds to words identified in a predefined dictionary. On the other hand, the phonetic-based approach distinguishes phonemes instead of words.

Video Analytics is also known as video content analysis, and it consists of various techniques for capturing, monitoring, and analyzing data from video formats. Compared to other types of analysis, video analysis is still incipient, but researchers developed many new methods in the last years (PRANIGRAHI ET AL., 2010). The large volume of video data is a crucial challenge. Primarily employed in surveillance and security systems, video analytics can detect breaches of restricted zones, identify objects and people, recognize suspected activities, and notify security personnel in real-time.

Social Media Analytics is the analysis of structured and unstructured data from social media platforms. Although the social network theory dates back to the 1920’s, social media analytics got emergent after Web 2.0 as a data-centric approach. Social media is a set of online platforms that allow users to create and share content. They can be social networks (Facebook, Linkedin), blogs (WordPress), microblogs (Twitter), social news (Reddit), media sharing (YouTube, Instagram), wikis (Wikipedia), question-and-answer sites (Yahoo! Answers), and review sites (TripAdvisor) (BARBIER & LIU, 2011; GUNDECHA & LIU, 2012). The relationship among users and the content they create are the sources of information in social media analysis. In content-based analytics, all text or media posted or shared by users provide insights about customers' feedback, reviews, and interests. In structure-based analytics, users' relationship is used to extract intelligence by creating graph models of nodes and edges representing users and their relationships (WASSERMAN & FAUST, 1994). These graphs are used to evaluate and monitor social dynamics, identify communities, evaluate users' or communities' social influence, and predict new relationships among users.

Predictive analysis is a general term for a set of techniques to predict future outcomes based on hidden patterns in a large volume of historical and current data. These techniques are primary statistical methods, regressions or machine learning techniques. Regression techniques consist of understanding the interdependencies between the variables and the outcomes and exploit them for future events prediction. Machine learning techniques are algorithms that can continuously redefine their sets, learning from the historical data, and making accurate projections to the future. Although predictive analysis is gaining importance, the statistical methods applied to big data are still developing to address large datasets. Conventional statistical methods are based on statistical significance. However, this concept does not fit well in big data because these massive datasets can represent the majority or even the entire analyzed population.

2.3 What is Machine Learning

Machine Learning is a central concept in this work. This section will present the historical development, concept and several approaches of this technology.

2.3.1 Natural and Machine Learning

Before starting to talk about Machine Learning, some aspects of the natural learning process can be reviewed to better understand machine learning processes. In general terms, learning is the process of transforming experiences or events into relevant knowledge. Shalev-Shwartz & Ben-David (2014) describe a good example of this process in rats. When rats found a new kind of food, they eat just a little piece, then analyze the flavor and its effects on their body. If the food causes some physiological issue, the rat associates the food with illness, and will not eat it again. The animal uses its previous experience to predict the future outcomes of its actions.

Besides that, rats can go further in memorization and generalize their expertise to other kinds of foods with similar looks and tastes like the bad ones. This process is referred to as inductive reasoning or inductive inference.

However, the pigeon experiment performed by the psychologist B. F. Skinner[2] shows that inductive reasoning could draw false conclusions. Pigeons were placed in a cage, which automatically delivered food to the animals at regular time intervals. The constant food delivery reinforced the activity that pigeons were engaged in at the first moment that food was dispensed. These events mislead pigeons to associate the food delivery with the behave they had for the first time. In this example, the central question is that humans rely on common sense and prior knowledge to understand that some variables are not correlated. This process is called inductive bias.

From the machine learning side, a classic example of inductive reasoning is the e-mail spam-filter task. A machine learning algorithm can memorize previous e-mails classified as spam and reproduce the same set of patterns to type a new e-mail as spam or not.

2.3.2 The concept of Machine Learning

Arthur Samuel (1901-1990) first defined Machine Learning as a "field of study that gives computers the ability to learn without being explicitly programmed" (SAMUEL, 1959). Other complementary definitions came from Simon (1983): "Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more efficiently and more effectively the next time". Mitchell (1997) describes a learning problem: a computer program is said to learn from experience “E” concerning some task “T” and some performance measure “P”, if its performance on “T”, as measured by “P”, improves with experience “E”. The development of a financial fraud detection system may be an example. The system has the task “T” to detect fraudulent transactions, and the measure “P” is the number of frauds detected. The system uses the analysis of previous transactions to get the experience “E”. Each new analysis feeds this experience aiming to achieve the minimum measure “P”. In other words,

---

[2] See at http://psychclassics.yorku.ca/Skinner/Pigeon

Machine Learning explores algorithms that can learn and make predictions based on inputted data.

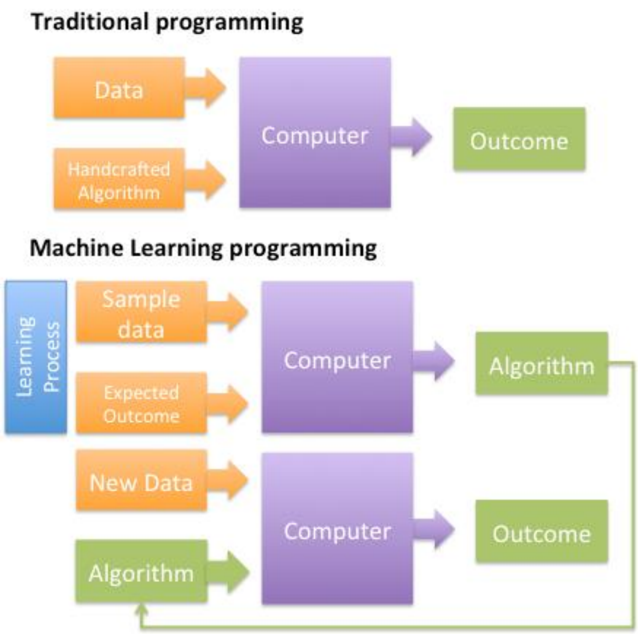


Figure 16 - Traditional programming vs. ML programming.

Source: based on Samuel (1959)

Highly complex problems and situations that need adaptivity require Machine Learning solutions. Everyday human activities like image understanding, speech recognition, and driving a car are too complex to be directly programmed and executed by an algorithm. Also, tasks beyond human capabilities, such as weather prediction and analysis of genomic data benefit from Machine Learning techniques. These tasks require an algorithm that can interact with the environment by learning through data to produce adapted results for the task. Decoding handwritten text is an example of an adaptivity task that the algorithm can adjust to different handwritten users (SHALEV-SHWARTZ & BEN-DAVID, 2014).

IBM Watson is an example of a Machine Learning application. It was initially conceived as a cognitive research experiment to teach a computer through significant volumes of data, like Wikipedia, newspapers, and other information, to provide evidence-driven answers in response to natural language questions. A public demonstration of this tool's potential was performed in 2001 on the quiz general knowledge game show "Jeopardy!" during which Watson defeated two human winners.

IBM developed Watson's new applications to support different industries and business challenges, mainly healthcare, education, and agriculture (CHEN ET AL., 2016).

2.3.3 Supervised and unsupervised learning

Machine Learning algorithms can be categorized according to the type of learning that it proposes. Learning is a wide field of study, but some learning paradigms are recognized as machine learning subdomains.

Supervised learning involves machine-learning algorithms that learn the “correct answer” by mapping inputs and outputs through examples inserted in the system. The learner system receives a set of training data for which the outcome variable is also provided. In this way, the algorithm can learn, find the patterns of data, create an internal rule, and predict a new entry outcome. In the spam-filter example, the algorithm receives a set of training e-mails labeled as spam or not-spam. This dataset provides the necessary information for the algorithm to “understand” which words, type of writing, or sender are associated with spam. Based on the founded patterns, the algorithm adapts its internal rule. In the mentioned example, the algorithm shows a binary classification. From the input x (information about the e-mail), the algorithm predicts the outcome y in two label values (spam or not spam).

However, different problems require different algorithms depending on the desired format of the outcome. If the algorithm product can take not only two but N label, this is called multiclass classification. In multivalue classification, the outcome y is labeled simultaneously in one or more of the N labels. In ranking problems, the outcome y provides an order on some set. Finally, structured prediction problems present the outcome y as an object like a graph that satisfies some internal rule.

Supervised learning algorithms can solve all these problems through tree-based techniques, regression techniques, support vector machines, neural networks, kernel techniques, and Bayesian classifiers.

On the other hand, unsupervised algorithms receive a set of training data with no outcome labels. It aims to look for some structure on sample data, group them on clusters of similar rules, and discover hidden patterns. E-mail anomaly detection task is an example of unsupervised learning. The learning algorithms receive the set of training e-mails with no outcome label, and the task is to detect unusual messages (SHALEV-SHWARTZ & BEN-DAVID, 2014). Dimension reduction methods are used in unsupervised learning to transform data from a high-dimension space to a low-dimension space, which contains the meaningful properties of original data. Clustering methods are used to find a partition of the data and classify new entry data with a predicting rule (JORDAN & MITCHELL, 2015).

The third machine learning paradigm is called reinforcement learning. Here, the training data is intermediate between supervised and unsupervised learning. In this case, instead of providing an output for input, training data indicates a reward or a punishment to the algorithm based on the settled goals. Learning by trial and error using rewards and punishments as feedback, the algorithms find a suitable solution by maximizing the total rewards. (JORDAN & MITCHELL, 2015; SHALEV-SHWARTZ & BEN-DAVID, 2014). Examples of reinforcement learning tasks are applied to games like IBM's Deep Blue computer, which plays chess based on maximizing the best moves.

2.3.4 Machine Learning Techniques

Several methods are currently employed in machine learning tasks. The most recognized techniques are Linear Regression, Logistic Regression, Decision Trees, Support Vector Machines (SVM), Naïve Bayes, Artificial Neural Networks (ANN), and Deep Learning. A brief description of the techniques is provided below.

- Linear and Logistic Regression

One of the most common statistical tools applied to machine learning is linear regression. It models the relationship between explanatory variables and the outcome as a linear function based on the training data. This linear function is the best approximation between the variables, as shows in the Figure 17 below.

Figure 17 - Generic linear regression

Source: Shalev-Shwartz & Ben-David, 2014

Logistic regression is similar to linear regression, except it predicts based on categorical variables using a logistic function to model each class's probability. For working with categorical data, logistic regression is frequently used in classification tasks (HASTIE ET AL., 2009; SHALEV-SHWARTZ & BEN-DAVID, 2014).

- Decision Trees:

Decision Tree is a machine learning predict model based on sequential splitting data depending on their features. A predefined set of splitting rules is applied at each node to identify regions with the most homogeneous features. At the end of the splitting edges (or branches), a label is provided (leaf). A classification tree provides a specific class at each leaf, while a regression tree provides the mean response for the observation in that split region (SHALEV-SHWARTZ & BEN-DAVID, 2014). The Figure 18 below shows an example of a decision tree partition at a two-dimensional feature space (X1 and X2). The output labels (Y1, Y2, Y3, Y4, Y5) are in the terminal nodes, and t1, t2, t3, and t4 are the split points (HASTIE ET AL., 2009). Decision trees have become popular because they can represent information intuitively and efficiently, deal with any variable type (numerical or categorical or binary) and require little data preparation effort. (ELITH ET AL., 2008).

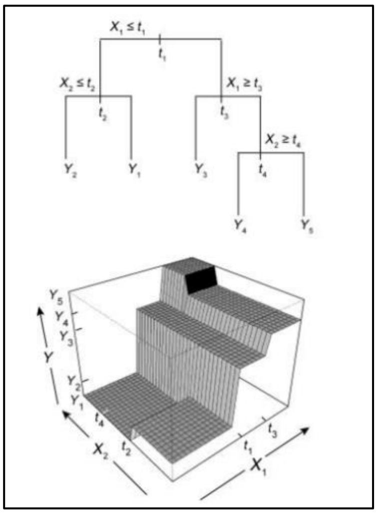


Figure 18 - Generic decision tree graph

Source: Elith et al. (2008)

- Support Vector Machine (SVM):

Support Vector Machine (SVM) is a supervised linear binary-classification algorithm. The algorithm creates a "large margin" separator between each class's elements from a set of training data. In other words, it models two separate hyperplanes that can classify new entry elements based on the greater distance to the nearest element of both classes. The larger is the margin between the closest elements of different classes, the lower is the algorithm error. Figure 19 shows an example of the elements from different categories separated by a linear separator. (SHALEV-SHWARTZ & BEN-DAVID, 2014; LIU ET AL., 2018)

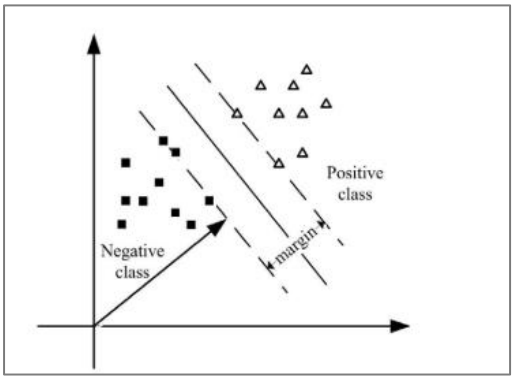


Figure 19 - Generic support vector machine application

Source: Liu et al. (2018)

- Naïve Bayes

Naïve Bayes is a probabilistic classifier method frequently used in text categorizing based on the words' frequency. It is naïve because it ignores the relationship between the features, considering only the probability of the features' presence. The algorithm learns from the training data the probability of each attribute on each class. Then it applies the Bayes rule to calculate the probability of each class, given new entry data. The classification is done by predicting the class with a higher likelihood (FRIEDMAN ET AL., 1997).

- Artificial Neural Networks (ANN)

Artificial Neural Networks or Neural Networks are machine learning computational models inspired by biological neural networks' structure. It consists of a set of basic computational devices (neurons or nodes) connected in complex ways. Each node receives input as a weighted sum of the direct connected neurons' outputs. The neuron's output is calculated through some non-linear function based on the input information. The connection edges typically have a weight that is adjusted as learning advances (SHALEV-SHWARTZ & BEN-DAVID, 2014).

Some different approaches have been developed in the neural network field. Feed-forward networks have modules aggreged in layers, and the signal runs in just one direction, from the input to the output. Convolutional Neural Networks (CNN) are feed-forward networks developed to analyze data in multiple arrays format, such as a colored image built as 2D arrays with pixel intensities for the three-color channels. CNN is characterized by the modules' local connection, shared weights, pooling, and multiple layers. It is mostly used for image recognition and video processing (LECUN ET AL., 2015).

On the other hand, in recurrent networks, some modules' output can feed modules in the same or former layers, including the module itself. Recurrent Neural Networks (RNN) are most used when it involves sequential inputs as speech recognition. RNNs can process an input sequence of one element simultaneously and keep in their hidden modules' information about the past elements' history (JORDAN & MITCHELL, 2015; LECUN ET AL., 2015).

- Deep Learning

Conventional machine learning techniques have some limitations in processing raw data into a representation that the learning system could find, recognize, and classify the input patterns. Deep learning methods are based on neural network architecture, transforming raw data by non-linear functions into multiple representation layers. For each layer, complex functions are learned to amplify some essential aspects of the input and suppress the irrelevant ones. Image recognizing algorithms are examples of deep learning applications. Generally, given the input as an array of pixel values, the first layer detects the edges and their location on the image. The motifs, represented by a particular arrangement of edges or borders, are detected on the second layer. The third and subsequent layers detect objects as a combination of the motifs seen before to get full image recognition as an outcome. (LECUN ET AL., 2015).

Deep learning is composed of a set of modules in each layer, called neural networks. In Figure 20, it is shown a neural network with two hidden layers and an outcome layer. The total input z is calculated as the weighted sum of the former layer's outputs in each layer. Then, to get the result of the layer, a non-linear function is applied to z. This step is continuously used until the last layer to get the algorithm output. The second neural network in Figure 20 shows the backward pass of the algorithm. It calculates each layer's derivative error and uses these errors to adjust the modules' weight for more authentic learning.

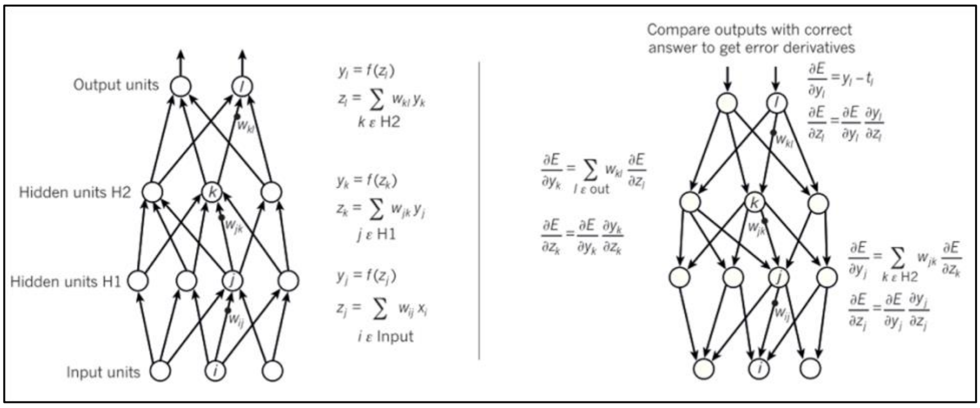


Figure 20 - Generic neural network

Source: Lecun et al. (2015)

Deep learning methods are useful in finding patterns in high-dimensional data, and it is used in several science fields. Beyond image recognition (TOMPSON ET AL., 2014), speech recognition (SAINATH ET AL., 2013), drug molecules activities (MA ET AL., 2015), genetics (XIONG ET AL., 2015), molecular physics, and brain-activity model (HELMSTAEDTER ET AL., 2013) are examples of deep learning algorithm applications.

2.4 Applications and Limitations of Big Data and Machine Learning

Big data and Machine Learning are applied in many technologies, supporting several sectors. Although there is an excellent expectation regarding such technologies' benefits, there are still significant barriers and bottlenecks. This section presents a brief description of the most relevant technologies based on Big Data or Machine Learning and some examples of industrial applications. The last subsections discuss the limitations.

2.4.1 Artificial Intelligence

Artificial Intelligence (AI) is an umbrella concept for many subfields, such as augmented intelligence, cognitive computing, AI robotics, and machine learning. It is broadly accepted that AI aims to automate cognitive intelligence tasks such as gathering information, planning, communicating, reasoning, and making decisions

(MULLER, 2017). In general, AI consists of intelligent systems that independently can execute tasks in complex environments and even improve their effectiveness by learning (BELKOM, 2020).

The concept of AI was first employed in 1956 in an academic workshop at Dartmouth College in Hanover, New Hampshire (MCCARTHY ET AL., 2006). Top computer scientists carried out the first artificial intelligence study. However, experts and academics could not set an objective definition of the AI concept, and the technology ran through two technological paths.

The first group of experts supported a top-down, rule- and logic-based approach for Artificial Intelligence. This approach, also called Symbolic AI, designs artificial intelligence to follow step-by-step or "if-then" processes to analyze the world's representations. Expert Systems (DURKIN, 1996) emerged from this knowledge-based research domain.

The second group of experts supported a data- and statistics-based approach, also called Subsymbolic AI. Subsymbolic refers that symbols and rules did not endorse AI, but learning from the environment. Subsymbolic AI has made Machine Learning emerges as a research domain, and it is, nowadays, the vast majority of AI research and applications (HAFEZI, 2020; BELKOM, 2020).

Hafezi (2020) presents three interdependent elements that support AI in recent years: advances in algorithms, big data, and computer capabilities. The first element refers to advances in machine learning techniques such as artificial neural networks (ANN) and deep learning. The second element, big data, emerged within the internet's digitalization and allowed opportunities for discovering new data-based insights. According to Hafezi (2020), examples of data sources are companies (production, inventory, customer, and financial data), the Internet of Things (data from sensors, network transmissions, and IoT-based applications), and bio-medical (clinical, medical, and pharmaceutical data). The third element refers to computing capabilities in power and storage and cloud computing. The interdependence among these elements is seen in several moments. Learning models could efficiently improve due to the explosion of available data for training and testing the algorithms. Big data was achieved after the evolution in computing capacity in storage and processing. Finally, the increase in computing processing power is driven by cloud computing, requiring better and faster hardware.

AI can be classified into three main categories based on different development levels: Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGI), and Artificial Super Intelligence (ASI).

Artificial Narrow Intelligence is a form of AI specializing in one or a few specific tasks such as playing chess, making recommendations, and recognizing speech or images. In other words, ANI consists of intelligent systems that cannot perform any task beyond its domains.

Artificial General Intelligence (AGI) is a human-level AI. It can recognize patterns, solve problems, and adapt itself to new environments. AGI is still beyond the current AI developments, and there is no agreement among AI experts about how far AI developments are far from AGI (FORD, 2018).

Artificial Super Intelligence (ASI) is intelligent system that surpass human intelligence in many domains, such as science, creativity, and social behavior. Some experts say that since AI development achieve AGI, it could cause an intelligence explosion, and ASI will become feasible (BOSTROM, 2016; BELKOM, 2020).

2.4.2 Fake news and Deep fakes

The problem of fake news has become very known since the 2016 US presidential election, in which the creation and spread of information that attacked or favored any of the candidates reached high volumes. In general words, fake news consists of disinformation, misinformation, and the human bias inherent in information produced by humans (LAZER ET AL., 2017; TORABI & TABOADA, 2019).

Jack (2017) states that disinformation is deliberately false or misleading information. In some cases, coordinated disinformation campaigns spread specific fake news, mainly through social networks such as Facebook, Twitter, and YouTube. On the other hand, misinformation refers to the spread of unintentionally inaccurate information. This situation can happen when the news is reported in error or when journalists misinterpret or do not verify their sources.

Disinformation and misinformation are not recent. However, the present-day fake news is much more distressing due to the speed with which it can spread. Via social media, fake news has the power to escalate and reach even more people than real information could do (VOSOUGHI ET AL., 2018).

Deep fake is a more robust type of fake news due to the employment of deep learning techniques. Deep learning's relevance in creating fake news is because the result of a doctored image, audio, or video designed by deep learning is much more realistic and challenging to detect than traditional counterfeit media. The field of digital forensics deals with the issue of automatically recognize doctored; however, deep fakes detection is still a challenge (CHESNEY & CITRON, 2018).

Deep learning algorithms use a set of training data to mimic someone's characteristics, such as the face, voice, or both. Through the training data, the algorithm can recognize the patterns and reproduce these patterns in different situations. Recent situations involving deep fakes are celebrity faces inserted in porn videos and the FakeApp[3], a photo-video editing application that disseminate the concept of deep fakes (ELLIS, 2018).

Chesney and Citron (2019) divide the harmful uses of deep learning into two categories: damage to individuals and organizations; and damage to society. Many kinds of deep fakes can harm people or organizations on several levels. Some can cause a simple irritation or embarrassment, while others can be much more violent and humiliating.

When focused on individuals or organizations, deep fakes can cause harm by exploitation or sabotage. In this case, exploitation refers to stealing people's identities for financial extort or other advantages. Blackmailers can use available data online from the victims to train deep fake algorithms and recreate high-quality photos, audios, and videos of the victim in disturbing situations. Victims could be forced to provide money, secrets, nude images, or sexual videos (also called "sextortion") to prevent the release of these fake media. When deep fakes aim to inflict psychological harm on people, it is classified as reputational sabotage. Even when the victims can debunk counterfeit media, it is often too late to remedy the initial damage due to the fast spread of fake information via social media (VOSOUGHI ET AL., 2018), and the boost of recommendation algorithms (CHESNEY & CITRON, 2018).

However, societal harms of deep fake could have an immense impact on social life. From the point of view of societal damage, deep fakes can affect public officials, politicians, or army officials, displaying disturbing action, speech, or photo-situation to attack the affiliated public institutions. The consequences to societal

[3] 38. FAKEAPP, https://www.fakeapp.org (last visited May 8, 2018).

stability are several, such as distortion of democratic discourse, manipulation of elections, destroying trust in public institutions, aggravate social divisions, and undermining public safety, diplomacy, national security, and journalism (CHESNEY & CITRON, 2018)

Many approaches have been developed to combat the spread and minimize the negative impacts of fake news and deep fakes. Lazer et al. (2018) propose two guidelines: empower individuals to recognize potential fake news and structural changes to reduce fake news exposition. Torabi & Taboada (2019) go even further on the possible actions to deal with fake news: (1) education to the people; (2) limit the spread of fake information; (3) manual check of fake news; (4) automatic check and classification of fake news.

Education towards media literacy, responsible citizenship, and civil and democratic values are essential to understand political influences and hidden interests under fake news. Some organizations are focused on encouraging people to critically check news sources and provide tools to verify the information[4].

Fake news spread fast due to its novelty (even if it is not true), its capacity to generate online attention, and its characteristic of confirming preexisting reader bias. The social bubble in social media makes it easier to present just one point of view of a story and reaffirm that point to all participants. Therefore, Lazer et al. (2017) propose frequently online communication among different ideological lines to avoid fake news reaching more people and becoming entrenched in their minds.

Manual fake news checking can be done in two ways: using a fact-checker website or manual checking by social media administrators. Fact-checking websites become very known for verifying news that users submit. Qualified journalists and other professionals can research and verify the information under these websites. Large tech companies have responded to social pressure and hired more content moderators to verify and respond to fake news reports, although the argument that the moderator's bias can be propagated in this case (CHEN, 2017).

Machine learning techniques are employed to create deep fake and also to detect misleading information. Natural Language Processing algorithms can automatically detect and classify fake news. Machine learning methods with this

[4] See the initiatives at the International Federation of Library Association (https://www.ifla.org/publications/node/11174) and the Shorenstein Center on Media, Politics and Public Policy of Harvard Kennedy School (https://shorensteincenter.org/free-course-identifying-misinformation/)

purpose can vary from probabilistic techniques (such as Naïve Bayes) to complex deep learning methods. Processing natural language using a probabilistic feature-based approach is a powerful model to identify and classify fake news. Correlations among features, such as n-grams[5], subjectivity and polarity markers, lexical-semantic classes, syntactic and discourse-level, are achieved via labeled training data. After recognizing patterns related to fake news, the algorithm can complete the task. This approach has been shown a limited performance gain as the volume of training data increases. Consequently, complex techniques, such as deep learning are getting more attention (TORABI & TABOADA, 2019).

Recurrent Neural Networks (RNN) can encode sequential information and be successfully applied for modeling short text semantics. Convolutional Neural Networks (CNN) can provide an abstraction of the input, and they are employed, in general, for more extended text classifications.

Most of automatic approaches for detecting fake news are data-based, and lots of training data are required to achieve good performance. This task's training data consists of diverse, balanced, and precisely labeled fake and real news articles. (LAZER ET AL., 2017; TORABI & TABOADA, 2019).

2.4.3 Limitations of the technologies

Big Data's limitations can be translated into four main points: low data quality, data manipulation, data interpretation and ethical issues (MITTELSTADT & FLORIDI, 2016; REIMSBACH-KOUNATZE, 2015).

- Low data quality in Big Data

Data quality is related to the relevance of the data for the application it proposes, the accuracy in data measurement, the credibility of the data source, the deadlines for accessing data, and the consistency of data (OECD, 2011). Inconsistent data may cause an increase in data cleaning and merging efforts (HAGEN ET AL., 2019). Also, biased data can create false causality among the parameters, compromising big data analysis. Harford (2014) mentions the example of 1936 USA elections, in which the stronger candidates were the Republican Alfred Landon and President Franklin Roosevelt. The magazine Literary Digest conducted a postal

[5] N-gram is a concept in Natural Language Processing, and it means a continuous sequence of n items in a text. These items can be phonemes, syllables, letters or words.

opinion poll to forecast the elections' result by assessing 10 million Americans, including 2,4 million who answered the call. The Literary Digest announced that Landon would win by 55%, but the American voters elected Roosevelt by 61%. Nevertheless, a small survey conducted with 3000 participants by the opinion poll pioneer George Gallup came closer to the real results than the Literary Digest. Gallup's results can be explained by the sample bias, which was much more intense in 2,4 million than 3000 participants. Sample error means the chance that a sample does not reflect the underlying population. Therefore, this example shows that an unbiased sample is more important than a large sample in opinion polls (HARFORD, 2014). In other words, not always lots of data means better forecast results. Adnan & Akbar (2019) and White & Breckenridge (2014) state that data sparsity, dimensionality, and diversity are very relevant elements in data quality issues.

- Data manipulation in Big Data

Data manipulation and biased algorithms are also cited as an ethical issue, even if it is intended or not. L'Heureux et al. (2017) point three moments of possible manipulations in the data analytics pipeline (Figure 21). Dimensionality reduction, instance selection, and data cleaning are three critical aspects of data manipulation. Dimensionality reductions are a process to map high-dimensionality spaces into low-dimensionality spaces without the loss of significant information using Principal Component Analysis (PCA) or several other techniques. Instance selection is composed of methods to select a better sub-dataset that can represent the whole population. Data cleaning, also referred to as pre-processing, is a process to remove and correct data noise, outliers, and data inconsistency. These processes may be responsible for biased future decisions.

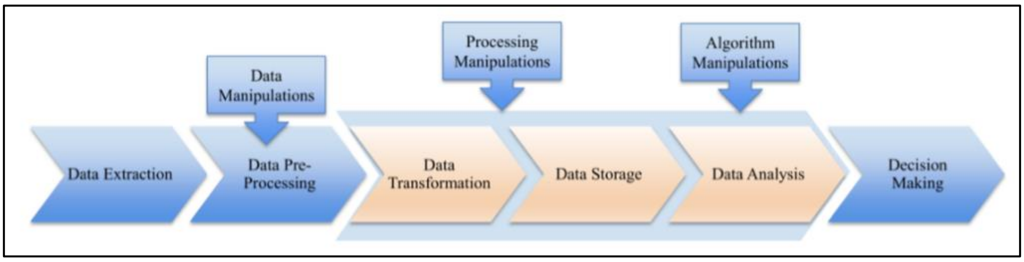


Figure 21 - Data manipulation in data pipeline

Source: L'Heureux et al. (2017)

- Data interpretability limitations in Big Data

The inappropriate use of data refers to the fact that some decisions can be made only by looking at the presented data analyses without fully understanding the whole context. Then, inaccurate analyzes and correlations that have no causal relation may result in misplaced decisions. Hagen et al. (2019) discuss the importance of contextual information by analyzing 311 citizens' requests performed by the researchers in their paper. They highlight the importance of the knowledge about regulative, residential, and political environments in analyzing the patterns of requests in Miami-Dade County to suggest better public policies. In other words, the changes that occur in the environment in which the data are collected can influence the result and are not addressed correctly in analyzing the data.

- Ethics and Privacy in Big Data

In the last years, ethical concerns have emerged as a critical issue in Big Data and Machine Learning field. Mittelstadt & Floridi (2016) defines five areas in this context: informed consent, privacy, ownership, epistemology, and big data divide.

Historically, informed consent is taken for only a single study, which may be a problem for big data. For example, Facebook carried out several experiments in manipulating users' news feeds to analyze how it affects the content users share (STAHL & WRIGHT, 2018). Flick (2016) criticizes Facebook's study due to the lack of ethical oversight and the negligence to obtain informed consent from the participants. Big data analytics is based on finding patterns among data from several sources, and, in general, it is not possible to predict the connections found in data before the research. As a result, the consent cannot be informed in terms of data subjectivity (future uses or consequences of this data) that are unknown when information is collected. From that point of view, the regulations around ethical issues related to informed consent can be seen as a barrier to implementing these technologies.

One of the most prominent ethical issues in Big Data is privacy and data protection (JORDAN & MITCHELL, 2015; STAHL & WRIGHT, 2018). Privacy in Big Data and Machine Learning are very frequently related to anonymization and confidentiality. Some authors (MARKOWETZ ET AL., 2014; MOORE ET AL., 2013; SHILTON, 2012) associate privacy to Big Data analysis invasiveness, particularly when it connects information from combined datasets, geolocation, and internet-based sources. Privacy norms are very pertinent for the new data generation sources such as social media. Data can be exposed and analyzed outside of the context in which it was created and violate data privacy (BOYD & CRAWFORD, 2012).

The concept of ownership refers to the rights regarding distribution, modification, and advantages of intellectual property or innovation from data analysis. Databases employ different forms of ownership, including two fundamental forms: the right to control the data, and the right to benefit from the data. Data control empowers data subjects to track and check their data manipulations to prevent "secret databases" and unacceptable uses of data (TENE & POLONETSKY, 2013). This control is even more relevant when there is the possibility of reidentification or hidden analysis (CHOUDHURY ET AL., 2014). Ownership in terms of benefit from the data requires that data custodians offer data subjects the rights to access their data in a suitable and machine-readable format for personal uses (TENE & POLONETSKY, 2013). The risks in both ownership approaches include the misinterpretation of the data without expert analysis, the relative uselessness of raw data and mistakes, and inaccurate modification by data subjects (MITTELSTADT & FLORIDI, 2016) .

Researchers reveal a connection between ethics and epistemology of Big Data and Machine Learning, based on algorithms and how humans fail to understand it, transforming them into black boxes. This situation is not a new problem, exemplified as patients who are unable to interpret their radiographs. However, the complexity added by big data makes understanding the algorithms more complicated (CALLEBAUT, 2012). As a result, questioning big data analytics' findings becomes more difficult, even by experts.

The digital divide refers to the unequal distribution of access and use of digital data in terms of skills, knowledge, and opportunity to access and use. A few large organizations are responsible for collecting and storing data. This situation deprives individuals and small organizations of accessing the logical methods that decision criteria are used in big data analysis and decision making (PUSCHMANN & BURGESS, 2014; TENE & POLONETSKY, 2013). Data for profiling and monitoring users can also be problematic from the point of view that deciding on the categories that a data fits into is unclear. Comprehension of why and how algorithms categorized information is the key to strengthening users' self-control over their data (MITTELSTADT & FLORIDI, 2016; TENE & POLONETSKY, 2013).

Stahl & Wright (2018) defend that every student learning AI, computer science, or data science should discuss ethics and security during their training to avoid ethical issues in performing data analysis.

2.5 Big Data and Machine Learning tools for futures studies

Several frameworks are employed to support data-driven futures studies processes. Mainly of these, focus on quantitative analysis. This section presents the more frequently used techniques within Big Data and Machine Learning for futures studies support, such as statistics, text mining, Big Data Analytics, neural networks, Deep Learning, cluster analysis and Social Network Analysis. The last subsection present examples of the methodological integration of BDML tools and foresight methodologies.

2.5.1 Statistics and futures studies

Most of new quantitative techniques in foresight use statistics as a way to analyze and understand data. Unlike forecasts, strategic foresight uses statistical methods to get a deep quantitative understanding of future trends provided by analytical data. Strategic foresight merges analytical data with qualitative analysis and expert's opinions to conduct a participatory process as valuable as the results. Statistics are used within algorithms analysis such as data mining, text mining, neural networks, and deep learning (HAFEZI ET AL., 2019). Patent analysis (UHM ET AL., 2017), time series analysis (JUN & PARK, 2013), and trend analysis (KIM & JU, 2019) are traditional methods in foresight that use statistical models. However, foresight methods can also help statistical approaches to deal with an increasingly turbulent context and offer different challenges and analysis opportunities. Kastrinos (2018) states that foresight could guide statistical offices to anticipate political and technological changes and judge the relevance of indicators they use for a phenomenon. The author analyses the implications of scenarios developed by the BOHEMIA[6] study for European statistical offices.

[6] See more at https://ec.europa.eu/info/research-and-innovation/strategy/support-policy-making/support-eu-research-and-innovation-policy-making/foresight/activities/current/bohemia_en

2.5.2 Text mining and Big Data Analytics

Several authors include Big Data Analytics in their foresight methodologies, especially when dealing with massive, structured, or unstructured data. Several techniques are used, such as data mining, text mining, clustering, and machine learning algorithms. Krittanawong & Kukin (2018) present an example of Big Data Analytics. The authors collected insights and defined patterns to better respond to treatment modalities in heart failure. The results supported an improvement suggestion in treatment specificity and efficacy through analyzing a massive collection of clinical studies data.

The relevant data for future-oriented activities can be found in several documents, such as scientific publications, patents, news articles, standards, social media, and websites.

Scientific and patent publications are used for foresight methods to examine the landscape of technology developments. In general, data is extracted from quality-assured databases such as Web Of Science and Scopus. In publication analysis, text mining can analyze the title, abstract, keywords, and other data fields, including even full texts. In patent analysis, text mining is often used to obtain information from unstructured text such as patent abstracts, claims, or descriptions. Patent infringement detection, monitoring R&D developments, technology transfer, technology planning, and patent classifications are current text mining applications in patent analysis.

News articles are sources of information about public concerns, beliefs, and reservations (ALBERT ET AL., 2015; AMANATIDOU ET AL., 2012; GLASSEY, 2012; PANG, 2010). Text mining is used in unstructured news texts for constructing analysis models and interpret the information by processing natural language and keyword matching. Yoon (2012) employ text mining to detect weak signals by analyzing web news in the field of solar cells. The author collected 28270 English web news articles from 2006 to 2011, defined the keywords, constructed a keyword emergence map, and identified weak signal topics by statistical analysis. He proposes his text mining method to be expanded to collect information from other web sources such as blogs and websites. Although systematic and automated text mining analysis of news articles can provide essential society-related insights for foresight study, its application is still rare (KAYSER & BLIND, 2017).

For societal discourse data, social media such as YouTube, Facebook, or Twitter are an essential source. They can provide insights by analyzing sentiment trends and social behavior. Many authors (ALBERT ET AL., 2015; AMANATIDOU ET AL., 2012; GLASSEY, 2012; PANG, 2010) applied this technique to gather information from several web channels to obtain technology-maturity models and trends detection. Text mining tasks can be used for several applications and combined with several foresight methods such as scenarios and technology roadmapping.

Websites are generally classified as semi-structured data and can be analyzed by text mining techniques. Innovation reports included in companies' websites provide handy insight for innovation indicators used in foresight activities (KAYSER & BLIND, 2017). An example of the use of websites as a source of future-relevant data is presented by Youtie et al. (2012). Data collected from small and medium-size companies' websites, based technology transition insights for nanotechnology field. However, general retrieval approaches to topic-related websites in text mining are not frequently used (KAYSER & BLIND, 2017).

2.5.3 Machine Learning in futures studies

Neural Networks, Cluster Analysis, and Deep Learning can support future-oriented activities by forecasting and evaluating new technologies. Trappey et al., (2019) offer a methodological approach for patent valuation, analyzing 6466 manufacturing IoT patents through principal component analysis (PCA) and deep neural networks. Patent valuation prediction based on PCA and Deep Neural Network approaches has shown improved accuracy compared to prediction based on the use of Back Propagation Neural Network for IoT patents.. Although the authors apply their framework systematically for IoT patents, it can also be applied in other fields to understand the value dynamics by a valuation process. Zhen & Yao (2020) use PCA, Decision Trees, and Deep Learning to evaluate companies' technological innovation capability and predict innovation indexes. The authors' conclusions point to a significant predictive power of the applied method related to firms' technological innovation capabilities and propose a framework to guide regional economic development. Zhou et al. (2020) forecast emerging technologies using limited patent data through Decision Trees, Deep Learning with data augmentation. Lee et al. (2017) predict the status of the pro-environmental consumption index of 13 countries based on Google search query data and World Bank indicators. The authors analyze the dataset applying recurrent neural networks with one, nine, and a hundred hidden layers.

Compared to several other techniques, the deep learning model has achieved the most accurate value to the actual pro-environmental consumption index. Thus, the authors state that deep learning technology and Google trend data are useful for environmental studies.

2.5.4 Deep Learning in Bibliometrics

Many studies explore the association of deep learning techniques with bibliometrics for foresight or forecast. Li et al. (2018) classify patents through a classification algorithm based on deep learning that overperformed all other patent classification algorithms using the same training data. Hassan et al. (2018) use deep learning to classify citation's importance based on features of the full-text article. This model presented a superior precision (92.57% accuracy) than other classifiers as decision trees (89% accuracy). Zhang et al. (2018) applied deep learning in natural language processing, aiming for topic extraction. The authors collected 4770 articles from Web Of Science and 557 academic proposals granted by the National Science Foundation (NSF) to train a deep learning algorithm to discover latent semantic structures in large-scale text.

2.5.5 Cluster Analysis

Cluster analysis is a technique that supports many different analyses in futures studies, such as patent (LEE ET AL., 2014) and topic analysis (ZHANG ET AL., 2016). Lee et al. (2014) depict a novel method for technology forecasting based on patent analysis. The authors collect patent data for quantitative analysis. They perform a principal component analysis (PCA) through a patent-keyword matrix to cluster these patents in subgroups. The final step is forecasting emerging technologies within the different clusters. The authors conclude that using various patent data types to predict emerging technology is more efficient than using only the number of patents in a technology cluster. Zhang et al. (2016) propose an analytic model for clustering terms in a specific technological topic and identify topical emphases changes. This topic analysis supports a Technology Roadmap by identifying emerging issues to compose the exercise.

2.5.6 Social Network Analysis (SNA)

In general, Social Network Analysis (SNA) is associated with different techniques to compose a set of tools to support foresight methodologies. SNA offers network metrics and visualization of authors, institutions, countries, and keywords to uncover emergent thematic or emergent research-related groups. SNA is often applied to bibliometric data of publications or patents; however, some research show the use of SNA to investigate the relationships between keywords from web blogs and on-line news. Kim & Ju (2019) analyze South Korean media data to look for statistical patterns in 13 novel industrial technologies in 2015. The authors extracted key concepts from on-line news and web blogs. They merged semantic networks of the 13 technologies (IoT, electric vehicle, autonomous vehicle, fintech, drone, artificial intelligence, robot, battery, healthcare, wearable, 3D printing, virtual reality, and big data) to forecast technological progress. The results point to more biased information in mass media than aggregated blogs; however, after analyzing the acceleration and distribution of term frequency over time and the centrality degree of critical concepts, the authors considered both sources in this research. The study shows that artificial intelligence, including robot technologies, is crucial to integrate technological competence. Jun & Park (2013) employ Apple's patent data to examine technological innovation via quantitative analysis (text mining, time series analysis, linear regression, and cluster analysis) and keyword networks (SNA). Their findings consist of the identification of technological trends and technological gaps in central technologies for Apple. Dotsika & Watkins (2017) identify disruptive technologies by analyzing bibliometric patterns from business and academic publication data on seven disruptive technologies (3D printing, big data, Bitcoin, Cloud, IoT, Social Media, and Massive Online Open Courses – MOOC). To calculate networks' characteristics, they used several network metrics, such as number of nodes, edges, density (ratio of the number of edges to the total possible edges), diameter (longest path among two nodes), average degree (average number of edges connected to a node) and clustering coefficient (ratio of connected neighbors' node to the total of these possible links).

To identify the potential influence of a node, the authors also applied centrality metrics such as degree centrality (based on the number of edges in a node), eigenvector centrality (based on the sum of a node edges weighted by the degree centrality of the connected node), closeness centrality (average short distance among the node and all the node in the graph), betweenness centrality (based on the number of times a node act as a bridge along the shortest path among two nodes) and eccentricity (maximum short distance among the node and any other node). As a result, the authors propose a new literature-driven method for forecasting emerging technologies by applying social network analysis.

2.5.7 Integration with foresight methodologies

Some authors explore the integration between Big Data and Machine Learning techniques combined with the foresight methodologies. Kayser and Blind (2017) employ text mining techniques into a roadmapping process to balance internal views and external trends. In general, roadmapping is used for internal strategy, and text mining outputs may provide an overview based on external data. Thus, text mining supports each step of roadmap activity with specific analysis. In step 1 in the Figure 22, text mining supports exploration and identification of relevant terms and missing aspects, helping to define the activity scope. Text mining supports the chronological order of market and technology developments by analyzing trends in the second step. In step 3, text mining supports the links among the roadmap objects by a network and association analysis. In the last step, text mining supports the validation of the final roadmap results. Kayser et al. (2014) reinforces the customer perspective and provide a framework for integrating text mining techniques in the roadmapping process with this approach.

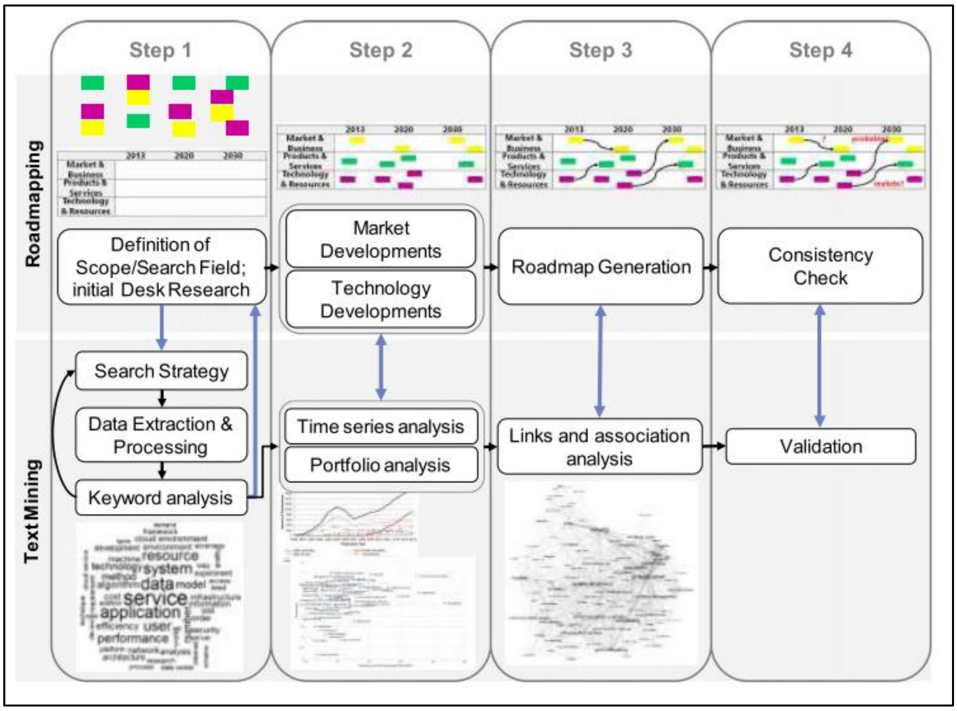


Figure 22 - Roadmapping supported by text mining

Source: Kayser et al. (2014) and Kayser & Blind (2017)

Different from roadmapping, the scenario method supposes that several different future scenarios are formulated to think about challenges and developments that can influence current decision-making. Automatic desk research or other text mining techniques applied to scenarios are rarely used. However, it can offer a broader understanding of the technological e social context in different scenarios when associated with literature analysis. In the Figure 23, Kayser & Shala (2020) present a framework for combining text mining techniques and scenario method.

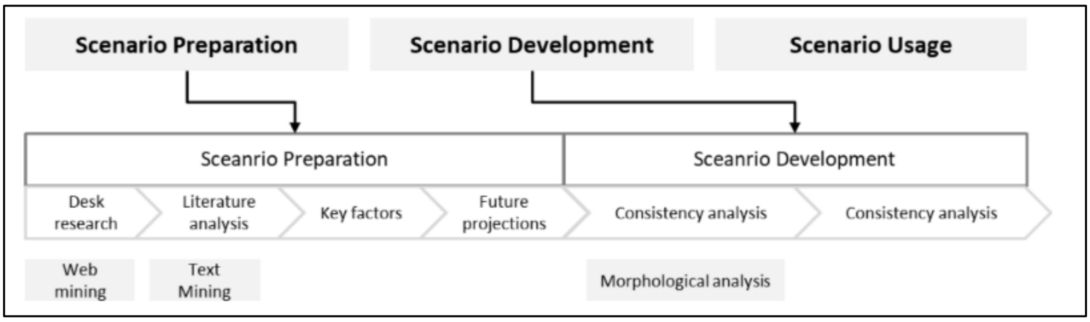


Figure 23 - Scenario supported by text mining

Source: Kayser & Shala (2020)

The authors used Twitter to delimit the thematic field and a web mining technique to collect data from the websites mentioned by the tweets. The text and web mining results associated with literature analysis provide a comprehensive overview of the topic and facilitate the scenarios' development. The steps of text mining in this research consist of retrieve data from Twitter via API, analyze the hashtags and their connection to each other, web mining information from mentioned websites, aggregate the content, develop a concept map based on the data retrieved, develop topic modeling, define influence factors, create future projections and create different scenario stories by morphological analysis (KAYSER & SHALA, 2020)

2.6 Conclusion

The second chapter's objective was to present the second part of the theoretical framework that supports this work. This chapter focused on the digital technologies that are the object of this work: Big Data and Machine Learning. Through a literature review, it was possible to understand the origins, concepts, and techniques related to Big Data and Machine Learning and their direct impact in several sectors.

The set of digital technologies that include Cloud Computing, Internet of Things, and Cyber-physical Systems make up the technological basis of Industry 4.0. Such technologies allow the integration of complex physical machines with the power of sensing and monitoring software. This technical composition is changing production patterns, creating wealth, and directly impacting society.

One of the concepts that underpin Industry 4.0 technologies is Big Data. Among the definitions of Big Data, the consensus among researchers is the characteristics of a large volume of data, a variety of data structure, and velocity of data generation, collection, and analysis (3 Vs). From the point of view of data analysis, Big Data Analytics techniques allow processing and analyzing structured, unstructured data, text, audio, video, and data from social networks.

Based on natural learning processes, Machine Learning algorithms stand out for their ability to understand nonlinear data patterns and predict this data's future behavior. Among the learning techniques, supervised and unsupervised learning stands out. In the first, the algorithm is trained from input and output training data. In unsupervised learning, the algorithm is trained only with input training data. Different techniques such as linear regressions, logistic regressions, decision trees, Support

Vector Machines, Naive Bayes, neural networks, and deep learning are used to search for better accuracy of algorithms' answers.

Despite the use of Big Data and Machine Learning tools and techniques in several areas, including the Artificial Intelligence field, the technologies still have significant limitations. The automatic creation and rapid dissemination of disinformation, fake news, and deep fakes generate substantial social impacts and deepen ethical discussions related to massive data. More specifically about the data, some authors point out the quality, manipulation, interpretability, and privacy as significant issues of these technologies.

The use of Big Data and Machine Learning tools in futures studies was explored in the last section of this chapter. In this chapter, the most used techniques were described, as it will be possible to verify through the results of the bibliometric analysis, later in the chapter 3 of this thesis. The most used technique applied in futures studies are statistics, text mining, social network analysis, and several machine learning techniques, such as deep learning and cluster analysis. The methodological integration of these tools in futures studies is explored by some authors, applying text analytics techniques to roadmapping and scenarios.

Thus, this chapter highlights the technological developments involving Big Data and Machine Learning and its direct application in futures studies. The next chapters delve deeper into understanding the dynamics of using BDML tools in futures studies and the future impact that these tools bring methodologically and conceptually.

PART 2 - THE FRONTIERS OF BIG DATA AND MACHINE LEARNING IN FUTURES STUDIES

The next two chapters of the thesis are dedicated to performing practical experiments and analyses regarding the use of Big Data and Machine Learning (BDML) tools for supporting foresight activities. This part is composed of two chapters. Chapter 3 intends to assess the panorama of futures studies practices in the last decade through a bibliometric analysis. Chapter 4 describes a survey analysis performed with foresight experts and practitioners about BDML tools' impacts in futures studies. The main objective of part 2 is to provide analytical information about the present and future perspectives of the adoption of BDML tools in futures studies.

Chapter 3 - Panorama of Futures Studies Supported by Big Data and Machine Learning

This chapter's main intention is to obtain an overview of the futures studies that employ Big Data and Machine Learning (BDML) tools or techniques in the past decades. The goal is to understand the dynamics of the use of different foresight methodologies, newly developed approaches, and established research networks around this topic. To achieve this goal, this chapter performs a bibliometric and social network analyses of publications regarding futures studies supported by Big Data and Machine Learning tools.

Thus, it is possible to obtain a panorama of futures studies that have been developed in the world recently. Regarding the methodology, the bibliometric analysis is quite adequate for reaching the objectives of this chapter since it can provide analytical data from academic papers. With the support of data analysis software such as Microsoft Excel, Vantage Point, and Gephi, it is possible to analyze the results obtained through data collection.

3.1 Bibliometrics and social network analysis

The need for accountability in scientific research emerged in 1960s when the peer review process was under increasing pressure. Driven by economic constraints, it lacked well-defined criteria to prioritize funding for new research since the newborn called “Big Science” required capital intensity. Also, new scientific collaborative and multidisciplinary fields had emerged, demanding more coordination.

Science is measured by internal and external criteria. Internal criteria consist of how well research is done; external criteria on why this particular field is essential to pursue. King (1987) places emphasis in the elements for internal criteria: the readiness of the research field for exploitation and the competence of scientists in a particular research field. The second question has driven the research on quantitative analysis handling bibliometric and other science indicators. The popularization of indicators to assess scientific outputs for policymakers and research managers has increased since the 1980’s.

Scientometrics is defined by Hess (1997) as a “quantitative study of science, communication in science, and science policy”. Mingers & Leydesdorff (2015) also, complement the definition by differentiating among scientometric subfields:

- Bibliometrics: application of analytical tools (mathematics and statistics) to books, papers, patents and media communication metadata.
- Informetrics: quantitative analysis of information science objects, including production, dissemination, and use of several forms of information.
- Webometrics: analysis of webpages as if they were documents, based on bibliometric and informetric approaches (PRIEM ET AL., 2012).
- Altmetrics: quantitative analysis of the activity in social media and online tools (PRIEM ET AL., 2012).

Science indicators include research output, such as publication counting and citation frequency, and research inputs such as funds, researchers, staff, and equipment. Bibliometric indicators have gained more relevance and interest since the development in the USA of Science Citation Index (SCI) and Computer Horizons Inc (CHI) database, which made bibliometric data widely available. The use of literature indicators in the Science Indicators reports, sponsored by National Science Foundations (NSF), was fundamental in stimulating the scientific interest in this field.

The primary sources of bibliometric analysis data are the Web of Science (WoS) and Scopus (MINGERS & LEYDESDORFF, 2015). Both databases can provide significant metadata about publications in a determined research field. Title, abstract, authors, affiliation, and keywords are examples of the type of information retrieved.

Several analyses can be performed with this data, such as descriptive statistics, text mining, social network, and cluster analysis.

A Social Network Analysis is a distinct research perspective because it is based on the assumption of the importance of the actors' relationship (WASSERMAN & FAUST, 1994). Thus, the relationship between actors is the fundamental component in this approach. The network models conceptualize the social, economic, or political structure and explicit the patterns of relations.

The key elements of network analysis are the actor, the relational tie, the dyad, the triad, the subgroup, the group, and the social network. The actor refers to the social entities, which means the discrete individual, corporate, or collective social units. If all the actors are the same type, it is called a one-mode network. If the same network actors are different units (for example, researchers and research institutes), it is called a two-mode network. The links among actors are the relational ties, and that could be the evaluation of one person to another (i.e., friendship), transfer of material resources, association or affiliation, behavioral interaction, movement between places, physical connection, formal relation or biological relationship (WASSERMAN & FAUST, 1994). Dyads are the relationship between two actors, and triads are the relationship between three actors. The relationship between three or more actors is called a subgroup. A group is a finite set of actors and relation ties through which networks measurement are made. Dryad, triad, subgroups, and groups are considered a unit of analysis in Social Network Analysis. Finally, the social network is a set or set of actors and the relationship among them.

Both approaches, bibliometrics and social network analysis are relevant in the context of this research as they can express quantitatively and qualitatively the research efforts of futures studies supported by Machine Learning and Big Data tools and techniques.

In this work, bibliometrics presents, through publications on the field, the most relevant approaches, the countries and institutions that stand out in this context, and the methodologies of futures studies that best adapt to the integration of BDML techniques. Social Network Analysis aims to show the collaborations between countries, institutions, and authors and the connection of these authors through the similarity of their conceptual bases. The modularity algorithm classifies the nodes in clusters by calculating the fraction of edges falling within the given groups minus the expected such fraction if edges were randomly distributed (DOTSIKA & WATKINS, 2017). This technique highlights the groups or clusters of nodes who have more intense connections, facilitating the consolidated analysis of the sub-themes explored in this field of research. Both approaches complement each other.

Some computational tools are essential for this analysis to be performed: Vantage Point, Microsoft Excel and Gephi. Vantage Point is a text-mining software chosen to perform the analysis and collect the desired outcome based on the occurrence and co-occurrence of the name of methodologies, countries of publication, and field of applications. Produced by "Search Technology Inc.", Vantage Point is defined as "a powerful text-mining tool for discovering knowledge in search results from patent and literature databases" (VANTAGE POINT, 2021).

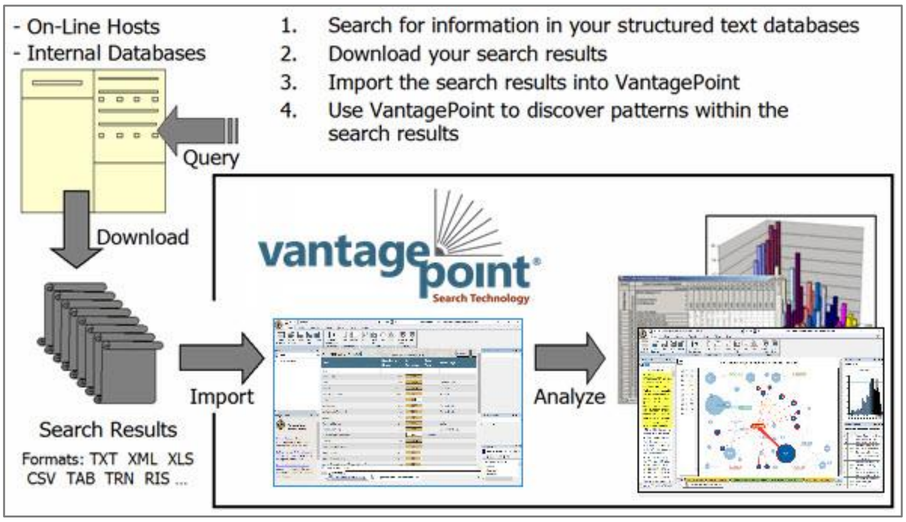


Figure 24 - Process of search and use of Vantage Point Software.

Source: Vantage Point (2021)

Microsoft Excel is a spreadsheet software developed by Microsoft and used for data analysis and visualization. Gephi is an open-source software for network analysis and visualization. Written in Java, it was developed by French students in 2008 and allows to create several types of graphs and networks. Gephi also contains several plugins to calculate social network metrics as centrality metrics (degree, betweenness, closeness), network density, path length, diameter, modularity, and clustering coefficient. Layout algorithms provided by Gephi can give shape to the graph to increase quality, efficiency, and readability.

3.2 Methodological steps: bibliometric analysis

Based on an established bibliometric methodology (HESS, 1997), seven steps were presented in the Table 6, to describe the activities and the employed tools in each phase.

Table 6 - Methodological steps for bibliometric analysis

#	Step	Tool
1	Definition of keywords for searching documents based on the previous literature review	Mendeley
2	Boolean query string definition for searching and collection peer review documents at Scopus and WoS databases	Scopus; WoS
3	Data corpus fusion, remove inconsistent and duplicate records	Vantage Point
4	Field fusion and data cleaning	Vantage Point
5	Classification of author and funding organizations	Vantage Point
6	Classification of methodologies based on Natural Language Processing (NLP) of Abstract fields and manual reading	Vantage Point
7	Data visualization: creation of tables, charts and network graphs	Microsoft Excel; Gephi

Source: Author's elaboration

3.2.1 Keywords definition for bibliometrics

Selecting appropriate keywords is crucial to define the database that will compose the bibliometric analysis. Thus, through the literature review, supported by Mendeley software[7], it was possible to select precise keywords to obtain the papers' desired corpus. The keywords were confirmed and reviewed by foresight and big data specialists.

It was expected to obtain the academic publications that focus on Foresight in Science, Technology, or Innovation (FSTI), which had some mention of Big Data or Machine Learning (BDML). Thus, it was operated three different searches to collect data from publications about "Foresight in STI" (FSTI database), publications about

[7] Mendeley is a Reference Management tool and academic social network provided by Elsevier

"Machine Learning and Big Data" (BDML database), and finally publications about "Foresight in STI with Machine Learning and Big Data" (FSTI+BDML).

Keywords referred to futures studies were associated with the keywords "science", "technology," or "innovation" to focus on foresight for STI. Expressions such as "future study" and "future research" were not employed because they can bring several documents that are not related to foresight or futures studies. Referring to Big Data and Machine Learning, the keywords were big data, machine learning, text mining, data mining, data analytics and deep learning. Table 7 presents the keywords.

Table 7 - Keywords for Foresight STI and BDML

Foresight in STI (FSTI)		Machine Learning and Big Data (BDML)
Foresight	STI	
- futuristic* - futurolog* - la prospective - foresight - technolog* forecast* - technolog* anticipation - technolog* prediction - future oriented stud* - future oriented analysis	- science - technology - innovation	- big data - text mining - data mining - machine learning - data analytics - deep learning

Source: Author's elaboration

3.2.2 Boolean query string definition for bibliometrics

Two widely known academic databases were chosen as data sources: Scopus and Web of Science. The portal Web of Science (WoS) is the referential database of Clarivate Analytics that indexes 161 million records across 254 subject areas. It includes full-text articles, reviews, editorials, chronologies, abstracts, proceedings (journals and book-based), and technical papers from more than 12,000 high impact journals and more than 160.000 conference proceedings. Scopus database indexes more than 75.000 items from more than 5.000 publishers and 16.000 authors profiles. A product of Elsevier, Scopus also offers several citation metrics and tools for analyzing high impact publication. These two databases comprehend a substantial part of academic research in the featured field.

Three Boolean string queries were constructed to embody all selected keywords that are related to each research theme. Table 8 presents all performed searches to collect the required data, indicating the source database, the name of the data corpus, the related Boolean query, the number of documents found, and the search data.

Table 8 - Boolean queries and document results

Source Database	Corpus Name	Query	# Docs	Search Date
Scopus	Foresight in STI (FSTI Scopus)	TITLE-ABS-KEY ("futuristic*" OR "futurolog*" OR "la prospective" OR "foresight" OR "technolog* forecast*" OR "technolog* anticipation" OR "technolog* prediction" OR "future oriented stud*" OR "future oriented analys*") AND TITLE-ABS-KEY ("science" OR "technolog*" OR "innovation") AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "re"))	13.369	July, 27th, 2020
Scopus	Machine Learning and Big Data (BDML Scopus)	TITLE-ABS-KEY ("big data" OR "text mining" OR "data mining" OR "machine learning" OR "data analytics" OR "deep learning") AND (LIMIT-TO (DOCTYPE ,"ar") OR LIMIT-TO (DOCTYPE, "re"))	208.903	July, 27th, 2020
Scopus	Foresight STI + Machine Learning and Big Data (FSTI+BDML Scopus)	TITLE-ABS-KEY ("futuristic*" OR "futurolog*" OR "la prospective" OR "foresight" OR "technolog* forecast*" OR "technolog* anticipation" OR "technolog* prediction" OR "future oriented stud*" OR "future oriented analys*") AND TITLE-ABS-KEY ("science" OR "technolog*" OR "innovation") AND TITLE-ABS-KEY ("big data" OR "text mining" OR "data mining" OR "machine learning" OR "data analytics" OR "deep learning") AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "re"))	235	July, 27th, 2020
WoS	Foresight STI (FSTI WoS)	TS=("futuristic*" OR "futurolog*" OR "la prospective" OR "foresight" OR "technolog* forecast*" OR "technolog* anticipation" OR "technolog* prediction" OR "future oriented stud*" OR "future oriented analys*") AND TS=("science" OR "technolog*" OR "innovation") Refined by: DOCUMENT TYPES: (ARTICLE OR REVIEW) Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.	2.884	July, 27th, 2020
WoS	Machine Learning and Big Data (BDML WoS)	TS=("big data" OR "text mining" OR "data mining" OR "machine learning" OR "data analytics" OR "deep learning") Refined by: DOCUMENT TYPES: (ARTICLE OR REVIEW) Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.	146.446	July, 27th, 2020
WoS	Foresight STI + Machine Learning and Big Data (FSTI+BDML WoS)	TS=("futuristic*" OR "futurolog*" OR "la prospective" OR "foresight" OR "technolog* forecast*" OR "technolog* anticipation" OR "technolog* prediction" OR "future oriented stud*" OR "future oriented analys*") AND TS=("science" OR "technolog*" OR "innovation") AND TS=("big data" OR "text mining" OR "data mining" OR "machine learning" OR "data analytics" OR "deep learning") Refined by: DOCUMENT TYPES: (ARTICLE OR REVIEW) Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.	135	July, 27th, 2020

Source: Author’s elaboration

3.2.3 Data fusion

The collected databases that refer to the same subject were integrated by using the software Vantage Point. Thus, it was identified and combined all duplicates title records. The number of unique documents belonging to each thematic data corpus is presented in Table 9.

Table 9 - Unique documents on each thematic data corpus

Base	New Corpus Name	Corpus Analyzed	Performed by	# Unique Docs
Scopus + Wos	**Foresight STI (FSTI)**	- FSTI Scopus - FSTI WoS	Vantage Point Software	14.461
Scopus + Wos	**Machine Learning and Big Data (BDML)**	- BDML Scopus - BDML WoS	Vantage Point Software	227.152
Scopus + Wos	**Foresight STI + Machine Learning and Big Data (FSTI+BDML)**	- FSTI+BDML Scopus - FSTI+BDML WoS	Vantage Point Software	270

Source: Author's elaboration

As is shown in Figure 25, the primary data corpus used to perform most of the analysis and to construct the networks are the intersection of the two data corpus: "Foresight in STI" (FSTI) and "Machine Learning and Big Data" (BDML).

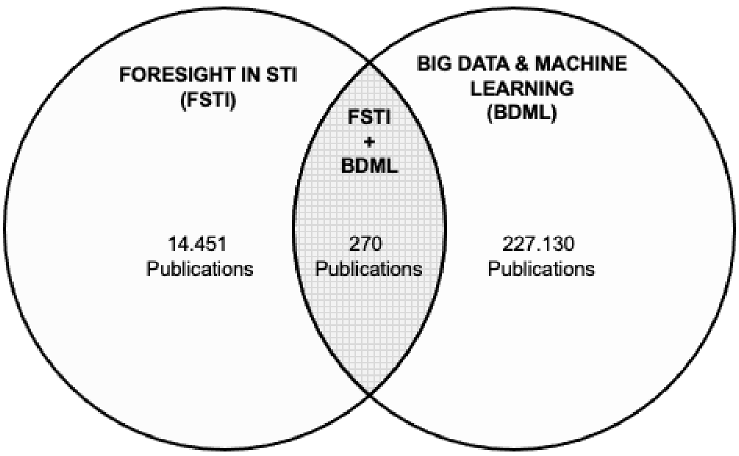


Figure 25 - Model of data collection results for FSTI, BDML and FSTI+BDML (1968-2019)

Source: Author's elaboration

3.2.4 Data cleaning

After data fusion, data fields that refers to the same information were merged. Each database exports similar data fields with different names, and merging these fields is necessary to analyze the data collected, regardless the source. After the fields were merged, the data was cleaned, which means the information was prepared for analysis by removing or modifying incorrect, incomplete, irrelevant, duplicate, and improperly formatted records in a fuzzy clustering technique. Data cleaning is an essential step of data analysis because, through this process, it is possible to assure that high-quality data is presented. High-quality data include accuracy, completeness, consistency, and uniformity to perform better the analysis proposed. Vantage Point supported this process.

The final fields names, after merging and cleaning, are presented below:

- <u>Abstract:</u> field that contains the full abstract of the documents
- <u>Abstract (NLP):</u> words and group of words (except stop words) found in abstract
- <u>Affiliations:</u> institution that the authors are part of or affiliated with.
- <u>Authors:</u> authors of the documents in the database.
- <u>Cited Authors:</u> authors of the cited documents by the papers
- <u>Cited References:</u> title, year, and authors of the cited documents by the database papers
- <u>Cited Year:</u> publication year of the cited documents by the database papers
- <u>Country:</u> country of the institution that the authors are affiliated
- <u>Authors' E-mail:</u> e-mail address of the main author
- <u>Funding Organizations:</u> institution that funded the research and was acknowledged in the paper
- <u>Journal:</u> journal that published the paper
- <u>Author Keywords:</u> indexed keyword suggested by the authors
- <u>Publication Year:</u> Year that the document was published
- <u>Title:</u> title of the paper
- <u>Total Times Cited:</u> number of citations of the document on Scopus or WoS platform.

3.2.5 Organizations' classification

Two fields contain organization names: Affiliation and Funding Organizations. These organizations were classified according to its natures, found via manual searching.

The categories in which the authors' affiliations were classified are shown below:

- Government: governmental institution as defense research institutes, governmental commissions, and national agencies for strategic research.
- University: both private and public research and higher education universities
- Research Institute: both private and public research institutes
- Company: private organizations including multinational enterprises and consultancy firms

The categories in which the funding organizations were classified are shown below:

- Government: direct governmental institution as ministries, national councils, and defense research institutes.
- Research Project: national and international thematic projects funded by different organizations.
- Foundation: domestic research foundations
- University: university scholarships and academic support.

3.2.6 Methodologies and BDML techniques classification

The methodologies were classified through Natural Language Processing of abstracts (Abstract NLP field) and manual reading. This field includes words and concepts present in the articles' abstract and provides more detailed and accurate information about the research content, such as its context, methodologies, impacts, and results.

Thus, terms corresponding to the methodologies were sought, based on the 33 foresight methodologies identified by Popper (2008) as qualitative, quantitative, and semi-quantitative methods. Techniques based on Big Data and Machine Learning were also classified by reading the full abstracts. These new methods were classified as BDML Techniques. The methodologies found in the papers are presented in the Table 10:

Table 10 - Methodologies classification

Qualitative Methods (Popper 2008)	Quantitative Methods (Popper, 2008)	Semi Quantitative Methods (Popper, 2008)	BDML Techniques
• Scenario • Expert Panels • Literature Review • Survey • Science Fictioning (SF) • Weak Signals	• Benchmark • Bibliometrics • Time Series Analysis • Modelling • Patent Analysis • Trend Extrapolation Analysis	• Cross-Impact Analysis or Structural Analysis • Delphi • Key or Critical Technologies • Multi Criteria Analysis • Roadmapping	• Big Data Analytics • Neural Networks • Literature-based Discovery (LBD) • Social Network Analysis (SNA) • Cluster Analysis • Deep Learning • Data Mining • Text Mining • Algorithm • Statistics

Source: Authors elaboration based on Popper (2008)

3.2.7 Data visualization

Data visualization is a graphic representation of data to communicate and highlight the relationship and the detected patterns between data values.

For data visualization, all data was transferred to Microsoft Excel to create tables and charts. Adjacency matrixes (authorship, bibliographic coupling and cooccurrence) were created on Vantage Point and then transferred to Gephi for network visualization and analysis. All graphs and images produced by the data collected in this research are presented in the next section.

3.3 Results and discussion

This section presents the results of the bibliometric analysis performed for this chapter. The first subsection will explore an overview of data results, showing indicators such as publications and citations per year. The subsequent sections explore different field analysis and also some correlational analysis.

It is analyzed countries, journals, affiliations, funding organizations, authors, methods, and keywords fields. The last section of this chapter presents the actual panorama of futures studies based on literature review and data analysis.

3.3.1 Overview

Before beginning to analyze in-depth the panorama of futures studies that use Big Data and Machine Learning tools, this research focused on the dynamics of publishing in two thematic areas: Foresight in STI (FSTI) and Big Data and Machine Learning (BDML) separately. Each of the databases analyzed follows its own publication dynamics according to the scientific community's interest around the research field.

The publications on Foresight in STI (FSTI) date back to the 1960s. One of the first published papers in this database is entitled "Project Sappho: A study in industrial innovation", published by Curnow R. C. and Moring G. G., affiliated to SPRU, published in volume 1, issue 2 of the Futures Journal (CURNOW & MORING, 1968). This publication indicates a strong connection between foresight and innovation studies. As shown in Figure 26, publications on this topic peaked in 2009, with 908 publications. However, they maintain approximately 440 publications per year in the past 20 years (1998-2019), which correspond to 37% of the total publication in FSTI.

Regarding publications about Big Data and Machine Learning (BDML), the annual average of publications in the last 20 years (1998-2019) is approximately 10,000 documents, with exponential growth from the first decade of the 21st century, reaching its peak in 2019, with more than 49,000 publications on the topic. The sum of documents published in the last 20 years on Machine Learning and Big Data correspond to 99% of all publications on these topics, showing the growing emergence of this field over the past two decades.

In this research's central database, which corresponds to the intersection between the two themes presented before, the first publications in Foresight in STI with Big Data and Machine Learning (FSTI+BDML) date back to 1996 and show an average of 11 documents published annually. However, FSTI+BDML publications follow an increasing curve of publications, peaking in 2019, with 48 new publications.

The annual dispersion of publications can be seen in Figure 26.

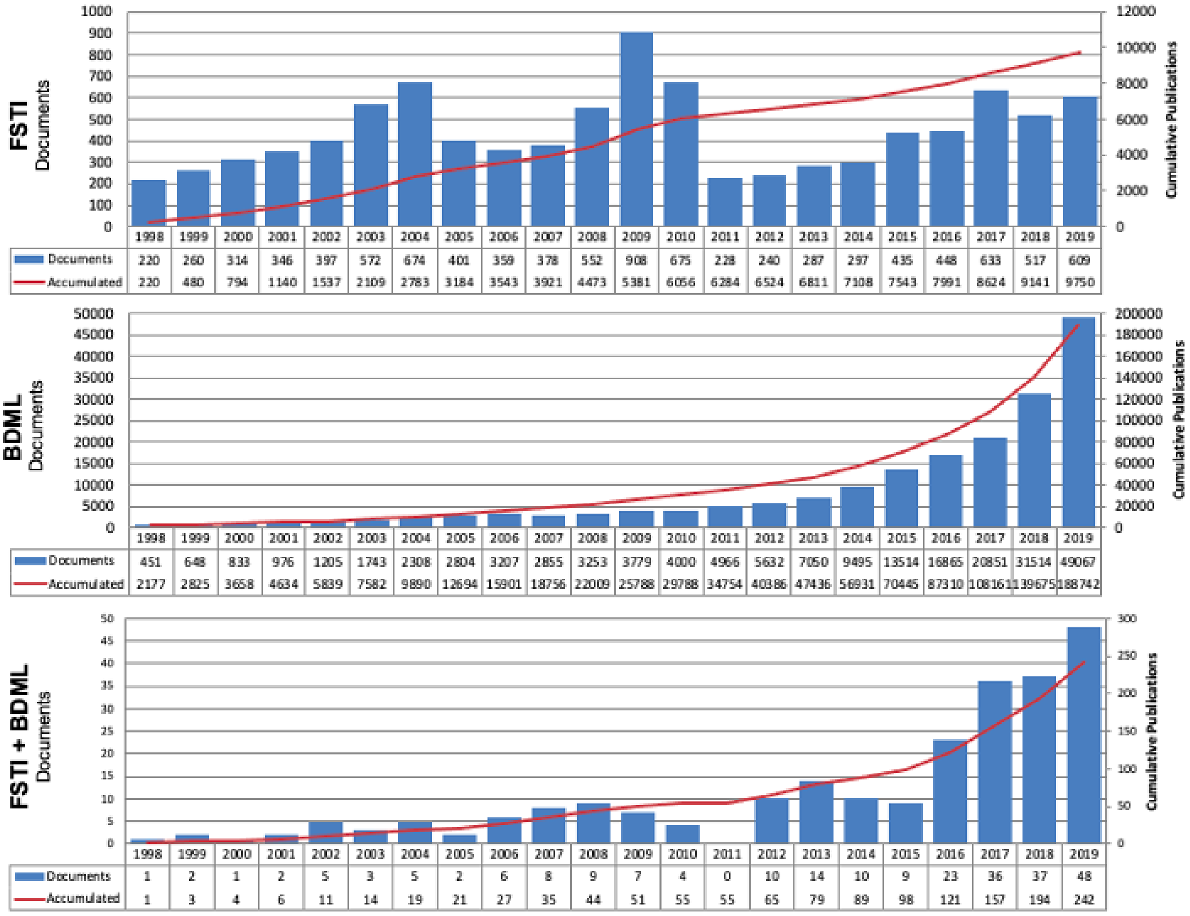


Figure 26 - Publications by year in FSTI, BDML and FSTI+BDML

Source: Author's elaboration

Although publications related to FSTI+BDML have evolved year by year, there is still a low volume of publications on the topic. While the developments of Big Data and Machine Learning for general use have grown enormously since the early '00s, publications involving foresight, STI, Big Data, and Machine Learning have been emerging after 2015.

From now on, we will analyze only the central database about Foresight in STI and BDML (FSTI+BDML).

3.3.2 Country analysis

Taking into account the 270 publications found in the intersection of two databases (FSTI + BDML, see on Figure 25) the USA stands out with 71 publications in which one or more authors belong to an American institution. This value corresponds to 26% of published studies. South Korea also stands out, with 38 publications (14%

of publications), followed by India, with 32 publications (12%), United Kingdom, with 24 and China, with 23 (both with 8% each). Brazil emerges for being the single South American in the top 10 countries in number of publications, with 13 publications (5%). The first five countries together (except duplicates) account for 165 publications, concentrating 60% of the total papers analyzed.

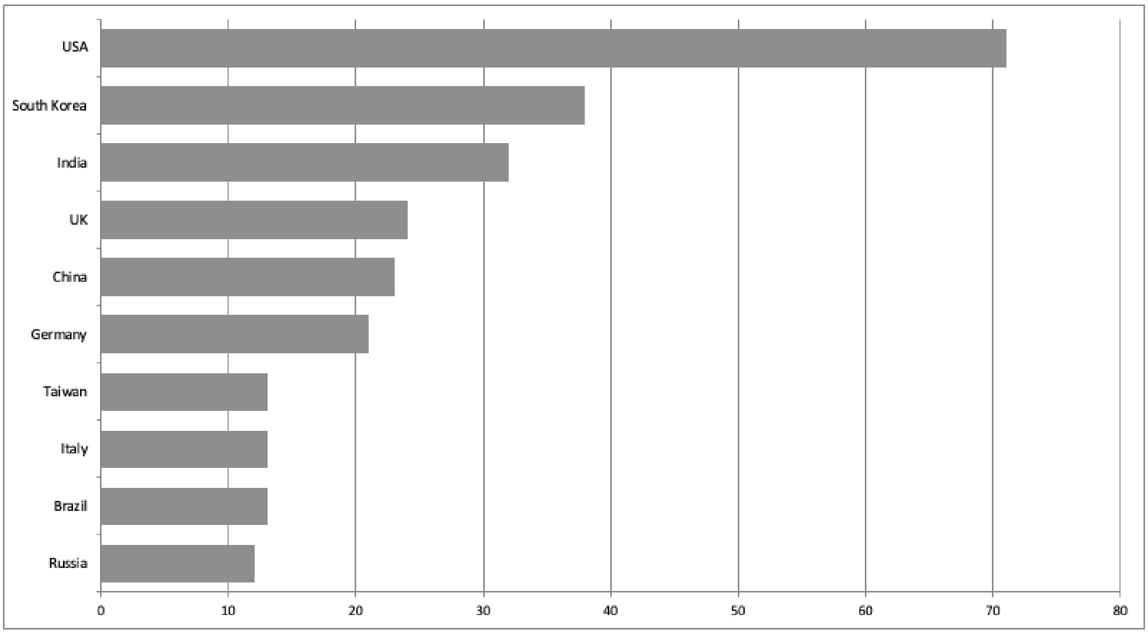


Figure 27 - Publication per country in FSTI+BDML papers

Source: Author's elaboration

According to total citations, the USA and South Korea emerged as the leading countries, with 2080 and 789 citations, respectively. Germany, Brazil, and Russia figure on the ten most productive countries (in the number of publications), but not gather on the top 10 most influential countries (in the total of citations). On the other hand, Slovenia, Netherlands, and Sweden are among the top 10 most influential countries, but they don't figure among the top 10 productive countries. Table 11 show the complete list of top 10 productive and influential nations and the total publications (TP) and total citations (TC) of the papers published by these countries

The most cited paper in this database is "Machine learning for medical diagnosis: History, state of the art and perspective", published by the journal "Artificial Intelligence in Medicine" authored by Igor Kononenko, affiliated to the Faculty of Computer / Information Science at University of Ljubljana, Slovenia, cited 585 times, which pulls this country to the 4th place on the list of most influential countries.

Table 11 - Productive and influential countries in FSTI+BDML papers

	Top 10 productive countries			Top 10 influential countries		
R	**Country**	**TP***	**TC****	**Country**	**TP***	**TC****
1	USA	71	2080	USA	71	2080
2	South Korea	38	789	South Korea	38	789
3	India	32	251	Taiwan	13	607
4	UK	24	495	Slovenia	1	585
5	China	23	291	UK	24	495
6	Germany	21	121	China	23	291
7	Brazil	13	155	India	32	251
8	Italy	13	200	Italy	13	200
9	Taiwan	13	607	Netherlands	9	182
10	Russia	12	62	Sweden	4	166

*** total publications**

**** total citations**

Source: Author‘s elaboration

Figure 28 shows the countries' co-occurrence network. In this network, each node represents a country that publishes about FSTI+BDML. The nodes' size indicates the value of the country's collaboration degree, that is, how many connections the node has with other nodes. The colors were defined by forming modularity clusters, calculated by the Gephi software, which indicates the groups of countries that most interact with each other.

On the network, it is possible to visualize the USA's vast relevance as a central country with connections between different clusters of countries. Also, it is relevant to note its strong collaboration with South Korea and China and other countries in the purple cluster.

The modularity analysis divides countries into six different clusters, in which only the group formed by Pakistan, Iraq, and Malaysia is not connected to the main network. The green cluster is formed mainly by European countries, which have a very connected knowledge production, emphasizing Italy, the UK, Germany, Holland, and France connections. The blue network is presented as a geographically heterogeneous network, composed of European, African, Asian, and Latin American countries. The last two clusters are more peripheral in the network. The dark green cluster is formed by Brazil, Portugal, and the Czech Republic, mostly connected with the USA. The orange cluster is formed by Iran, Belgium, and Vietnam, connected with the USA and European countries.

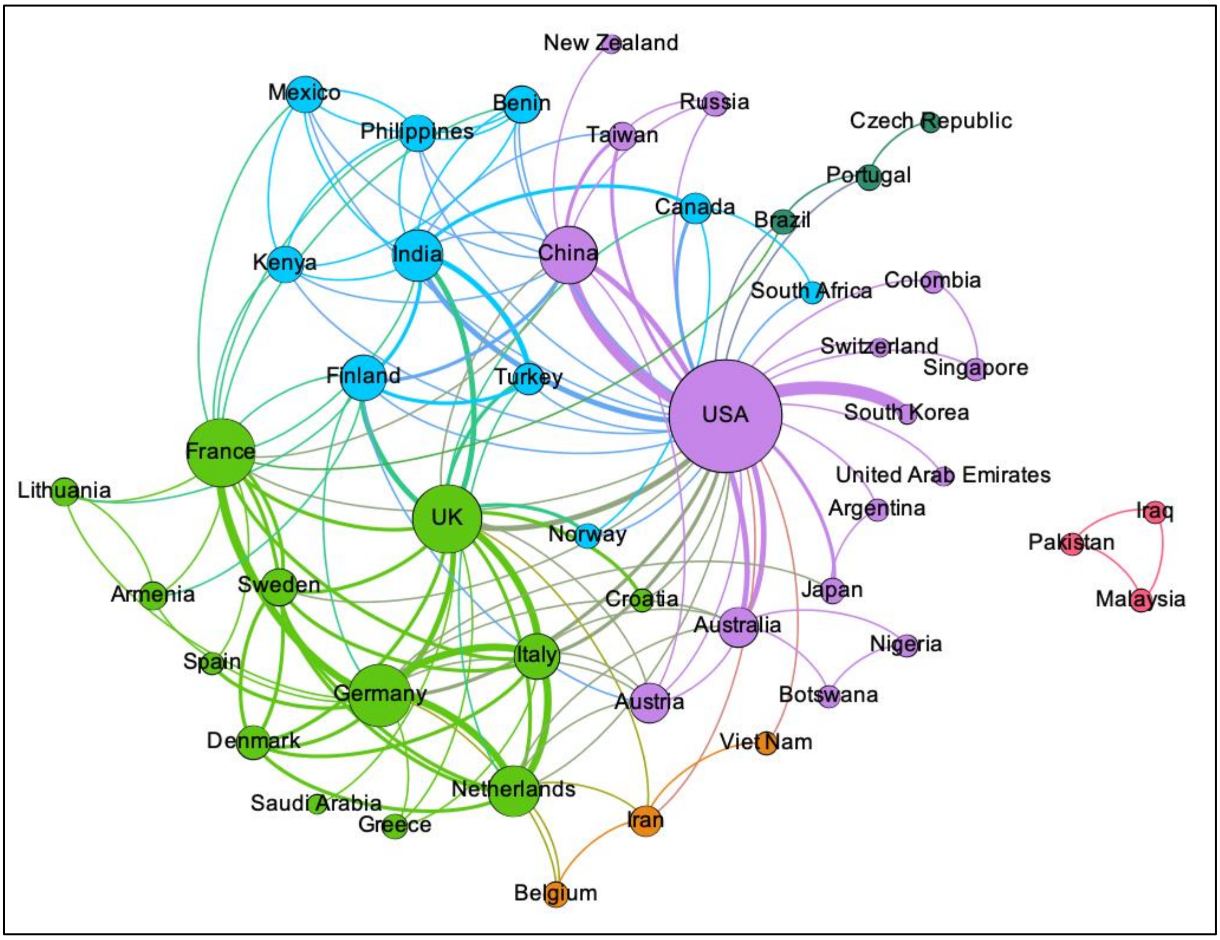


Figure 28 - Countries' co-occurrence network in FSTI+BDML papers

Source: Author's elaboration

3.3.3 Journal analysis

Technological Forecasting and Social Change (TFSC) is the most relevant journal in the database. It is the first journal in the number of publications (corresponding to 17.6% of total publications) and the number of citations (32% of the sum of citations). TFSC is also one of the journals with the highest Impact Factor among the journals analyzed, showing that the publications analyzed are high impact researches.

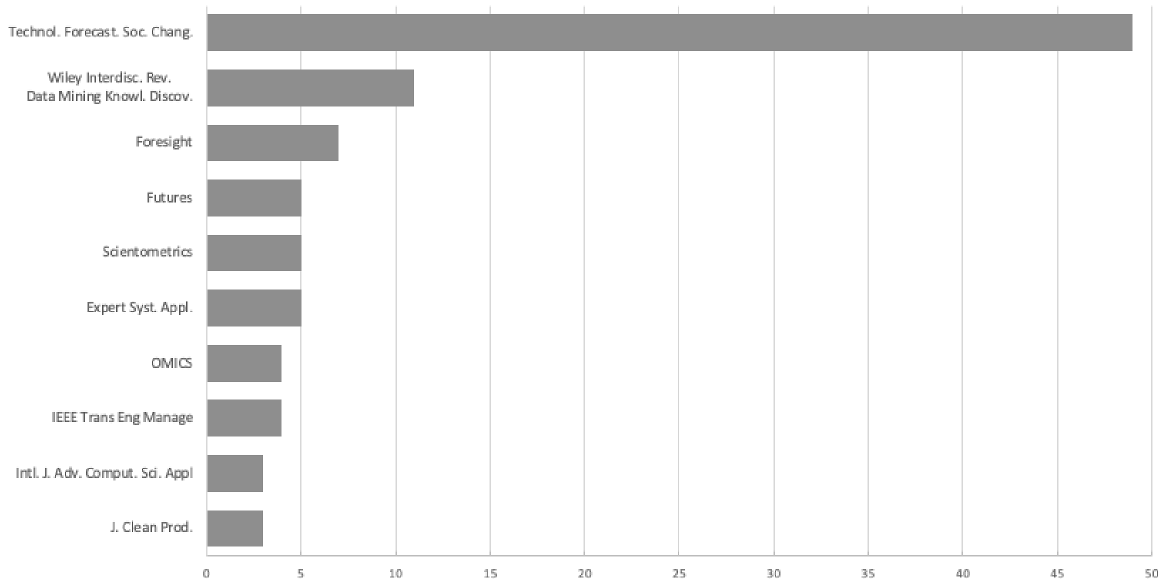


Figure 29 - Publications per journal in FSTI+BDML papers

Source: Author's elaboration

The top 5 journals in the number of publications are "TFSC", "Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery", "Foresight", "Expert Systems with Applications", "Scientometrics" and "Futures", with five or more publication on this analysis, and together they correspond to more than 30% of total publications. The presence of the journals "Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery" and "Expert Systems with Applications" shows that the use of Big Data and Machine Learning for futures studies are getting attention even in journals not dedicated to the subject.

The journals "Artificial Intelligence in Medicine" and "Technology in Society" figure on the top 10 list of most influential journals, both with a unique, highly cited publication.

In addition to TFSC, the following journals also stand out as productive and influential journals: "Expert Systems with Applications", "IEEE Transactions on Engineering Management and Technovation".

Table 12 shows the complete list of the top 10 most productive and influential journals.

Table 12 - Productive and influential journals in FST+BDML papers

R	Top 10 productive journals				Top 10 influential journals			
	Journal	TP*	TC**	IF**	Journal	TP*	TC**	IF***
1	Technol. Forec. Soc. Chang.	49	1719	5,85	Technol. Forec. Soc. Chang.	49	1719	5,85
2	Wiley I. R. Data Min Knowl.	11	12	2,54	Artif. Intell. Med	1	585	4,38
3	Foresight	7	53	5,00	Technol. Soc	1	268	2,41
4	Expert Syst. Appl.	5	222	5,45	Expert Syst. Appl.	5	222	5,45
5	Scientometrics	5	44	2,87	Technovation	3	170	4,10
6	Futures	5	21	2,77	J. Retail.	1	158	4,50
7	IEEE Trans Eng Manage	4	90	2,05	Sci. Eng. Ethics	2	131	2,79
8	OMICS	4	47	2,51	Int J Prod Econ	1	105	2,84
9	Technovation	3	170	4,10	Adv. Electron. Mater.	1	104	5,99
10	J. Clean Prod.	3	23	7,25	IEEE Trans Eng Manage	4	90	2,05

* total publications

** total citations

*** impact factor

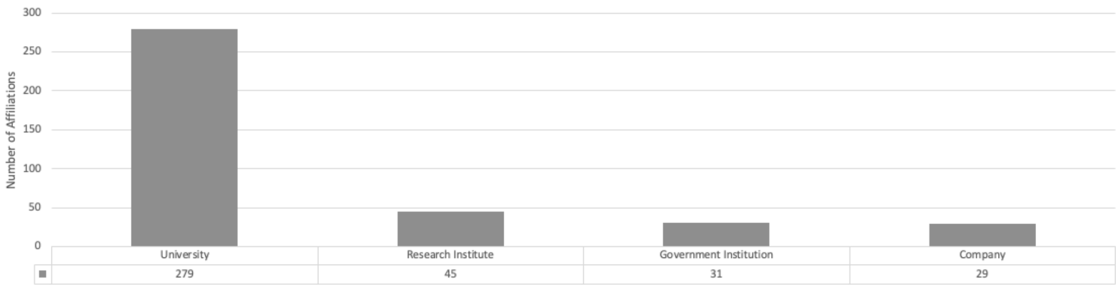
Source: Author‘s elaboration

3.3.4 Affiliation analysis

Cheongju University in South Korea is the most productive institution, with 11 publications. It is followed by the American governmental institution Office of Naval Research, with nine publications. Georgia Institute of Technology in the USA, National Research University (HSE) in Russia, and Seoul National University in South Korea are the next on the list, with eight publications each.

The Korea Institute of Science and Technology Information (KISTI) stands out as the most relevant research institute in the database. Among companies, DDL-OMNI Engineering and Search Technology are both the most productive institutions, with three papers each. DDL-OMNI Engineering is an American company, founded in 1967, acquired in 2018 by the company American Systems, responsible for offering government services of strategic solutions for American national programs in defense, intelligence, and healthcare. Search Technology is an American service company that provides the text mining software Vantage Point, which is employed in this analysis.

The graph below shows the proportion of each type of affiliation institution. It is possible to observe that universities make up the largest group of institutions (as expected by collecting academic papers), with 279 universities responsible for 234 publications (84% of the total publications). The 45 research and 31 government institutions account for 44 publications each (16% each). Companies correspond to 29 affiliations, accounting for 28 publications (10%).



R	Affiliation - University	TP	Affiliation - Research Inst.		Affiliation - Govn. Inst.		Affiliation - Company	
1	Cheongju University	11	Korea Inst. Sci.Tech.Inf. (KISTI)	6	Office of Naval Research	9	DDL-OMNI Engineering	3
2	Georgia Institute of Technology	8	Fraunhofer Institute	5	European Commission	4	Search Technology, Inc.	3
3	National Research University HSE	8	German Cancer Res. Cen. (DKFZ)	3	Naval Surface Weapons Cen.	3	Hewlett Packard Enterprise	2
4	Seoul National University	8	Danish Cancer Society Res. Cen.	2	CERIS CNR	2	IBM Corp	2
5	Beijing University of Technology	7	European Institute of Oncology	2	-	-	Robert Bosch LLC	2
	Total University Pub.	234	Total Research Institute Pub.	44	Total Government Inst. Pub.	44	Total Company Publications	28

Figure 30 - Top affiliations by classification in FSTI+BDML papers

Source: Author's elaboration

The affiliation network (Figure 31) shows the co-occurrence of affiliations on published papers. In this analysis, Georgia Institute of Technology (USA), Indian Institute of technology Kharagpur (India), the German Cancer Research Centre (Germany), University of Groningen (Netherlands), Indian Institute of Technology Bombay (India), Gaziantep University (Turkey) and the Massachusetts General Hospital (USA) stand out as the institutions with the highest degree, that is, institutions that are the most connected with other institutions. Georgia Institute of Technology is a central institution in this subject, with collaboration with companies, research institutes and other universities. These institutions are highlighted on the network.

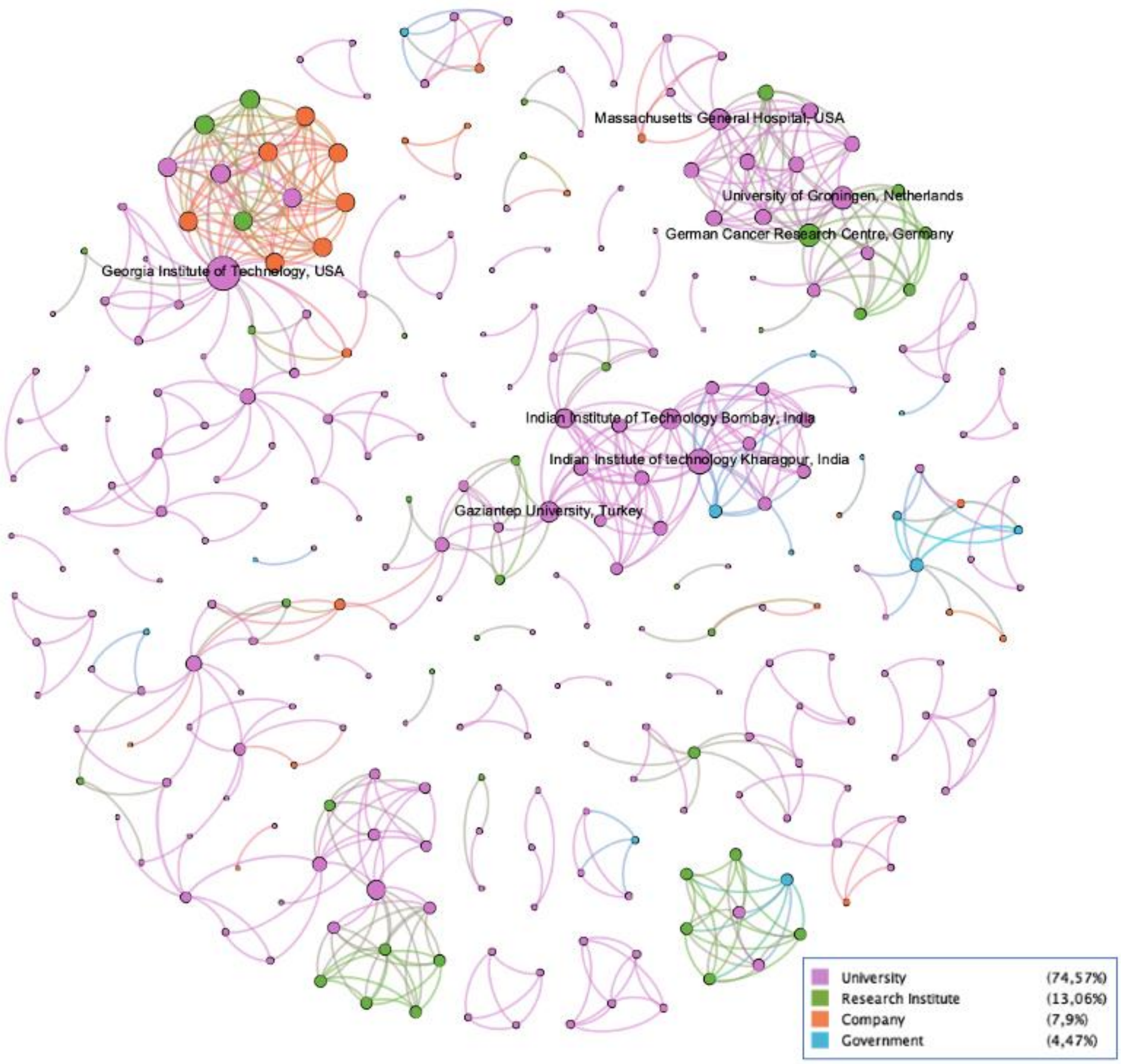


Figure 31 - Affiliation's co-occurrence network in FSTI+BDML papers

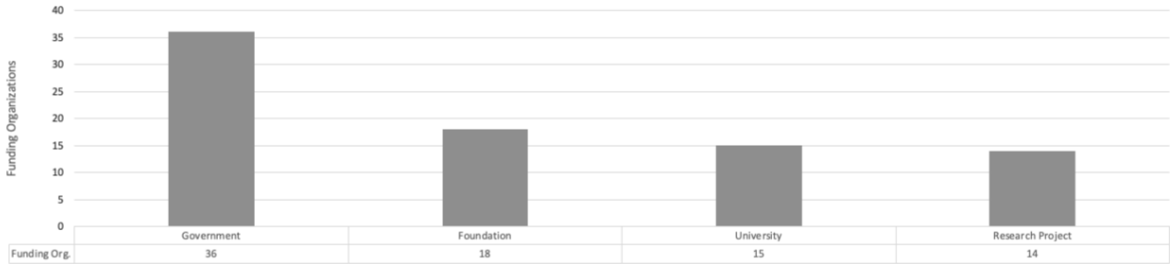
Source: Author's elaboration

3.3.5 Funding organization analysis

The funding organizations are the institutions that fund or support the projects and research presented in the papers. Among the documents analyzed on FSTI+BDML, 97 papers (35% of total publications) mention the funding organizations that financed the projects. Among 83 mentioned funding institutions, 36 of these are government institutions (43% of funding organizations), such as ministries, national councils, and defense research institutes, and account for financing 45 papers (16% of total publications).

National research foundations represent 21% of the funding Organizations (18 institutions) and are responsible for funding 44 publications (16% of total publications).

Directly university funds (15 universities) are responsible for financing 18 publications (6%) and other research projects account for 14 publications (5%).



R	Funding Org. - Governemnt	TP	Funding Og. - Foundation	TP	Funding Org. - University	TP	Funding Org. - Research Project	TP
1	Russian Min. Of Educ. and Science	5	National Res. Found. of Korea (NRF)	14	Beijing University of Technology	3	BK21 Plus	1
2	CNPq - Nat. Council for Sci. Tech.	4	National Nat. Sci. Found. China (NSFC)	10	European Commission	3	Masdar Initiative	1
3	German Fed. Min. Educ. Res. (BMBF)	3	FAPESP - SP Research Foundation	4	Ulsan National Inst. Sci. Tech. (UNIST)	2	NARO Brain	1
4	Sci. Tech. Res. Council of Turkey	3	National Science Foundation (NSF)	3	Kyung Hee University	1	IMSKA-MAT	1
5	Australian Research Council (ARC)	2	CAPES Higher Educ. Improv. Coord.	2	Masdar Institute of Sci. Tech.	1	SBIR	1
	Total University Pub.	45	Total Research Institute Pub.	44	Total University Fund. Pub.	18	Total Company Publications	14

Figure 32 - Top funding organization by classification in FSTI+BDML papers

Source: Author's elaboration

The funding organizations' co-occurrence network (Figure 33) shows the degree as the nodes' size and presents 4 clusters, classified by the level of modularity by Gephi. Each of the groups features the most central institutions of the subnet, highlighting the National Science Foundation of the USA (orange cluster), National Council for Scientific and Technological Development in Brazil (blue cluster), National Research Foundation of Korea (green cluster), Russian Ministry of Education and Natural National Science Foundation of China composing, both of them, one single cluster (purple cluster).

The orange cluster headed by the National Science Foundation also counts with the Australian Research Council, the National High Technology Research and Development Program of China, and the Department of Energy United States Department of Energy (DOE).

The blue cluster headed by the Brazilian Agency National Council for Scientific and Technological Development (CNPq) also includes the Brazilian organizations Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP), Coordination for the Improvement of Higher Level Personnel - Brazil (CAPES), Radiocommunication Reference Center (RNP) and the International Association for Hydrogen Energy (IAHE), the latter being the only international organization on this cluster.

The green cluster headed by the National Research Foundation of Korea (NRF) also includes other Korean organizations such as Future Strategic Fund of Ulsan National Institute of Science and Technology (UNIST), Brain Korea 21 PLUS Project, Korea Institute of Science, Technology Evaluation and Planning (KISTEP) and the Japanese organization Global Research Laboratory Program of the Ministry of Science.

The purple cluster headed by the National Natural Science Foundation of China and Russian Ministry of Education and Science also includes Chinese organizations Beijing University of Technology, Ministry of Education Social Science Foundation of Beijing, China Scholarship Council, China's National Key R&D Program, and the European Commission.

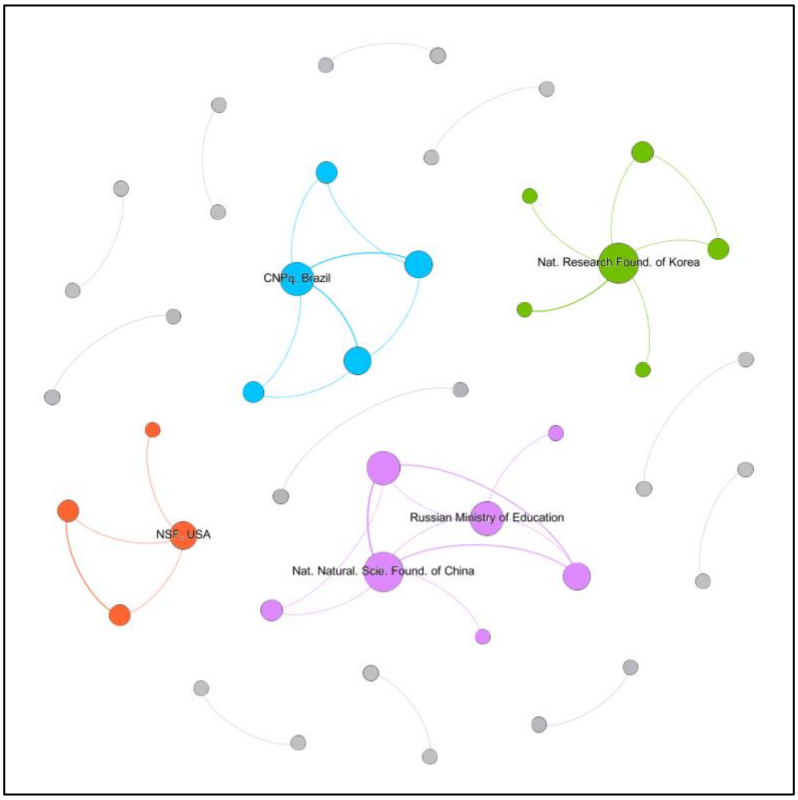


Figure 33 - Funding organizations' co-occurrence network in FSTI+BDML papers

Source: Author's elaboration

3.3.6 Author analysis

The author with the highest number of published futures studies supported by Big Data and Machine Learning (11 publications) is professor Sunghae Jun, from the Statistics Department at Cheongju University in Seoul, South Korea. Ronald N. Kostoff, affiliated to the USA Office of Naval Research, counts with eight publications, followed by Alan L. Porter, from Georgia Tech and Search Technology, with six publications. The Korean researcher Yongtae Park, together with Kostoff, Yoon, and Porter, are among the most influential authors. Among the authors with the highest number of citations, most authors have a single one high impact publication in the field.

Table 13 - Productive and influential authors in FSTI+BDML papers

R	Top 10 productive authors				Top 10 influential authors			
	Author	TP*	TC*	CPP***	Author	TP*	TC**	CPP***
1	Jun, Sunghae	11	184	16,7	Kononenko, I.	1	585	585,0
2	Kostoff, R. N.	8	441	55,1	Kostoff, R. N.	8	441	55,1
3	Porter, Alan L	6	298	49,7	Yoon, Byungun	5	331	66,2
4	Yoon, Byungun	5	331	66,2	Porter, Alan L	6	298	49,7
5	Park, Sang Sung	5	128	25,6	Rygielski, C.	1	268	268,0
6	Ozdemir, Vural	5	72	14,4	Wang, J.-C.	1	268	268,0
7	Daim, Tugrul	4	31	7,8	Yen, D. C.	1	268	268,0
8	Zhou, Yuan	4	24	6,0	Park, Y.	3	254	84,7
9	Trappey, A. J. C.	4	5	1,3	Boylan, R.	1	208	208,0
10	Park, Y.	3	254	84,7	Simons, G. R.	1	208	208,0

* total publications

** total citations

*** citation per publication

Source: Author's elaboration

The total 270 analyzed documents are published by 745 authors, of which 680 (91% of the total authors) have only one single publication, and 40 authors (5%) have published two papers. The aggregated 25 authors with three or more publications (3% of the authors) account for 64 published articles (24% of the total publications).

To better understand how authors organize themselves around major research topics, a Bibliographic Coupling network of the 65 authors with two or more publications was framed. In this network, each node represents an author. The size of the node is proportional to the author's degree. The edges represent the proportion of cited authors that both connected researchers mention in their publications. Through a Bibliographic Coupling network, it is possible to visualize how much the research published by specific author is linked with other researchers based on the bibliography included in their work. Such an approach shows connections among authors who share the same conceptual bases, even if they do not have coauthored papers.

In this way, the modularity algorithm was run to find the clusters of authors. It means modularity can aggregate authors around the same methodology, research object, or research topic. Figure 34 represents the authors' network, and the colors of their nodes represent the authors' clusters.

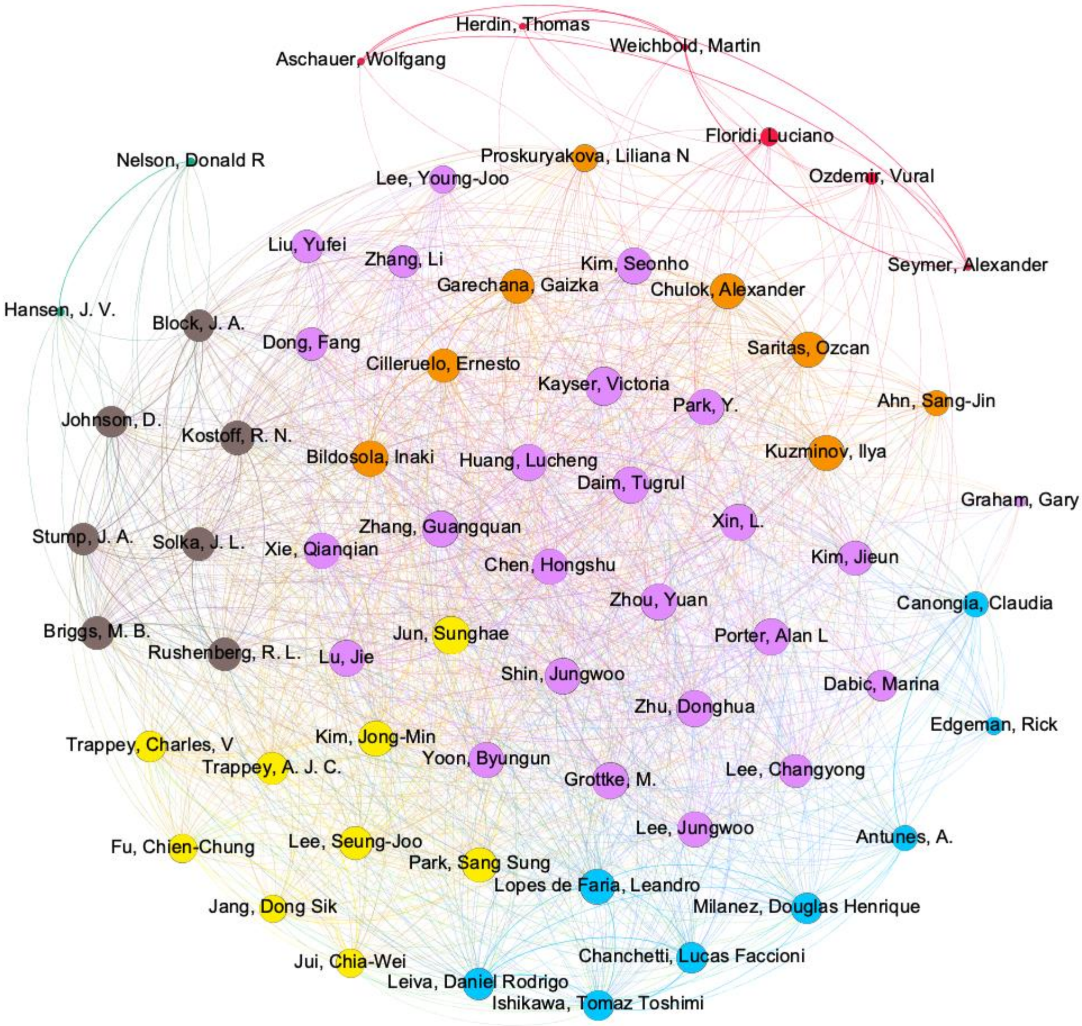


Figure 34 - Bibliographic coupling network in FSTI+BDML papers

Source: Author's elaboration

In the network were found 7 clusters, classified with different colors. In Table 14 it is possible to see the top 10 authors with the highest value of the degree, Betweenness Centrality, and the cluster to which this author belongs.

Table 14 - Author's degree, betweeness centrality and cluster in FSTI+BDML papers

Author	TP	Degree	Betweeness Centrality	Cluster
Kayser, Victoria	3	59	105,55	Purple
Kim, Seonho	2	56	95,02	Purple
Jun, Sunghae	11	59	31,94	Yellow
Floridi, Luciano	2	26	24,58	Red
Bildosola, Inaki	3	56	18,97	Orange
Kostoff, R. N.	8	52	14,86	Brown
Kuzminov, Ilya	2	55	14,60	Orange
Ozdemir, Vural	4	17	12,96	Red
Briggs, M. B.	3	51	11,91	Brown
Rushenberg, R. L.	3	51	11,91	Brown

Source: Author‘s elaboration

It is possible to recognize that the researchers Victoria Kayser and Seonho Kim (both belonging to the purple cluster) are the main bridges of the network, which means they are the authors who make the most connections between the different clusters. Sunghae Jun, Luciano Floridi, Inaki Bildosola, and Ronald Kostoff are also "bridge" authors in their respective clusters.

Table 15 presents the aggregated data for each thematic cluster, the total number of publications in the cluster, the average degree of the nodes, the number of authors belonging to the cluster, the cited keywords, the primary methods, the countries, and the types of affiliation.

The purple cluster is consolidated as the largest cluster in the network, embracing 25 authors and 39 publications. It is also the cluster with the highest average degree, which means that each node in this cluster connects on average with more than 52 other nodes. Among the most cited keywords by the cluster, "technological forecast" and "text mining" stand out. Regarding the methods, patent analysis and text mining are the most employed approaches for the purple cluster. This cluster is formed mainly by Korean, American, and Chinese authors, mostly affiliated with universities.

The yellow cluster has nine authors, 18 publications, and an average degree of 49.67. Its keywords focus on patent analysis, which is also the most used method in its papers. It is mainly composed of Korean authors affiliated with universities.

The blue cluster is composed of eight authors, ten papers, and an average degree of 43. Among the keywords of this cluster, "Bibliometrics", "S&T Indicators," and "Biotechnology" stand out. The methods used are mainly patent analysis, text mining, and time series analysis. The authors in the blue cluster are majority Brazilians, affiliated with universities.

The orange, like the blue cluster, also has eight authors and ten papers. However, they have an average degree of 50.13, higher than the blue cluster. This cluster's keywords that differ from the other clusters are "Horizon Scanning" and "Trend Analysis". Among the methodologies used, text mining and bibliometrics stand out. The authors of the orange cluster are mostly Russian, affiliated with universities.

The Brown cluster has eight publications and seven authors, connected on average to 49.86 other nodes. This cluster differs from the others by using “Literature-based Discovery” and “Expert Panels” methods in its publications. All authors of this cluster are Americans and are mainly linked to governmental institutions.

The Red cluster has six authors and eight publications, with a low average grade compared to the previous clusters, of only 11.83. It is a cluster that discusses "Big data, Evaluation" and "Ethics" through “Big Data Analytics” and “Scenario” as methods or techniques. It is composed of authors from different countries such as India, UK, Canada, Austria, and Turkey, mainly linked to universities.

The last and smallest cluster, green, has only two authors, only two publications, and the lowest average grade of the network, connecting on average with ten other nodes. This cluster differentiates itself by using “Time Series Analysis”, “Statistics” and “Neural Network” as research approaches. The two researchers in this cluster are American and linked to universities.

Table 15 - Authors' clusters keywords, methods, countries and affiliations in FSTI+BDML papers

Cluster	TP	Avg Degree	# Authors	Main Keywords	Main Methods	Main Countries	Affiliation Type
Purple	39	52,56	25	Technological forecasting [18, 46%]; Text mining [18, 46%]; Big data [8, 21%]; data mining [7, 18%]; Emerging technologies [6, 15%]	Patent Analysis [17, 44%]; Text Mining [17, 44%]; Data Mining [8, 21%]; Trend Extrapolation [6, 15%]; Cluster Analysis [6,15%]	South Korea [14, 36%]; USA [13, 33%]; China [11, 28%]; Germany [5, 13%]; UK [4, 10%]	University [37, 95%]; Research Institute [8, 21%]; Government [7, 18%]; Company [3, 8%]
Yellow	18	49,67	9	Technological forecasting [9, 50%]; Patents and inventions [4, 22%]; Patents [4, 22%]; Patent analysis [4, 22%]; patent clustering [4, 22%]	Patent Analysis [17, 94%]; Text Mining [9, 50%]; Statistics [9, 50%]; Algorithm [5, 28%]; Big data analytics [4, 22%]	South Korea [13, 72%]; Taiwan [5, 28%]; USA [3, 17%]; China [1, 6%]	University [17, 94%]; Research Institute [1, 6%]; Government [1, 6%]
Blue	10	43	8	Bibliometrics [4, 40%]; Technological forecasting [3, 30%]; S&T indicators [2, 20%]; Hydrogen Storage [2, 20%]; biotechnology [2, 20%];	Patent Analysis [7, 70%]; Text Mining [6, 60%]; Time Series Analysis [5, 50%]; Literature Review [2, 20%]; Bibliometrics [2, 20%]	Brazil [7, 70%]; Denmark [2, 20%]; France [1, 10%]	University [9, 90%]
Orange	10	50,13	8	Technological forecasting [4, 40%]; Foresight [4, 40%]; Bibliometrics [3, 30%]; Text mining [3, 30%]; Horizon scanning [2, 20%]; Trend analysis [2, 20%]	Text Mining [6, 60%]; Bibliometrics [4, 40%]; Time Series Analysis [3, 30%]; Scenario [2, 20%]; Survey [2, 20%]	Russia [5, 50%]; Spain [3, 30%]; South Korea [2, 20%]; USA [1, 10%]	University [10, 100%]; Research Institute [2, 20%]; Government [2, 20%]
Brown	8	49,86	7	Technological forecasting [8, 100%]; Literature-based discovery [8, 100%]; Text mining [8, 100%]; data mining [7, 88%]; Science and Technology [7, 88%]	Lit.based Discovery (LBD) [8, 100%]; Expert Panels [3, 38%]; Text Mining [2, 25%]; Roadmapping [2, 25%]; Cluster Analysis [1, 13%]	USA [8, 100%]	Government [8, 100%]; University [3, 38%]; Company [3, 38%]
Red	8	11,83	6	Big data [4, 50%]; Evaluation study [2, 25%]; Security perceptions [2, 25%]; ethics [2, 25%]; Artificial intelligence [2, 25%]; Social media [2, 25%]	Big data analytics [3, 38%]; Scenario [2, 25%]	India [4, 50%]; UK [3, 38%]; Canada [2, 25%]; Austria [2, 25%]; Turkey [2, 25%]	University [8, 100%]; Research Institute [2, 25%]
Green	2	10	2	Technological forecasting [2, 100%]; time series [2, 100%]; forecasting [2, 100%]; Time series analysis [2, 100%]; Exponential smoothing [1, 50%];	Statistics [2, 100%]; Time Series Analysis [1, 50%]; Modelling [1, 50%]; Algorithm [1, 50%]; Neural Networks [1, 50%]	USA [2, 100%]	University [2, 100%]

Source: Author’s elaboration

3.3.7 *Methods analysis*

After analyzing the articles' abstracts, it was possible to extract the foresight methodologies and the BDML techniques used in each study. Figure 35 shows the number of publications that use each of the analyzed methodologies, and Figure 36 the publications that use mentioned BDML techniques.

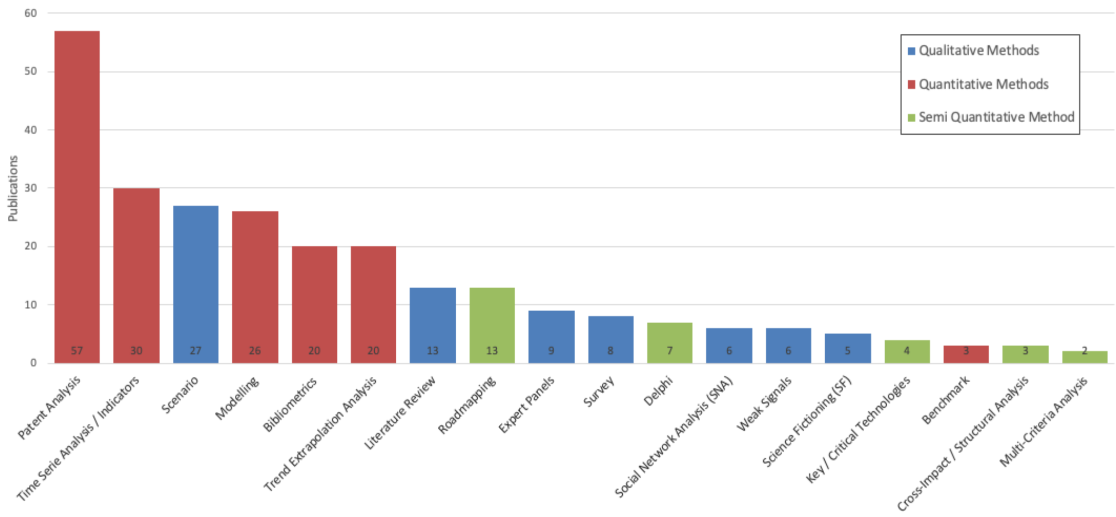


Figure 35 - Publications per methodological approach in FSTI+BDML papers

Source: Author's elaboration

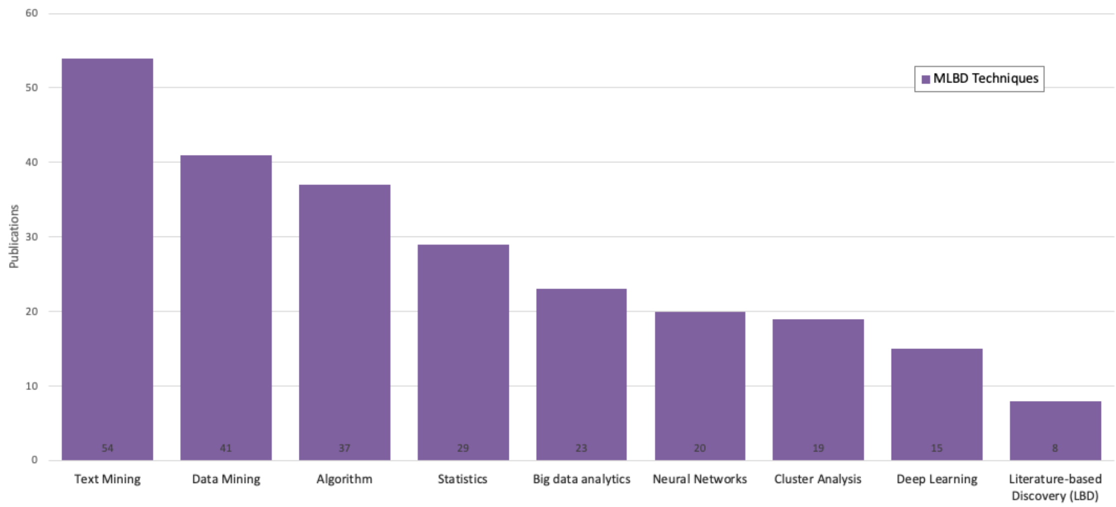


Figure 36 - Publications per BDML technique in FSTI+BDML papers

Source: Author's elaboration

The most used methodology is Patent Analysis, mentioned in 57 papers (21% of total). The most cited qualitative method among publications is

Scenario, with 27 publications (10% of total), and the most cited semi-quantitative method is Roadmapping, with 13 publications (5% of total). Regarding BDML techniques, the Text Mining technique appears in 54 papers (20% of total), Data Mining up to 41 papers (15%), and Algorithms up to 37 (14%).

BDML techniques are explicitly mentioned in 169 papers (63% of total). Quantitative methods are present in 112 publications (41% of the total), qualitative methods are presented in 54 publications (20% of total), and Semi-Quantitative methods in 27 publications (10% of total).

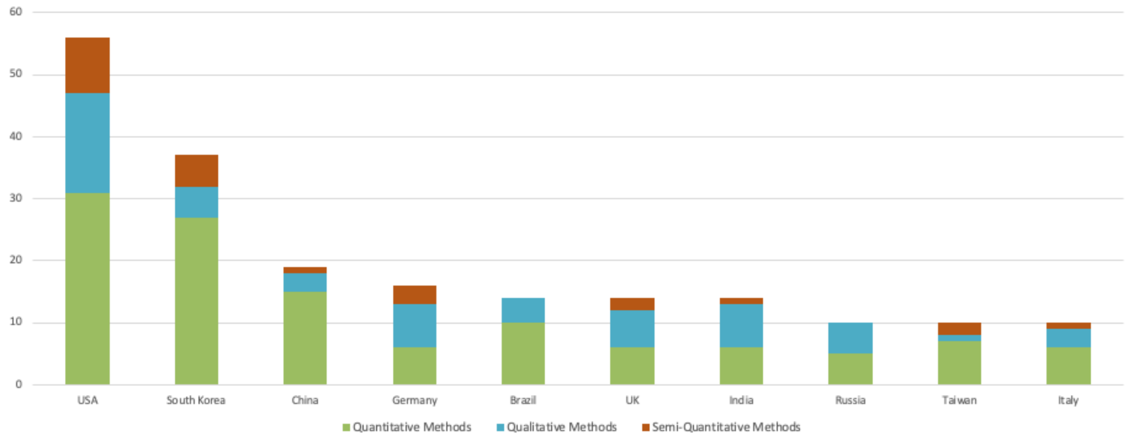


Figure 37 - Countries per type of methodological approach in FSTI+BDML papers

Source: Author's elaboration

Table 16 - Countries methodological approaches in FSTI+BDML papers

R	USA	TP	South Korea	TP	India	TP	China	TP	UK	TP
1	Patent Analysis	14	Patent Analysis	22	Scenario	4	Patent Analysis	9	Modelling	3
2	Modelling	11	Trend Extrapolation	5	Modelling	3	Trend Extrapolation	6	Scenario	2
3	Time Series	7	Modelling	3	Bibliometrics	2	Time Series	5	Literature Review	2
4	Roadmapping	6	Time Series	3	Survey	2	Modelling	3	Bibliometrics	2
5	Scenario	5	Delphi	3	Patent Analysis	1	Scenario	2	Roadmapping	2

R	Germany	TP	Brazil	TP	Taiwan	TP	Italy	TP	Russia	TP
1	Scenario	4	Patent Analysis	8	Patent Analysis	6	Bibliometrics	3	Scenario	3
2	Patent Analysis	2	Time Series	5	Trend Extrapolation	3	Patent Analysis	2	Literature Review	3
3	Modelling	2	Scenario	2	Time Series	2	Time Series	2	Survey	2
4	Time Series	2	Bibliometrics	2	Roadmapping	2	Modelling	2	Time Series	2
5	Weak Signals	2	Literature Review	2	Scenario	1	Literature Review	2	Trend Extrapolation	2

Source: Author' elaboration

Figure 37 shows the types of methods employed in publications for the 10 most productive countries on the topic. Through the figure, it is possible to state that only Germany and India published more papers applying Qualitative than Quantitative methods while Russia and the UK have the same proportion of both types of methods. Brazilian and Russian publications did not mention any Semi-Quantitative methods. Most countries' publications focus on quantitative methods, which shows the connection between data-driven methods and BDML techniques in foresight studies. Deepening on the methodologies used by each country (Table 16), half of the ten most productive countries use Patent Analysis as the most used method in their studies, with emphasis on South Korea with 22 publications with this approach (58% of the total published by Korean institutions).

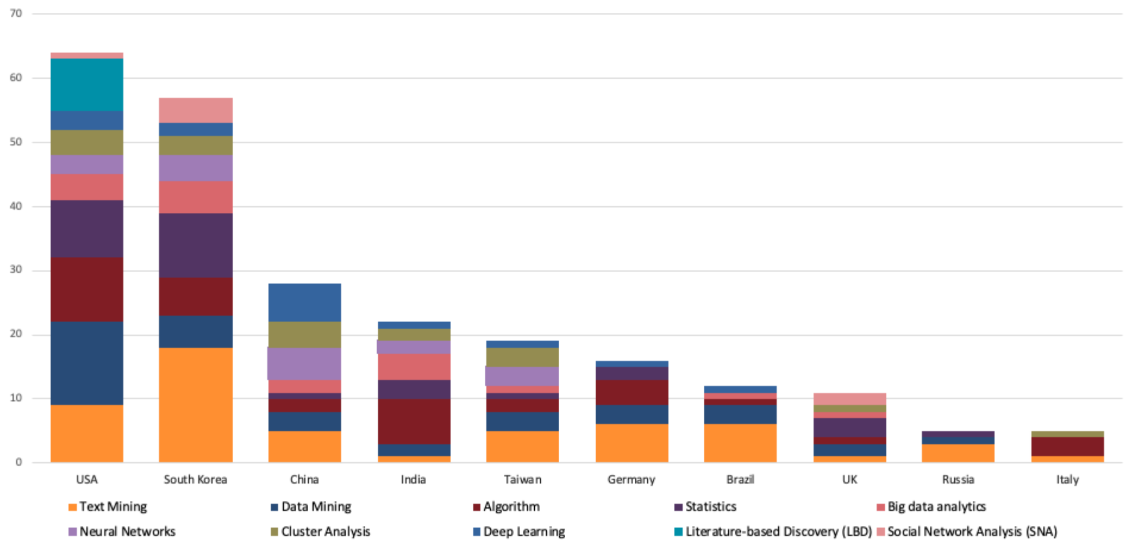


Figure 38 - Countries per BDML techniques in FSTI+BDML papers

Source: Author's elaboration

Regarding BDML techniques, Text Mining is the most used in publications from South Korea, Taiwan, Germany, Brazil, and Russia. Most Indian and Italian publications use Machine Learning Algorithms. The UK mentions Statistics the most, China mentions Deep Learning and the USA mentions Data Mining.

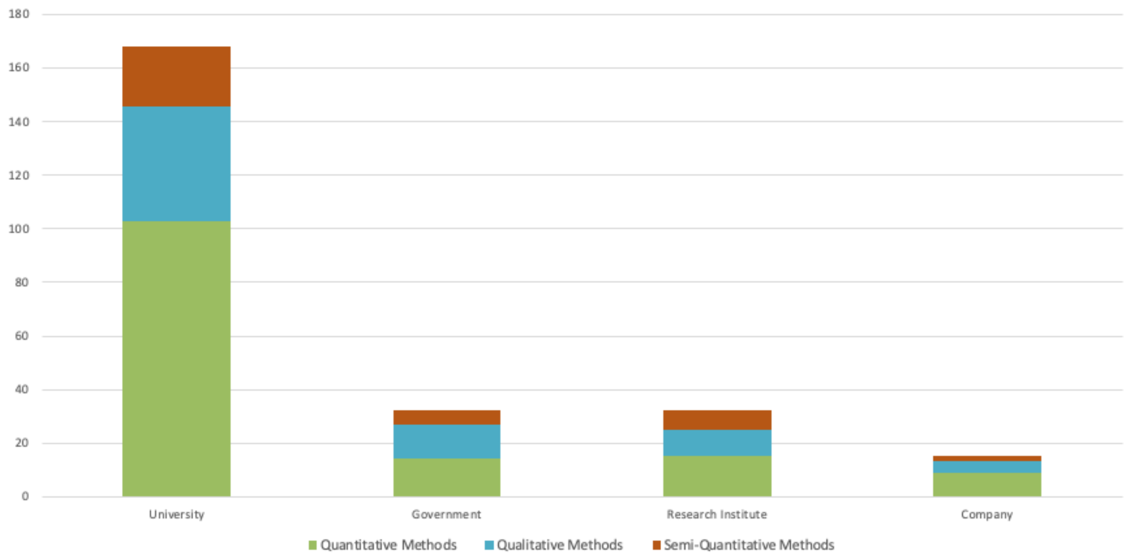


Figure 39 - Type of affiliation per type of methodological approach in FSTI+BDML papers

Source: Author's elaboration

Regarding the aggregated types of classified affiliation (University, Government, Research Institute and Company), quantitative methods are the most used approach in all types of affiliation, followed by qualitative methods and, finally, semi-quantitative methods. Patent Analysis is the most used method in Universities' and Research Institutes' papers, as shown in Table 17. The most used method in the Government's papers is Expert Panels. Companies' papers frequently employ Time Series Analysis.

Table 17 - Methodological approach per type of affiliation in FSTI+BDML papers

R	University	TP	Government	TP	Research Institute	TP	Company	TP
1	Patent Analysis	53	Expert Panels	6	Patent Analysis	7	Time Series	5
2	Time Series	29	Patent Analysis	5	Time Series	5	Patent Analysis	3
3	Modelling	25	Time Series	5	Scenario	5	Modelling	3
4	Scenario	22	Scenario	4	Bibliometrics	3	Scenario	2
5	Bibliometrics	18	Bibliometrics	4	Modelling	3	Bibliometrics	2

Source: Author's elaboration

Patent Analysis is the most used methods for authors affiliated to Universities and Research Institutes. Companies publish studies using Algorithms, followed by Data Mining and Statistics. Emphasis is given to Literature-based Discovery and Expert Panels, used in futures studies published by authors affiliated to governmental institutes, showing the government's focus in consolidating knowledge by expert opinions and published insights.

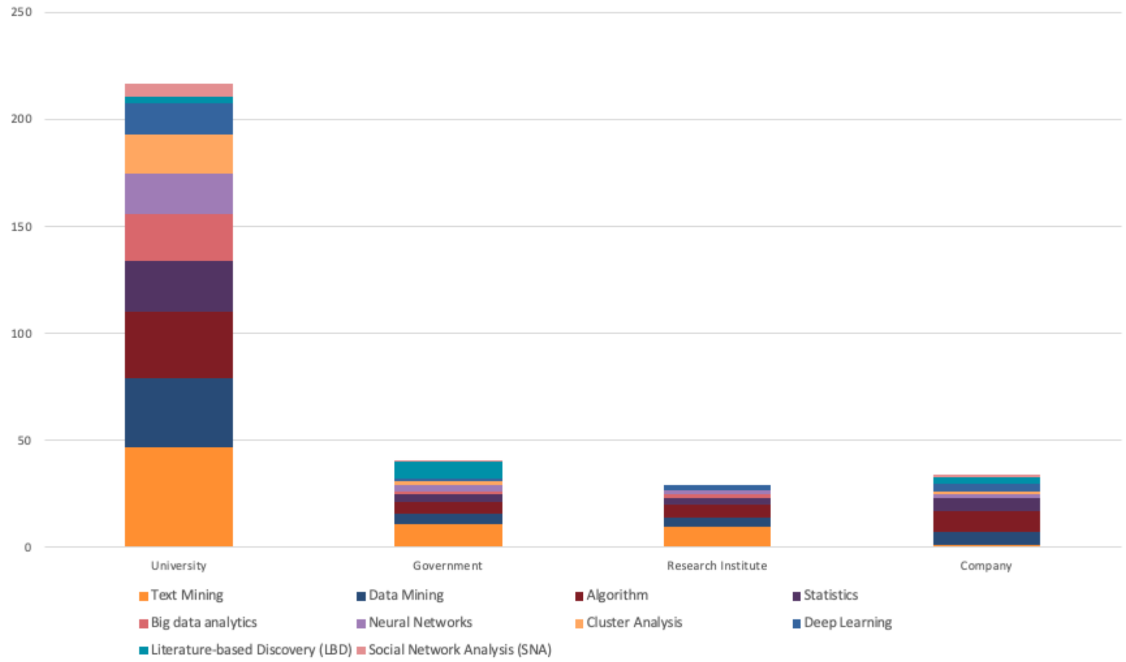


Figure 40 - BDML techniques per type of affiliation in FSTI+BDML papers

Source: Author's elaboration

In general, a futures study employs more than one single method. According to Popper (2008), foresight projects combine an average of 6 methods. Thus, the co-occurrence network of the methodologies and BDML techniques (Figure 41) shows the relationship between the futures studies' methodologies and the BDML techniques. Each node in this network represents a method or technique, the size of the node represents the number of papers that use this method or technique, and the colors of the nodes distinguish the type of method.

Through network analysis, it is possible to verify that Patent Analysis is often used with Text Mining techniques, Time Series Analysis, Algorithms, and Statistics. Kayser et al. (2014) confirm that patent analysis is frequently analyzed by text mining in futures studies, also associated with roadmapping or scenario methods. The network also shows that quantitative methods are often complemented with BDML techniques by concentrating most of these methods and techniques in the same network area. The qualitative methodology that has a slightly more intense connection with quantitative and other techniques is

Scenario. Semi-quantitative methods behave like peripheral methods, except for Roadmapping and Dephi.

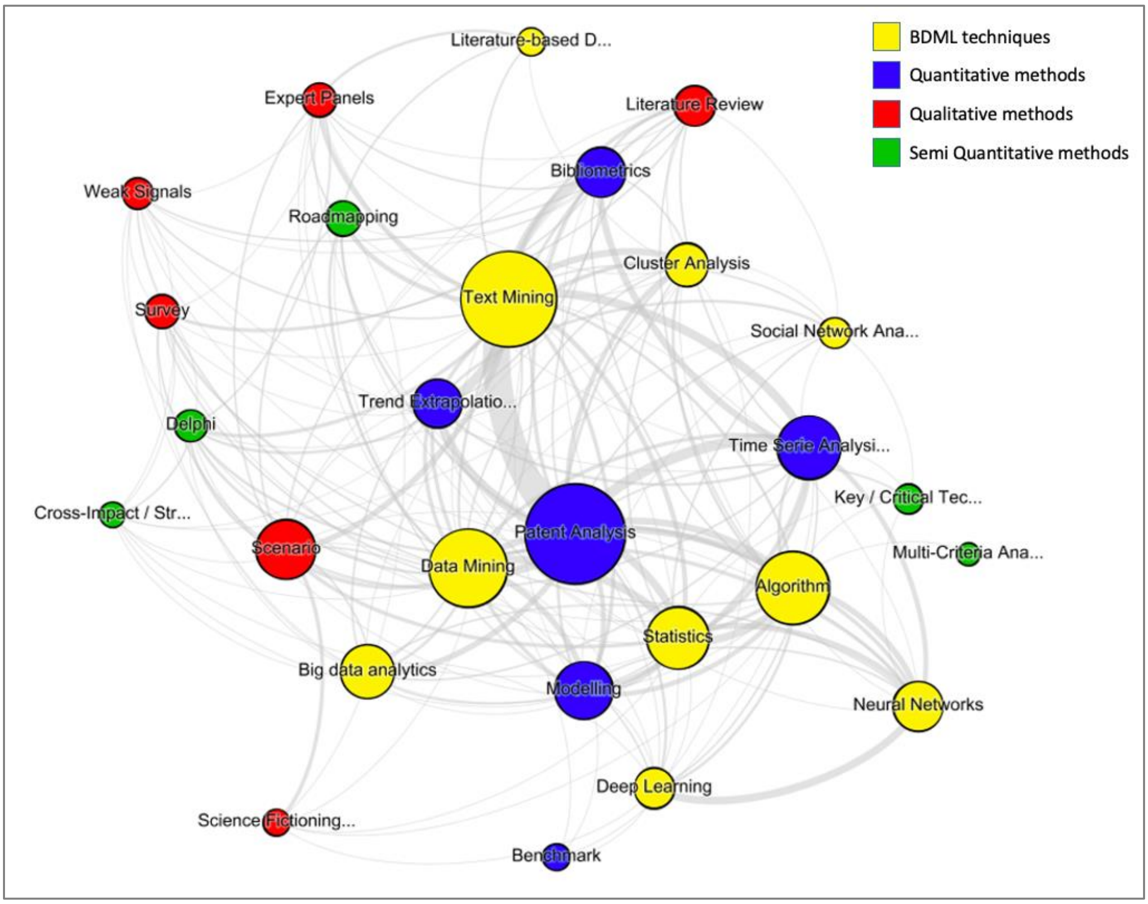


Figure 41 - Methodological approach's co-occurrence network in FSTI+BDML papers

Source: Author's elaboration

3.3.8 Keyword analysis

Figure 42 below shows the evolution of most-cited authors' keywords per year over the past ten years. The most cited keywords in the database are "Technological Forecasting" (87 papers), "Data Mining" (53 papers), "Big Data" (44 papers), "Text Mining" (44 papers), and "Machine Learning" (32 papers). One hundred seventy-nine papers cite at least one of these five keywords, corresponding to 66% of total publications. The keyword "Human" is mostly associated with futures studies in medical and diagnosis contexts.

It is possible to see an increase in Big Data, Text Mining, Machine Learning, and Artificial Intelligence mentions from 2014, which shows the recent relevance of these terms on futures studies research.

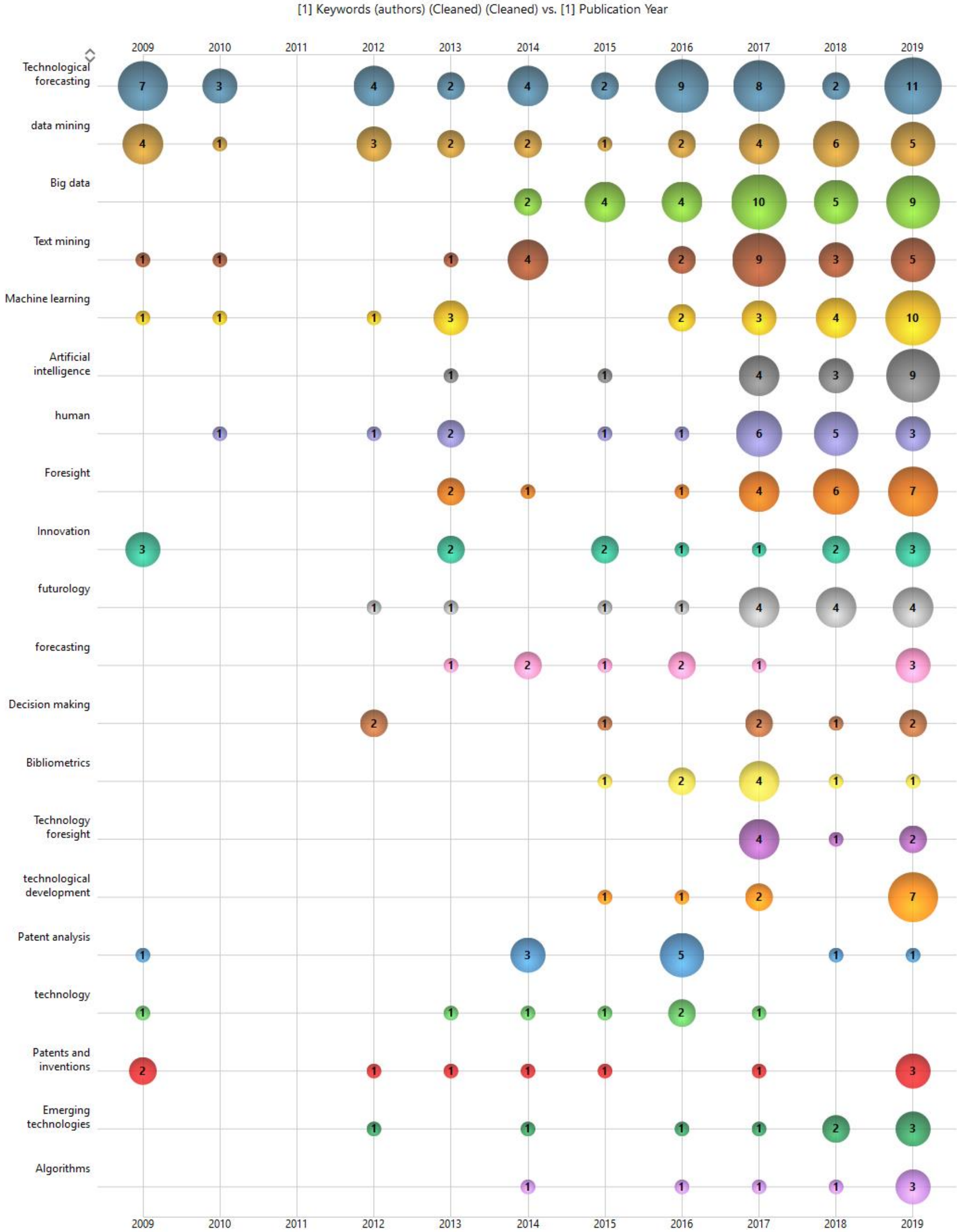


Figure 42 - Occurrence of author's keywords per year in FSTI+BDML papers

Source: Author‘s elaboration

Table 18 shows the evolution of main research approaches and topics year by year, from 2015. It also highlights the rank variation of each keyword in the year according to the previous year.

Patent analysis is the most cited keyword in 2015, followed by “Big Data” and “Technology foresight”. In 2016, there is a highlight for words that also did not appear in the 2015 ranking, such as "Artificial Intelligence", "Machine Learning", "Trend Analysis" and "Data Mining". This shows that since 2015 there has been significant growth in papers that use or explore more complex data analysis techniques for futures studies. This trend continues in the following years with the appearance of the keywords "Learning Systems" and "Deep Learning" in 2018 and "Internet of Things" in 2019.

Table 18 - Rank of author's keywords per year in FSTI+BDML papers

R	2015	Var.	2016	Var.	2017	Var.	2018	Var.	2019
1	Patent analysis	↑4	Technological forecasting	↑2	Big data	↑5	Machine learning	↑16	Technological forecasting
2	Big data	↑8	Text mining	↓-1	Technological forecasting	↑8	Artificial intelligence	↑1	Big data
3	Technology foresight	↓-1	Big data	↑69	technological development	↓-2	Big data	↓-1	Artificial intelligence
4	Information technology	↑5	Bibliometrics	↑21	human	↑3	data mining	↑1	Foresight
5	Technological forecasting	↑14	futurology	⇨0	futurology	↑3	Foresight	↓-1	data mining
6	forecasting	↑52	Artificial intelligence	↑1	Machine learning	↑3	Text mining	↑4	deep learning
7	technology	↑5	Machine learning	↑2	data mining	↓-3	human	↓-6	Machine learning
8	technological development	↑191	Trend analysis	↑27	Foresight	↓-3	futurology	↓-2	Text mining
9	Bibliometrics	↑7	data mining	↓-7	Text mining	↑157	Learning systems	↓-2	human
10	Text mining	↑10	Emerging technologies	↓-4	Artificial intelligence	↑175	deep learning	↑12	Internet of things

Source: Author‘s elaboration

The term “uncertainty” was mentioned only 10 times in the abstracts or keywords of the selected papers, mostly published by TFSC. It shows that the discussion over the conceptual basis of futures studies is overlooked in these studies, focusing on the potential of new tools to mitigate uncertainty.

Figure 43 shows that “Robotics”, “Industry 4.0”, “Nanotechnology” and “Biotechnology” are relevant research areas in the use of Big Data and Machine Learning tools in foresight activities.

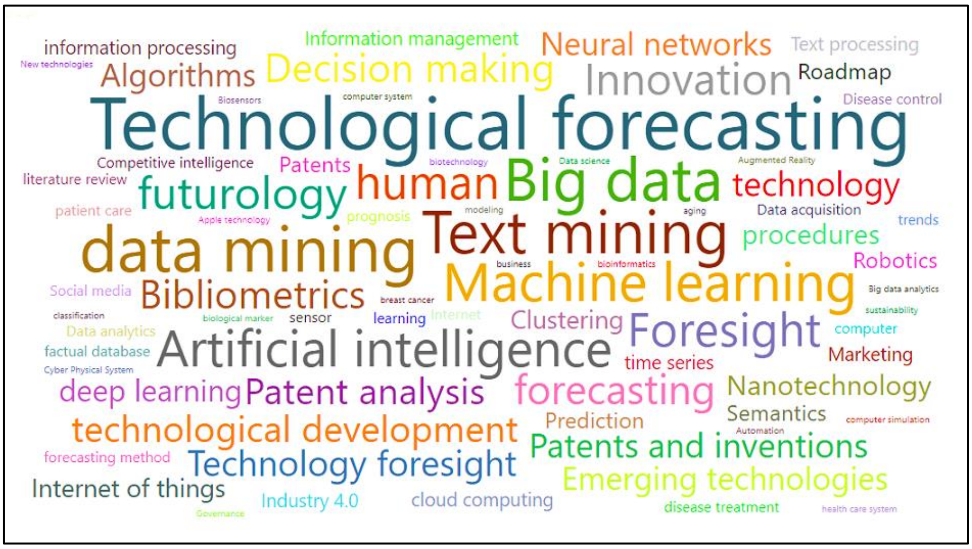


Figure 43 - Wordcloud of author's keywords in FSTI+BDML papers

Source: Author's elaboration

3.4 Conclusions

This chapter analyzed publications of futures studies that employ Big Data or Machine Learning (BDML). The idea is to get the panorama of the employed methods around the world. Bibliometrics and Social Network Analysis supported by Microsoft Excel, Vantage Point, and Gephi was employed to collect, clean, systematize, and visualize data. It was collected 270 papers related to Foresight in STI and Big Data and Machine Learning (FSTI+BDML) from two of the most important academic papers repository (Scopus and Web Of Science). To complement the analysis and compare the dynamics of publications, it was also collected 14.461 papers related to Foresight in STI (FSTI) and 227.152 papers related to Big Data and Machine Learning (BDML).

Big Data and Machine Learning tools in foresight studies are increasingly getting attention since the developments in Big Data and Machine Learning were embraced by other research fields. The USA is consolidated as the most productive, influential, and "bridge" country in collaborative research that applies these tools on foresight, mainly collaborating with South Korea and China. They consolidate the largest collaborative cluster of countries in the number of publications, with 149 publications. The European cluster counts with 68 publications and 16 countries. It is the second-largest cluster on the network, followed by the Indian cluster (blue cluster in Figure 28) with 50 publications and

12 countries. Brazilian (dark green), Pakistanis (red), and Belgian (orange) cluster together counts with 23 publications and nine countries, three countries each cluster. The collaboration between countries in this research field is concentrated in some countries that rely on the USA, the European countries, and India. The peripherical groups, in general, have internal collaboration and with the USA.

The leading journal that publishes futures studies supported by Big Data and Machine Learning tools is Technological Forecast and Social Change, one of the most important journals to discuss methodology and practice of technological forecast and futures studies. The second most productive journal is Willey Interdisciplinary Reviews – Data Mining and Knowledge Discovery that discuss data mining and knowledge discovery through an interdisciplinary approach. A journal focused on data mining on the list of top 10 productive shows that foresight is getting attention even in not specialized journals.

Most of the authors are affiliated to universities (279 authors), however research institutes, government institutions, and companies account for 84 publications (31% of total publications). The interest of the American government in futures studies dates back to the ’60s (GEORGHIOU ET AL., 2008) and is still intense with the Office of Naval Research and the Naval Surface Weapons Center, which respond to nine futures studies supported by Big Data and Machine Learning. DDL OMNI is a company that offers strategic solutions for American national programs and reinforces the presence of the American government in futures research.

Governmental institutions and national foundations are responsible for funding 76 futures studies supported by Big Data and Machine Learning (28% of total and 78% of the papers that cite the funding organization). This highlights the role and the increasing interest of public funding in futures studies, not only in the USA but also in Korea (with National Research Foundation of Korea), China (National Natural Science Foundation of China), Russia (Russian Ministry of Education) and Brazil (CNPq, CAPES, and FAPESP).

Seven clusters of authors were identified in the bibliographic coupling network (Figure 34). Patent Analysis is the primary method for purple, yellow, and blue cluster. The brown cluster is characterized by governmental-affiliated American authors, using Literature-based Discovery and Expert Panels as methods. The red cluster is focused on Big Data and Big Data Analytics research conducted mainly by Indian researchers. The various methods used in the clusters are not very dispersed, mainly quantitative methods and other techniques, following the same trend of the overall analysis.

Patent Analysis and Text Mining are the most mentioned BDML methods in this dataset. BDML techniques and future-oriented quantitative methods are frequently used together and show that BDML tools are being adapted to support traditional futures study methodologies.

The overall analysis shows that the support of Big Data and Machine Learning in foresight methodologies is an emerging field of study, still with few researches published by a small group of authors focusing on their potential of mitigate uncertainty.

Chapter 4 - The Future of Futures Studies

This chapter aims to assess the impacts of Big Data and Machine Learning (BDML) in futures studies, collecting elements to draw methodological and conceptual futures in futures-oriented activities. Chapter 4 presents the methodology, results, and discussion of an expert-consultation regarding the future influence of BDML techniques in future and foresight studies. A survey was performed with 479 foresight specialists to understand their perceptions about 12 near-future projections based on Machine Learning and Big Data's impacts. It consists of 4 sections. Section 4.1 will introduce the methodological approach. Section 4.2 presents the survey results. The following section will discuss the survey results to understand how BDML tools impact future-oriented activities. The last section presents the conclusions of the survey analysis.

4.1 Methodological approach: survey analysis

Web surveys or online surveys are widely used as a research tool in future-oriented exercises (CABRAL ET AL., 2019; KARACA & ÖNER, 2015; KELLER & VON DER GRACHT, 2014; YODA, 2011). Thus, this methodological approach fits this chapter's objectives in collecting expert opinion data to assess the future. This approach also allows access to a large number of specialists at a low cost (SAUERMANN & ROACH, 2013).

An online survey with foresight experts was performed to collect data, experiences, and expectations regarding BDML tools in foresight activities. The survey projections in this research were based on Keller & von der Gracht, (2014). The survey assessment methodology was based on Mota et al. (2020) and it includes the online survey design, the collection of e-mail addresses of authors from a bibliometric data, sending inviations and analyze the results. Participants assessed 12 projections about the influence of Big Data and Machine Learning for prospective activities having 2025 as time-horizon. Three aspects were demanded in the survey: expected probability of the projection (EP); desirability of the projection (DE); and impact on the “foresight industry” (IF). All aspects have been measured in a 5-point Likert-scale.

The projections are classified in the five foresight steps proposed by Miles (2002) and Popper (2008): Pre-Foresight, Recruitment, Generation, Action, and Renewal. The short time horizon of 2025 for the projections was defined due to the exponential technological evolution in the BDML field.

The methodological steps adopted to collect and analyze data are shown in Table 20.

Table 19 - Methodological steps for survey analysis

#	Step	Tool
1	Development of projections of Futures Studies regarding Machine Learning and Big Data	Literature review
2	Online survey design	Survey Monkey
3	Defining target respondents: experts' e-mail address collection	Vantage Point / Python
4	Send, follow and monitor survey answering	Survey Monkey
5	Analyze results and data visualization	Vantage Point / Microsoft Excel

Source: Author's elaboration

4.1.1 Development of projections

Keller & von der Gracht, (2014) developed 20 projections about the future and the influence of Information and Communication Technologies (ICT) tools in future foresight processes. These projections assessed different perspectives that involve futures studies, such as foresight practice, methodological trends, data, market, and decision-making impacts. 177 foresight experts answer their opinions of probability, desirability and impact for the future projections posed by the authors. They concluded that ICT contributes to more efficient and accurate foresight processes. Also, it supports the accessibility to information, the spread of collaboration tools, and process optimization.

Inspired by Keller & von der Gracht, (2014), our approach developed and classified 12 future projections about the implications of Big Data and Machine Learning in five steps of a generic foresight model, proposed by Miles (2002) (Figure 44).

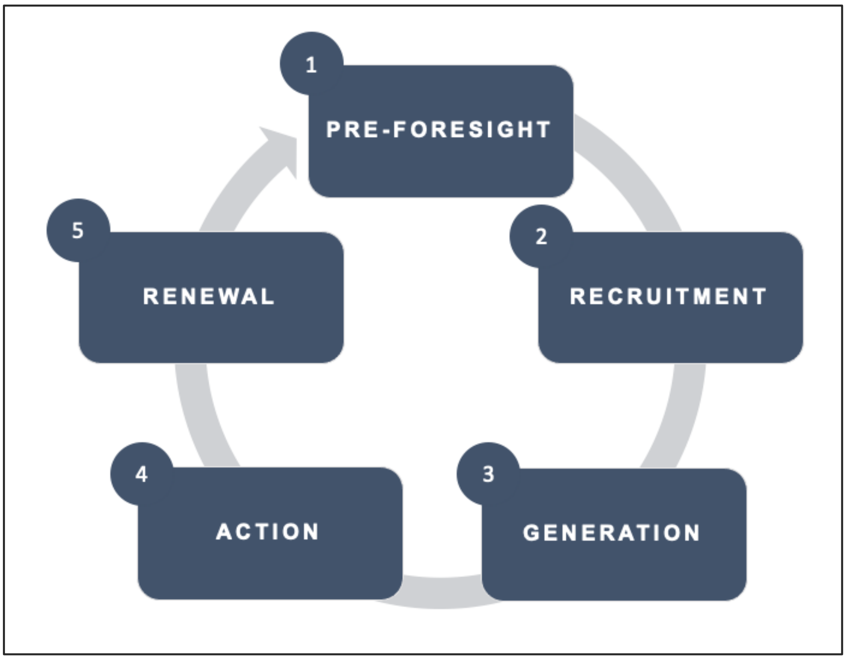


Figure 44 – Generic foresight process

Source: Author's elaboration based on Miles (2002) and Popper (2008)

All projections were conceived to be assessed in three aspects: 1) Expected Probability (EP), 2) Desirability (DE); and 3) Impact on Foresight Industry (IF). The first aspect expresses the expected probability of the projection's occurrence. The second expresses the desirability of the projection's occurrence. The last aspect expresses the impact on foresight industry if the projection occurs. The three aspect were presented on a 5-point Likert scale, with the categories: 1) very low; 2) low; 3) neither low or high; 4) high; 5) very high (LIKERT, 1932).

The 12 projections about the impact of Big Data and Machine Learning tools on futures studies are presented in Table 20 and explained as following:

Table 20 - Description of 12 projections about influence of Big Data and Machine Learning

#	Foresight step	Short title	Projection
1	Pre-Foresight	Support objectives definition	Future-oriented activities' objectives will be easily defined using Big Data and Machine Learning tools for its development in 2025.
2	Pre-Foresight	Guide methodological choice	The possibility of using Big Data and Machine Learning tools will make the methodological choice for futures studies easier in 2025.
3	Recruitment	Require analytical skills	Data Scientists or coding, analyzing, and data visualization skills will be required to develop consistent futures studies in 2025.

4	Recruitment	Support data collection	The relevant data for future-oriented activities will be less time-consuming and easily accessed using Big Data tools in 2025.
5	Recruitment	Improve data quality	The quality of future-relevant data will be significantly enhanced by the application of Big Data and Machine Learning tools or techniques in foresight studies in 2025.
6	Generation	Enhance data analysis	The quality of future-relevant data analysis will be significantly enhanced by the adoption of Big Data and Machine Learning tools and techniques in foresight studies in 2025.
7	Generation	Drive qualitative methods	Big Data and Machine Learning tools and techniques will be critical to support qualitative analysis in foresight activity in 2025.
8	Generation	Increase data manipulation	The use of Big Data and Machine Learning solutions for futures studies will increase the frequency of manipulated (biased) data in 2025.
9	Action	Support results transfer	Big Data and Machine Learning tools and techniques will increase futures studies' embeddedness to strategic planning or decision-making in 2025.
10	Action	Increase decision accuracy	Big Data and Machine Learning tools and techniques will increase the accuracy of decision-making based on foresight in 2025.
11	Renewal	Support study evaluation	Evaluating and monitoring future-oriented activities will be easily reached when using Big Data and Machine Learning tools in 2025.
12	Renewal	Support objectives' reach	Futures studies' objectives will be easily reached when Big Data and Machine Learning tools are used for its development in 2025.

Source: Author's elaboration, based on Keller & von der Gracht (2014)

- Pre-Foresight projections

Defining the objectives of a foresight study is a crucial activity that drives the whole foresight process, and it is essential to ensure that the outcomes can support future decisions. Typical objectives of foresight activities are: 1) fostering science, technology, and innovation (STI) cooperation and network; 2) orienting policy formulation and decisions; 3) recognizing key barriers and drivers of STI; 4) encouraging strategic and future thinking; 5) supporting STI strategy- and priority-setting; 6) identifying research/investment opportunities; 7) generating visions and images of the future; 8) triggering actions to promoting public debate (MILES AT AL., 2008). Pre-analysis of large databases may support objectives' definition in future-oriented activities.

The methodology or the combination of methodologies that will drive the foresight activity should be decided to meet the objectives. Foresight's methodological approach is affected by available resources: budget, team expertise, political support, technological and physical infrastructure, and available time (POPPER, 2009). The availability of Big Data and Machine Learning tools can guide the methodological approach for future-oriented activities and affect the definition of the objectives.

So, the first two projections related to the Pre-Foresight step are presented:

- Projection 1: Future-oriented activities' objectives will be easily defined using Big Data and Machine Learning tools for its development in 2025 (Support objectives definition).
- Projection 2: The possibility of using Big Data and Machine Learning tools will make the methodological choice for futures studies easier in 2025 (Drive methodological choice).

- Recruitment projections

The Recruitment step in the foresight process is associated with choosing participants (both experts and non-experts) to assess, and databases necessary for the activity, which means collecting all necessary information to develop prospective processes.

Gary & von der Gracht (2015) studied the future of foresight team through a global Delphi survey. In this study, they developed three possible scenarios for 2030 regarding three different pathways for professional foresight: 1) Assimilation: foresight capabilities and tools will be absorbed by other professions, and professional futurists will no longer be recognized as a professional identity by the market; 2) Academicization: foresight support for futures work is mainly offered by universities, that concentrates professional training and specialists to conduct future-oriented studies; and 3) Certification: foresight profession has self-organized and formalized by commoditizing foresight tools and offering "futurists" certifications. In a digital revolution context, some foresight team competencies are essential in any of the three future scenarios presented above. Abilities as coding, mathematical, statistical, and data visualization skills, as well as domain knowledge, will be needed in any futures study supported by BDML tools (TETLOCK AND GARDNER, 2015).

Choosing the right team is imperative to extract useful information and knowledge from a large volume of data is fundamental for any application and future actions (WU ET AL., 2014). Data mining techniques have developed a significant role in Big Data analytics in identifying and collect relevant data from several heterogeneous sources (KAYSER & BIERWISCH, 2016; KAYSER & BLIND, 2016). However, besides the collection, data quality is also very relevant for a successful study. Some authors (Hazen et al., 2014; Lee et al., 2002, 2004; Wang et al., 1995) define data quality in two dimensions: intrinsic and contextual. Intrinsic dimension is related to objectives and native attributes to the data, like accuracy (data are free of errors?), timeliness (data is up to date?), consistency (data are presented in the same format?), and completeness (data has necessary missing values?). Contextual dimension is related to subjective and judgmental attributes that are dependent on the context in which data are used, like relevancy, value-added, quantity, believability, accessibility, and reputation of the data. Data quality is a prominent issue in Big Data and Big Data Analytics.

Thus, related to the Recruitment step, the following projections were developed:

- Projection 3: Data Scientists or people with coding, analyzing, and data visualization skills will be required to develop consistent futures studies in 2025 (Require analytical skills).
- Projection 4: The relevant data for future-oriented activities will be less time-consuming and easily accessed using Big Data tools in 2025 (Support data collection).
- Projection 5: The quality of future-relevant data will be significantly enhanced by the application of Big Data and Machine Learning tools or techniques in foresight studies in 2025 (Improve data quality).

- Generation projections

From a sample of 130 case studies from 15 countries, Popper (2008) found an average of five to six methods used to compose the methodological framework of future-oriented activities. The author claims that there is no “ideal” or "best" combination of methods for a foresight study, but, in general, they consist of a mix of quantitative, semi-quantitative, qualitative, creativity-based, evidence-based, interaction-based, or expertise-based methods. Data-driven futures studies rely not only on a big volume of data or simulations but also on qualitative methodologies and data interpretation to better comprehend future paths. Coates (2010) affirms that only data is not enough to predict the future, and it is necessary to stimulate creativity and interpretative skills to get real information from this source. Data availability may become a non-issue to futures studies, but a wise interpretation and analysis may challenge new research.

However, when a topic becomes relevant as a future trend, controversies are developed, increasing data scrutiny. Also, malicious attacks can happen, and data can be manipulated or biased to prove a different perspective. The reliance on cloud-computing and cloud-databases can expose foresight practitioners to security risks of their data (KAUFMAN, 2009). Data security and manipulation is already a concerning issue in the discussion about fake news and deep fakes.

In Generation step, the following projections are proposed:

- Projection 6: The quality of future-relevant data analysis will be significantly enhanced by the adoption of Big Data and Machine Learning tools and techniques in foresight studies in 2025 (Enhance data analysis).
- Projection 7: Big Data and Machine Learning tools and techniques will be critical to support qualitative analysis in foresight activity in 2025 (Drive qualitative methods).
- Projection 8: The use of Big Data and Machine Learning solutions for futures studies will increase the frequency of manipulated (biased) data in 2025 (Increase manipulated data).

- Action projections

Some approaches are described in the literature to formulate actions, strategies, and policies based on futures studies outcomes (AMANATIDOU, 2008; KILDIENE ET AL., 2011; SALO ET AL., 2003; ZAVADSKAS & TURSKIS, 2011). The results of a prospective study are not necessarily ready for strategy or policy planning. Different methods are used to transfer the results of a prospective activity to strategies, policies, or any decision-making process. Among them, Campanella & Ribeiro (2011) formulated a multi-criteria framework for dynamic circumstances. All data employed in futures studies' development can provide a better understand and facilitate transferring foresight results into strategies or policies.

Decision-making is the final target of any futures study. The decision theory by Shackle (1969) associates the concept of uncertainty and potential surprise, which means high uncertainty about a phenomenon is associated with high potential surprise and, at its limit, drives to the maximum surprise and the perfect impossibility of the phenomenon to occur. On the other hand, if the degree of uncertainty is low, the potential surprise is also low and, at its limit, drives to null surprise and the perfect possibility of the phenomenon to occur. Support of Big Data and Machine Learning can make potential surprise decrease by using lots of information and making decision-maker more confident in terms of decision accuracy.

These assumptions drive the next two projections, presented below:

- Projection 9: Big Data and Machine Learning tools and techniques will increase futures studies' embeddedness to strategic planning or decision-making in 2025 (Support results transfer).
- Projection 10: Big Data and Machine Learning tools and techniques will increase the accuracy of decision-making based on foresight in 2025 (Improve decision accuracy).

- Renewal projections

The whole foresight process proposed by Miles (2002) is a cyclic process that includes an ex-post evaluation. Monitoring and evaluate the foresight process become a relevant topic and is addressed by some researchers (GEORGHIOU & KEENAN, 2006; MAZURKIEWICZ ET AL., 2013). Different metrics are proposed to evaluate foresight depending on the study's objectives (CALOF & SMITH, 2012). Big Data and Machine Learning tools may be incorporated into this process to maximize the efficiency of evaluating and monitoring foresight.

In evaluating futures studies, the central question is that the proposed general objectives studies were achieved. Futures studies can have several different objectives, and to achieve the goals, a data collection process, a critical review and an in-depth analysis may be required for it.

The last two projections, related to the Renewal step in the foresight process are:

- Projection 11: Evaluating and monitoring future-oriented activities will be easily reached when using Big Data and Machine Learning tools in 2025 (Support study evaluation).
- Projection 12: Futures studies' objectives will be easily reached when Big Data and Machine Learning tools are used for its development in 2025 (Support objectives' reach).

4.1.2 Online survey design

Supported by the Survey Monkey platform, the online survey was structured in four parts: 1) Introduction; 2) Instructions; 3) Projections; and 4) Demographic Questions.

In the first part, the introduction presented a succinct abstract, the research's objectives, the participant's contribution, and confidentiality and privacy terms. Regarding confidentiality and privacy, this research has been approved by the Ethics Committee of the University of Campinas by the number 25592219.6.0000.8142 as notified at the Free and Informed Consent Form (See Annex 1). Finally, the institutions that supported this research were mentioned: São Paulo Research Foundation (FAPESP), the project System Innovation: Organizational Strategy, Research, and Innovation Policy Governance (InSySPo), the Science and Technology Policy Department (DPCT), and the University of Campinas (Unicamp).

The main concepts were explained in the instructions part, as well as the structure of the questions and projections. Two preliminary questions were asked in this part of the survey: "Do you have experience as a futurist or practitioner of futures studies supported by ICT, Big Data, or Machine Learning techniques?" and "Do you believe that Big Data and Machine Learning tools and techniques will reduce uncertainty related to futures studies in 2025?". These questions aim to understand the experts' conversance with the subject.

The projections were presented in groups according to the foresight steps and each participant were asked to evaluate, on a 5-point Likert scale, the above referred three aspects: Expected Probability (EP), Desirability (DE), and Impact on Foresight Industry (IF).

The last part of the survey is composed of demographic questions: "What kind of institution are you affiliated with?" (University, consultancy, research center or institution, company, government, other); "How many years of experience do you have in future-oriented activities?" (Under 1 year, 1-5 years, 5-10 years, 10-15 years, 15+ years); "What is your region?" (Africa, Asia, Europe, Latin America and the Caribbean, Northern America, Oceania).

The survey questionnaire is presented in Annex 2.

4.1.3 Defining target respondents: foresight experts

Following the survey guidelines of Mota et al. (2020), the experts were defined according to authors identified on the 14.461 publications of “Foresight in STI” (FSTI), from the bibliometric analysis in chapter 3 of this work.

The total number of authors in the FSTI database is 26.654. For 7.753 of these authors, in addition to the e-mail address, it was possible to extract more information such as author name, title, and year of publication. Mentioning the name and other information about the expert in the invitation mail for the survey is essential to raise the response rate (SAUERMANN & ROACH, 2013). The survey was also sent to the World Futures Studies Federation (WSFS)[8] and Millennium Project (MP)[9] members, through group mail. The e-mail address of 18.901 authors was not mentioned, and they were not included in this survey participants.

In total, 7.753 foresight experts were directly invited through e-mail via the Survey Monkey platform.

4.1.4 Send, follow and monitor survey answering

The survey was available to receive answers from October 19th to October 26th, 2020. The first invitation was sent to the 7.753 foresight experts on

[8] WFSF is a UNESCO and UN consultative partner with members in over 60 countries. It is a forum for discussing ideas, visions and plans for alternative futures (more information access https://wfsf.org/)
[9] Millenium Project is a global participatory think tank with 67 nodes around the world (more information access http://www.millennium-project.org/)

October 19th, and two reminder e-mails were sent on October 22nd and October 24th via the Survey Monkey platform. On October 19th, the survey invitation was sent to WSFS and MP e-mail groups.

4.1.5 Analyze results and data visualization

Descriptive statistics were employed to analyze the variables related to the projections. Frequency, mean, median, and mode are examples of measures used to understand the aggregated data.

Finally, the software Microsoft Excel and Vantage Point were used to create tables and charts for data visualization.

4.2 Survey results

A total of 661 researchers and foresight experts have participated in this study. However, complete answers correspond to 72,5% of the total (479 experts) and were considered valid for analysis. Considering that the foresight experts' population is 26.654 (based on the total of authors in the FSTI database), for achieving 95% of confidence level and 5% of margin of error in a normal distribution, the minimum sample size is 379 complete responses (FLOREY, 1993). Although it was not possible to compare population and sample characteristics due to data limitation, the minimum sample size has achieved.

The Figure 45 below shows the sample demographic profile of the experts.

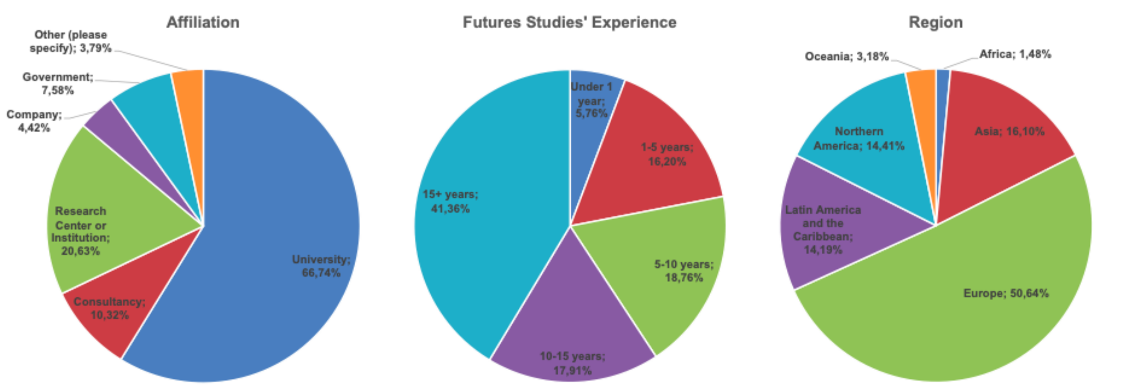


Figure 45 - Experts demographic profile

Source: Author's elaboration

Most of consulted experts are from universities (317, 66,2%), followed by experts affiliated with research institutions. Almost 60% of experts (278) has shown an extensive background with futures studies, with more than 10 years of experience.

Regarding the geographic region, more than half of consulted experts are from Europe (239, 49,9%), followed by Asia (76, 15,9%), Northern America (68, 14,2%), and Latin America and the Caribbean (67, 13,9%).

4.2.1 Overview

More than half of the experts (263) have reported practical experience with futures studies supported by Machine Learning and Big Data tools and techniques. Also, 60% (287) have declared that BDML tools and techniques could reduce uncertainty related to futures studies (Figure 46).

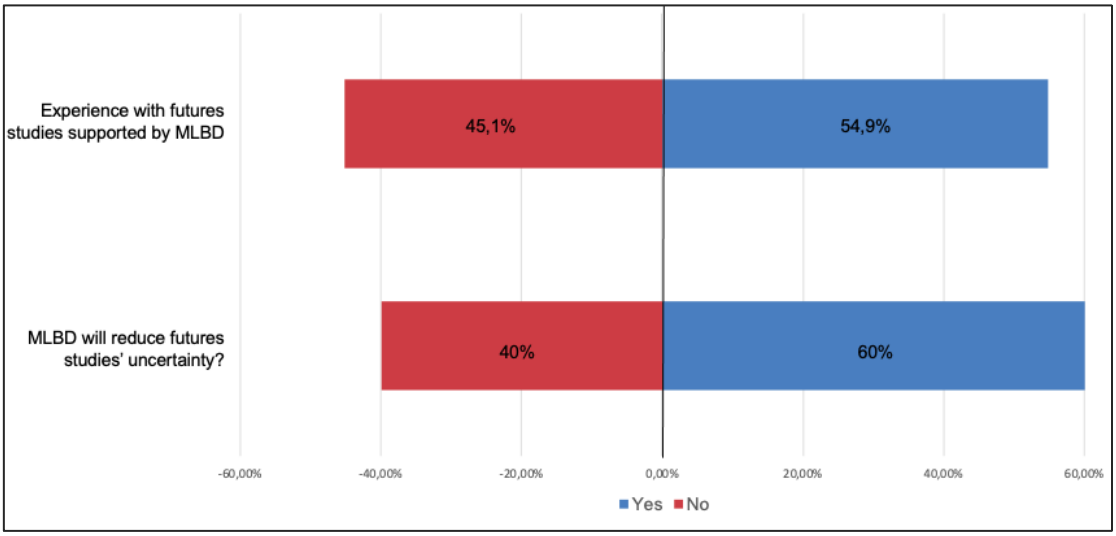


Figure 46 - Experience and implications of BDML tools

Source: Author‘s elaboration

Table 21 presents an overview of the consulted experts’ opinions. The median and mean values are shown for Expected Probability (EP), Desirability (DE), and Impact in the Foresight Industry (IF), for each of the projections.

Table 21 - Median and mean values of survey projections

Step	#	Projection	EP		DE		IF	
			Median	Mean	Median	Mean	Median	Mean
Pre-Foresight	1	Support objectives definition	3	2,89	3	3,29	4	3,36
	2	Guide methodological choice	3	3,23	4	3,43	4	3,40
Recruitment	3	Require analytical skills	4	3,69	4	3,65	4	3,69
	4	Support data collection	4	3,44	4	3,83	4	3,66
	5	Improve data quality	4	3,37	4	3,76	4	3,62
Generation	6	Enhance data analysis	4	3,44	4	3,74	4	3,60
	7	Drive qualitative methods	4	3,40	4	3,55	4	3,54
	8	Increase data manipulation	4	3,51	2	2,41	3	3,42
Action	9	Support results transfer	4	3,55	4	3,61	4	3,65
	10	Increase decision accuracy	3	3,24	4	3,63	4	3,52
Renewal	11	Support study evaluation	3	3,15	4	3,51	3	3,37
	12	Support objectives' reach	3	2,88	4	3,34	3	3,16

Source: Author's elaboration

The projections related to the foresight objectives (P1 – support objective definition and P12 – support objectives' reach) were those with the lowest aggregate expected probability values. On the other hand, P3 (require analytical skills) and P9 (support results transfer) are considered to have a higher probability of occurrence.

Regarding desirability, the least desirable projections are P8 (increase data manipulation) and P1 (support objectives definition). Experts recognize data manipulation as a limitation to foresight data support, which means it decreases data reliability. Also, they considered that it is not desirable the influence of BDML tools in objectives definition for futures studies. In general, all other projections had a median value considered to be highly desirable.

The projections with the most significant impact on the foresight industry are P3 (require analytical skill) and P4 (support data collection). In general, almost all projections related to Recruitment, Generation, and Action show a high impact in the foresight industry, except for P8 (Increase data manipulation). Both Renewal projections show no significant impact on the foresight industry.

Figure 47 shows the distribution of answers to each question of the survey.

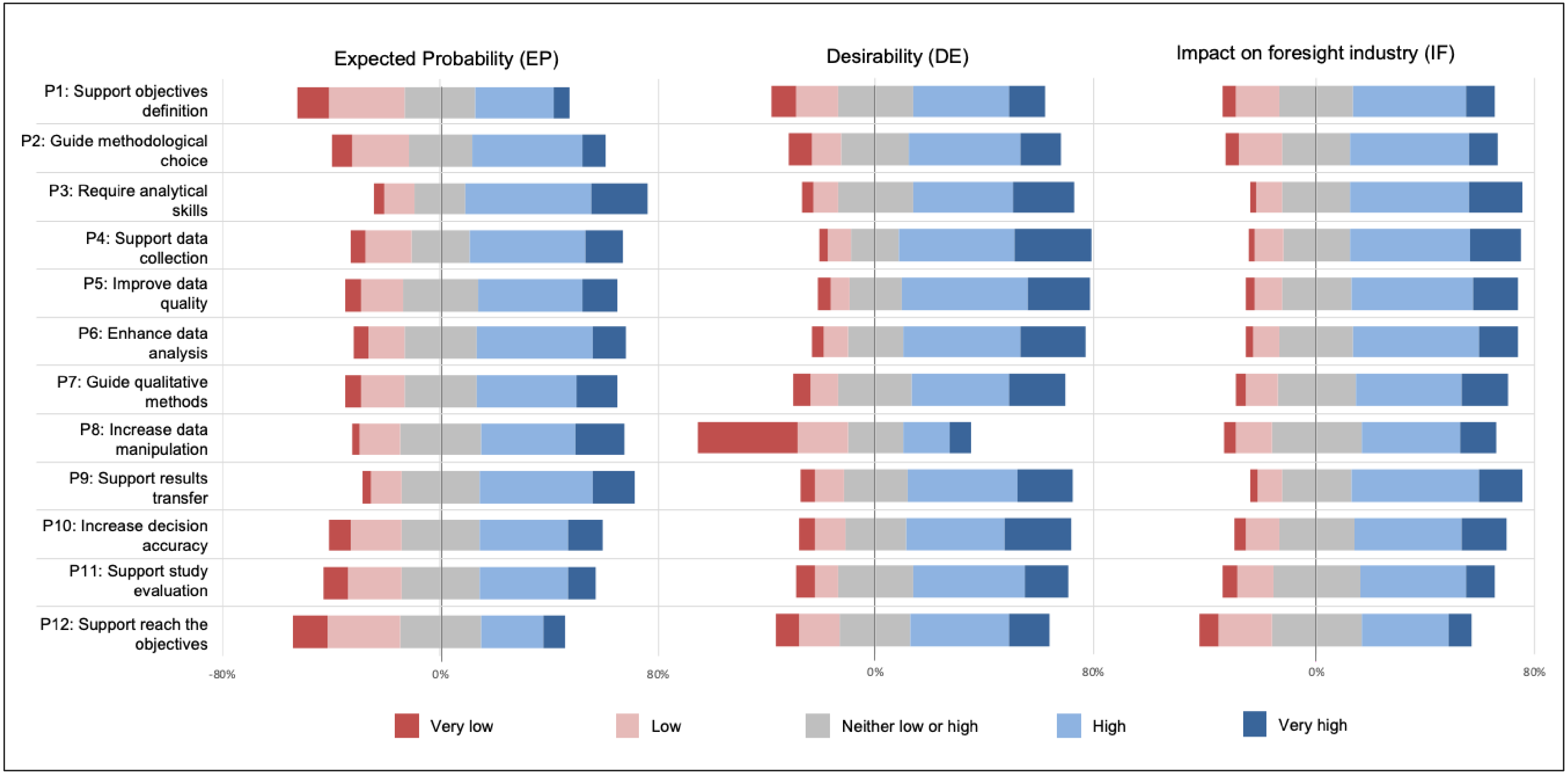


Figure 47 - Distribution of answers in EP, DE and IF dimensions

Source: Author's elaboration

In Figure 48, the projections were plotted with the mean value of IF on the y-axis, the mean value of EP on the x-axis, and the mean value of DE as a proportion of the bubbles' size.

Projections were divided into four different groups based on the expected probability of occurrence. The first group consists of the projections P3 (require analytical skills), P9 (support results transfer) e P8 (increase data manipulation), with a high probability of occurrence (mean values are greater than 3,5). The projections P4 (support data collection), P5 (improve data quality), P6 (enhance data analysis), and P7 (drive qualitative methods) were classified with a medium-high probability of occurring (mean values are greater than 3,25). The projections P2 (guide methodological choice), P10 (increase decision accuracy), and P11 (support study evaluation) were classified with a medium-low probability of occurrence (mean values are greater than 3). Finally, the projections P1 (support objectives definition) and P12 (support objectives' reach) were classified with a low probability of occurring (mean values are less than 3).

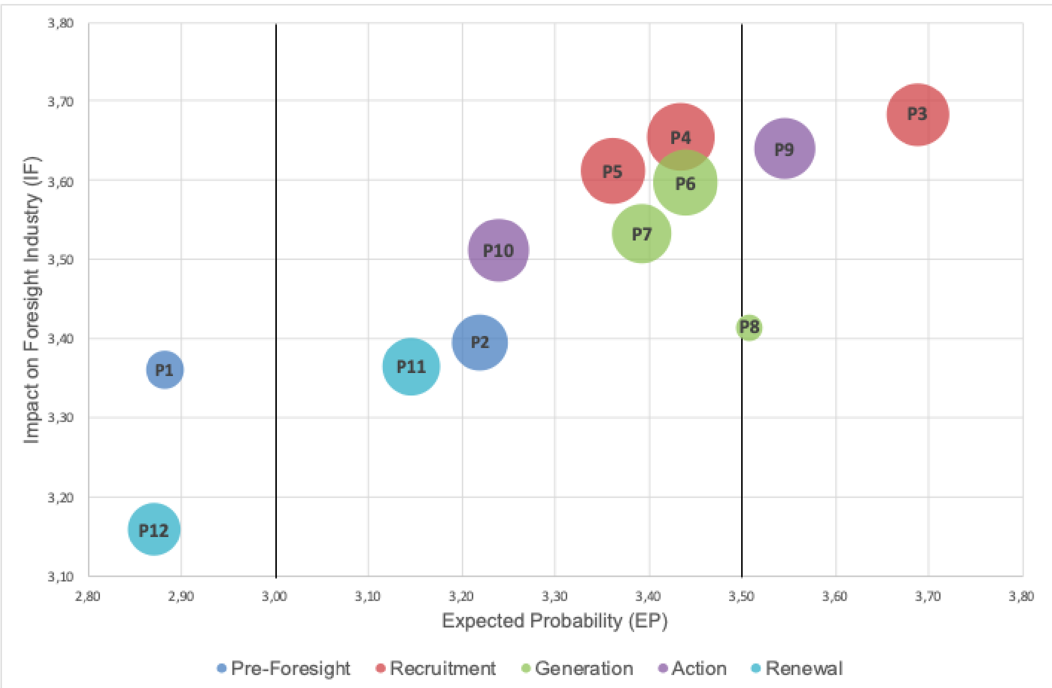


Figure 48 - Scatterplot of projections according to probability, impact and desirability for 2025

Source: Author's elaboration

4.2.2 Pre-foresight projections

Two projections are analyzed in the Pre-Foresight phase:

- Projection 1: Support objectives definition
- Projection 2: Guide methodological choice

The first projection states that BDML tools will support foresight process since the very first step, objectives definition. Table 22 shows the distribution of the answers refer to the three aspects of projection 1, namely expected probability (EP), desirability (DE), and impact on foresight industry (IF).

Table 22 - Distribution of answers for P1 – Support objective definition

P1	very low	low	neither low or high	high	very high	Total
Expected Probability	56 (11,81%)	132 (27,85%)	**124 (26,16%)**	136 (28,69%)	26 (5,49%)	**474**
Desirability	41 (8,63%)	74 (15,58%)	**130 (27,37%)**	167 (35,16%)	63 (13,26%)	**475**
Impact on Foresight Indutry	24 (5,05%)	76 (16%)	129 (27,16%)	**197 (41,47%)**	49 (10,32%)	**475**
					Answered	**476**
					Skipped	**3**

Source: Author's elaboration

The distribution of the expected probability of projection 1 indicates that the experts do not foresee (low expectation) that Machine Learning and Big Data will influence the objectives' definition. The point "3 - neither high or low" is the median of this expected probability and desirability. For impact on the foresight industry, the median value is "4 – high", according to more than 40% of the respondents. Although the objectives' definition is not seen as influenced by Big Data and Machine Learning tools, it will have a certain impact on the foresight industry if it occurs.

Projection 2 affirms that BDML tools will direct the methodological choice for the study. The distribution of answers to that projection is presented in Table 23.

Table 23 - Distribution of answers for P2 – Guide methodological choice

P2	very low	low	neither low or high	high	very high	Total
Expected Probability	35 (7,34%)	97 (20,34%)	**113 (23,69%)**	192 (40,25%)	40 (8,39%)	**477**
Desirability	41 (8,6%)	51 (10,69%)	120 (25,16%)	**195 (40,88%)**	70 (14,68%)	**477**
Impact on Foresight Indutry	22 (4,62%)	78 (16,39%)	118 (24,79%)	**206 (43,28%)**	52 (10,92%)	**476**
					Answered	**477**
					Skipped	**2**

Source: Author's elaboration

Experts see the expected probability that Machine Learning and Big Data tools support futures studies' methodological choice as "3 – neither high or low" (median point) at the Likert scale, and it was classified as medium-low expectations. However, the experts considered high desirability (more than 40% of experts) and high impact on the foresight industry (more than 43% of experts) in this projection.

4.2.3 Recruitment projections

Three projections are analyzed in Recruitment phase:

- <u>Projection 3:</u> Require analytical skills
- <u>Projection 4:</u> Support data collection
- <u>Projection 5:</u> Improve data quality

Projection 3 asserts that it will be a necessary analytical skill to deal with data in future-oriented activities. Table 24 shows the distribution of P3.

Table 24 - Distribution of answers for P3 – Require analytical skills

P3	very low	low	neither low or high	high	very high	Total
Expected Probability	17 (3,58%)	53 (11,16%)	89 (18,74%)	**218 (45,89%)**	98 (20,63%)	**475**
Desirability	19 (4%)	43 (9,05%)	132 (27,79%)	**175 (36,84%)**	106 (22,32%)	**475**
Impact on Foresight Indutry	12 (2,53%)	45 (9,49%)	118 (24,89%)	**205 (43,25%)**	94 (19,83%)	**474**
					Answered	**475**
					Skipped	**4**

Source: Author's elaboration

The experts considered that it is highly expected that a foresight team will include data experts or analytical competencies to develop futures studies. Almost 67% of assessed researchers considered the expected probability of P3 high or very high and led this projection to the top projection in terms of the expected occurrence. The desirability and the impact on the foresight industry were also expressed as high or very high for most respondents.

Projection 4 affirms that the BDML tool will support data collection, and it will be less time-consuming. Table 25 shows the distribution of answers for P4.

Table 25 – Distribution of answers for P4 – Support data collection

P4	very low	low	neither low or high	high	very high	Total
Expected Probability	25 (5,25%)	80 (16,81%)	102 (21,43%)	**201 (42,23%)**	68 (14,29%)	**476**
Desirability	17 (3,58%)	40 (8,42%)	84 (17,68%)	**201 (42,32%)**	133 (28%)	**475**
Impact on Foresight Indutry	12 (2,52%)	50 (10,5%)	115 (24,16%)	**212 (44,54%)**	87 (18,28%)	**476**
					Answered	**476**
					Skipped	**3**

Source: Author‘s elaboration

The expectation that data collection supported by BDML tools will be less time-consuming and easily accessed is classified as medium-high due to the mean value of the answers in P4. It is very high desirable (highest desirability score among all 12 projections) that BDML supports data collection. The impact on the foresight industry if it occurs is also high.

Projection 5 states that the BDML tool will increase the quality of future-relevant data. Table 26 shows the distribution of answers in this projection.

Table 26 - Distribution of answers for P5 – Improve data quality

P5	very low	low	neither low or high	high	very high	Total
Expected Probability	30 (6,28%)	72 (15,06%)	132 (27,62%)	**183 (38,28%)**	61 (12,76%)	**478**
Desirability	23 (4,82%)	33 (6,92%)	90 (18,87%)	**221 (46,33%)**	110 (23,06%)	**477**
Impact on Foresight Indutry	16 (3,36%)	47 (9,87%)	121 (25,42%)	**214 (44,96%)**	78 (16,39%)	**476**
					Answered	**478**
					Skipped	**1**

Source: Author‘s elaboration

The experts have pointed to a medium-high value of expectation that BDML tools will enhance the data quality. This projection's desirability has the second-highest score, thus increasing data quality through BDML tools is seen as very highly desirable. The impact on foresight industry if this projection occurs is also high.

4.2.4 Generation projections

Three projections are analyzed in Generation phase:

- Projection 6: Enhance data analysis
- Projection 7: Drive qualitative methods
- Projection 8: Increase manipulated data

Projection 6 states that BDML tool will enhance the quality of data analysis in a foresight study. The distribution of P6 is shown in the Table 27.

Table 27 - Distribution of answers for P6 – Enhance data analysis

P6	very low	low	neither low or high	high	very high	Total
Expected Probability	24 (5,05%)	64 (13,47%)	125 (26,32%)	**203 (42,74%)**	59 (12,42%)	**475**
Desirability	20 (4,24%)	41 (8,69%)	96 (20,34%)	**202 (42,8%)**	113 (23,94%)	**472**
Impact on Foresight Indutry	13 (2,75%)	46 (9,75%)	127 (26,91%)	**218 (46,19%)**	68 (14,41%)	**472**
					Answered	**476**
					Skipped	**3**

Source: Author's elaboration

The expected probability of BDML enhance data analysis for futures studies is classified as medium-high. The desirability and the impact on the foresight industry for data analysis were considered high by the participants.

Projection 7 states that BDML outputs will guide the methodological steps in a futures study process. The distribution is shown in Table 28.

Table 28 - Distribution of answers for P7 – Drive qualitative methods

P7	very low	low	neither low or high	high	very high	Total
Expected Probability	28 (5,92%)	74 (15,64%)	126 (26,64%)	**174 (36,79%)**	71 (15,01%)	**473**
Desirability	29 (6,14%)	50 (10,59%)	127 (26,91%)	**167 (35,38%)**	99 (20,97%)	**472**
Impact on Foresight Indutry	18 (3,81%)	56 (11,86%)	134 (28,39%)	**185 (39,19%)**	79 (16,74%)	**472**
					Answered	**475**
					Skipped	**4**

Source: Author's elaboration

More than half of the consulted researchers (51,8%) pointed out that it is high or very high the expected probability of BDML tools drive the qualitative steps in a foresight study. Additionally, around 55% of the survey participants consider high or very high the desirability and the impact on the foresight industry of this projection.

Projection 8 affirms that BDML tools applied in futures studies will increase the frequency of manipulated data. The distribution of P6 is shown in Table 29.

Table 29 - Distribution of answers for P8 – Increase data manipulation

P8	very low	low	neither low or high	high	very high	Total
Expected Probability	13 (2,77%)	69 (14,68%)	140 (29,79%)	**162 (34,47%)**	86 (18,3%)	**470**
Desirability	172 (36,75%)	**86 (18,38%)**	93 (19,87%)	81 (17,31%)	36 (7,69%)	**468**
Impact on Foresight Indutry	21 (4,49%)	60 (12,82%)	**155 (33,12%)**	168 (35,9%)	64 (13,68%)	**468**
					Answered	**471**
					Skipped	**8**

Source: Author's elaboration

The frequency of manipulated or biased data employed in futures studies is highly expected for more than 52% of participants (high or very high expected probability). However, Projection 8 has achieved the lowest value in desirability, showing a significant limitation of BDML tools for future-oriented activities (more than 55% of respondents choose low or very low desirability). Researchers expect a medium impact on the foresight industry.

4.2.5 Action projections

Two projections are analyzed in Action phase:

- Projection 9: Support results transfer
- Projection 10: Improve decision accuracy

Projection 10 affirms that BDML tool will increase futures studies' embeddedness to the decision-making, which means it will support results transfer. Table 30 shows the distribution of answers in the survey.

Table 30 - Distribution of answers for P9 – Support results transfer

P9	very low	low	neither low or high	high	very high	Total
Expected Probability	15 (3,17%)	54 (11,42%)	134 (28,33%)	**198 (41,86%)**	72 (15,22%)	**473**
Desirability	25 (5,27%)	50 (10,55%)	110 (23,21%)	**192 (40,51%)**	97 (20,46%)	**474**
Impact on Foresight Indutry	13 (2,77%)	43 (9,15%)	119 (25,32%)	**220 (46,81%)**	75 (15,96%)	**470**
					Answered	**474**
					Skipped	**5**

Source: Author's elaboration

Futures studies' embeddedness in strategic planning or decision-making is highly expected with the advent of BDML tools. The expectation score of this projection is the second highest. Futures studies supported by BDML should be used in strategic decision making. Its impact on the foresight industry is also high.

Projection 10 states that BDML tools will increase the accuracy of foresight-based decision-making. The distribution of the answers to this projection is shown in Table 31.

Table 31 - Distribution of answers for P10 – Increase decision accuracy

P10	very low	low	neither low or high	high	very high	Total
Expected Probability	36 (7,63%)	87 (18,43%)	**137 (29,03%)**	152 (32,2%)	60 (12,71%)	**472**
Desirability	29 (6,13%)	51 (10,78%)	106 (22,41%)	**171 (36,15%)**	116 (24,52%)	**473**
Impact on Foresight Indutry	21 (4,48%)	56 (11,94%)	131 (27,93%)	**184 (39,23%)**	77 (16,42%)	**469**
					Answered	**473**
					Skipped	**6**

Source: Author's elaboration

The increase of confidence and accuracy of decision-making based on foresight supported by BDML has medium-low expectation to happen, with a median value as “3.- neither low or high”. However, the desirability of this accuracy’ gain is high, as well as the impact of this future projection on the foresight industry.

4.2.6 Renewal projections

Two projections are analyzed in Renewal phase:

- Projection 11: Support study evaluation
- Projection 12: Support reach the objectives

Projection 11 states that BDML tools will support monitoring and evaluation of future-oriented activities. The distribution is shown in Table 32.

Table 32 - Distribution of answers for P11 – Support study evaluation

P11	very low	low	neither low or high	high	very high	Total
Expected Probability	43 (9,11%)	93 (19,7%)	**135 (28,6%)**	154 (32,63%)	47 (9,96%)	**472**
Desirability	33 (7,01%)	40 (8,49%)	129 (27,39%)	**194 (41,19%)**	75 (15,92%)	**471**
Impact on Foresight Indutry	25 (5,35%)	62 (13,28%)	**148 (31,69%)**	182 (38,97%)	50 (10,71%)	**467**
					Answered	**472**
					Skipped	**7**

Source: Author‘s elaboration

The experts have pointed to a medium-low expectation that BDML tools will support the evaluation and monitoring of future-oriented activities. As well as the expected probability, the impact on the foresight industry of this projection is also low. Nonetheless, the desirability is considered “4 – high”, as a median.

Projection 12 states that BDML tools will support to reach foresight objectives. The distributions of the answers to P12 are shown in Table 33

Table 33 - Distribution of answers for P12 – Support objectives' reach

P12	very low	low	neither low or high	high	very high	Total
Expected Probability	60 (12,74%)	125 (26,54%)	**140 (29,72%)**	108 (22,93%)	38 (8,07%)	**471**
Desirability	39 (8,32%)	70 (14,93%)	123 (26,23%)	**167 (35,61%)**	70 (14,93%)	**469**
Impact on Foresight Indutry	33 (7,08%)	90 (19,31%)	**154 (33,05%)**	148 (31,76%)	41 (8,8%)	**466**
					Answered	**471**
					Skipped	**8**

Source: Author's elaboration

The less expected projection in this survey is P12. For almost 70% of the researchers, the expected probability that BDML tools will support reaching the study's objectives are "very low", "low", or "neither low or high". The overall impact of P12 on the foresight industry is considered "3 – neither low or high". However, the desirability is on the same level as most of the projections, "4 – high".

4.2.7 Qualitative comments on the survey

Table 34 presents selected commentaries sent by respondents. The selection of the commentaries was based on the relevance of the opinion to the discussion of this analysis.

Table 34 - Selected comments per foresight step

Step	# Comments	Selected Comments
Pre-Foresight	89	*"ML is useful for some tasks, such as scanning and clustering data. But that does not address 'objectives' nor will it reduce uncertainty. Given that all agents will be using such technologies it could increase uncertainty and reduce trust"* *"Projection 1: I see potential on new Big Data and Machine Learning tools, not only for defining Future-oriented activities' objectives, but to provide solid evidence of change based on data. Projection 2: Although Big Data and Machine Learning tools will be proven powerful tools, I expect a higher adoption rate by new researchers. Many researchers are expected to use the current tools that they are familiar with. In a longer-term, beyond 2025, the impact of Big Data and Machine Learning tools tends to rise."*

		"Big Data and Machine learning add input to futures research, as do judgments from Futures Wheels, Cross-Impacts, and Scenarios. BD&ML do not make it easier or more difficult to define objectives. Just another input. Same with methodological choice." *"These tools might increase the range of methods available but are unlikely to make the choice of methods easier. This will still have to rely on human intelligence and objectives."*
Recruitment	61	*"Projection 4: Big Data tools are designed to solve problems related to excess of data. Using Big Data tools means that you have a lot of data to process. Such analysis tend to be time-consuming and hard. I consider that Big Data tools enables future-oriented activities to be performed using huge datasets, but this analysis will be hard to be performed, requiring a Data Scientist."* *"The capabilities needed for foresight need to be rooted in different disciplines, also including a "humanistic" touch for understanding macro societal dynamics. I believe that professional/scholars developing future studies need to be sufficiently familiar with technological opportunities (data literacy) but not necessarily able to code themselves. There might be more availability of data, but they might be proprietary databases."* *"The ability to process more data automatically will make futures studies more complex, not simpler. The use of automated big data analyses and machine learning tools is also very prone to biases, which will make the process of using them and interpreting their results more complex, which does not necessarily mean that results will be "better"."* *"While computers may enable more rapid convergence on consistency, users must still fully understand the interactions in a consistent future to be able to lead/manage the change the future represents. This is a cognitive process which is not easily sped up. This makes it less likely Big Data will significantly speed up studies in future. Some increase in projection accuracy may potentially be enabled via big data, as long as the machine learning is not trained on biased data, which may enhance the basis of prediction."*
Generation	59	*"The ethics involved must be established on manipulated or biased data and the conscious human should be strong in order*

		to avoid the biased data. The personal that manage this kind of data should be conscious of this, of their higher responsibility." *"Human intervention to bias results will be always present and should be traced to avoid fraud."* *"Futures work needs to be and maybe transformed by big data, predictive analytics and AI--but bias and developing a objective holistic IT framework is beyond the industry players today."* *"Expert knowledge and communication skills remain critical for qualitative work in foresight studies and cannot be replaced by these data techniques."* *"Projection 6: Big Data tools are expected to enable the analysis of huge datasets. Machine Learning tools may be applied to improve the ETL (Extract Transform Load), thus, reducing inconsistencies. However, in the current small dataset paradigm, a well performed ETL is enough to deliver a good-quality data analysis. I consider Big Data and Machine Learning solutions tools especially useful for new types of analysis, not replacing the tools that already work."*
Action	52	*"Hard to say. In my opinion it enhances use of forecasting instead of foresight, but with the exception when it is used in a qualitative way to better understand what drives what. In general, a model that becomes more detailed, does not necessarily become more accurate. In fact I think the other way around. Often aggregated estimates will prove to outperform detailed ones on many occasions."* *"Big data and ML are essentially about trajectories of the past. They are mostly relevant for planning / forecasting, with less impact on possibilistic or constructivist foresight."* *"P9 - This will depend on acceptance by decision-makers, not just researchers. That will likely take longer. P10 - This is more likely to be true for near-term foresight (1-3 years) than longer-term foresight."* *"I don't see foresight requiring less human intelligence any time soon. I rather see a big danger that people rely too much on*

		"artificial intelligence", which contains embedded biases that are not transparent."
Renewal	46	*"Many factors impact on achieving foresight study objectives (in the system where the foresight is focused), beyond the scope of particular studies (i.e. power, politics, culture, change management, etc)."* *"Extraction of objectives from trends in big data and machine learning requires robust data exploration and visualisation techniques in parallel, as well as thorough scans of the design space being considered. This is likely to mean this process will still be complex, but easier than it is today."*

Source: Author's elaboration

4.3 Discussion: future of futures studies

Results shown in this chapter allow us to aggregate the projections in four groups based on the mean value of Expected Probability (EP) and Impact in the Foresight Industry (IF). These groups consolidate the elements of future-oriented activities influenced by the adoption of Big Data and Machine Learning tools and techniques. The first group is formed by projections whose expected probability (EP) means are greater than 3,5 (high probability of occurrence) and includes the projections P3 (require analytical skills), P9 (support results transfer) and P8 (increase data manipulation). Group two is composed of projections whose expected probability (EP) means are greater than 3,25 and less than 3,50 (medium-high probability of occurrence) and includes projections P4 (support data collection), P5 (improve data quality), P6 (enhance data analysis) and P7 (drive qualitative methods). The third group includes projections whose expected probability means are less than 3,25 and greater than 3,00 (medium-low probability of occurrence), with the projections P2 (guide methodological choice), P10 (Increase decision accuracy) and P11 (Support study evaluation). The last group is formed by projections whose expected probability means are less than 3,00 (low probability of occurrence). Figure 49 illustrates the four groups of projections and the mean values of EP (x-axis), IF (y-axis) and DE (bubble size) .

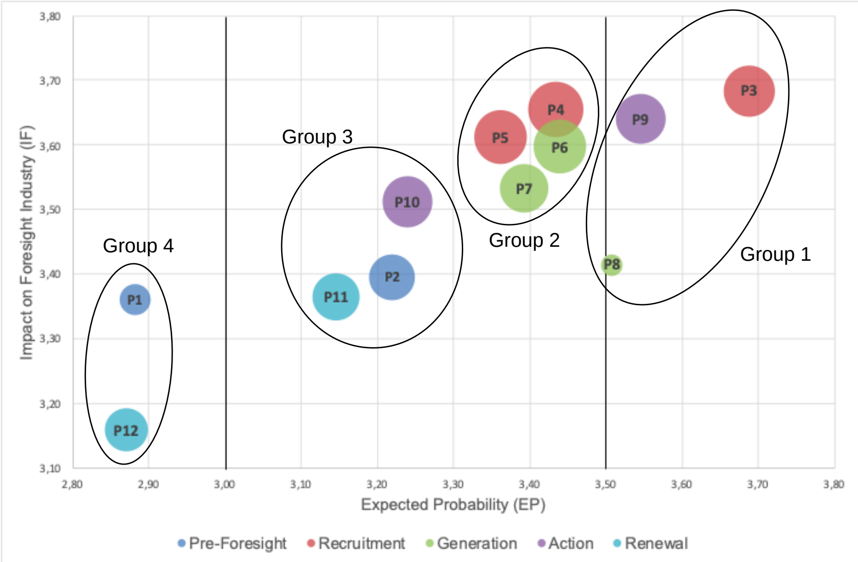


Figure 49 – Group of projections based on EP and IF

Source: Author's elaboration

4.3.1 Foresight Team, Results Transfer and Data Uncertainty

The projections in the first group have shown high expectation of occurrence. Except for P8 (increasing data manipulation), new required skills to deal with data, and the role of BDML tools in transferring foresight results to action, are highly desirable and highly impacting projections in a near future.

The foresight core team is an essential element to assure the quality of a future-oriented activity's process and results. Keller & von der Gracht (2014) already discussed the practitioner's role and the foresight core team in managing ICT tools and Foresight Support Systems. They claim that team competencies allow better communication among stakeholders and a better combination of ubiquitous data from different research fields. Pankratova & Savastiyanov (2014) also state that the foresight process's quality and results in a big data era is highly dependent on human abilities, mainly domain and analytics competencies. Data scientists are necessary for performing data collection, storage, cleaning, and analysis. They are critical to ensure an accurate interpretation of the analyzes, together with domain experts. Hines et al. (2017) explore the necessary infrastructure for foresight and state that, farther than ICT tools, foresight consultancy firms have developed considerable knowledge in databases, processes, and analytical tools to deal with big data in futures studies.

According to the survey results, foresight experts agreed that analytical competencies are fundamental to deal with future-oriented activities in the future. Based on the three scenarios for the future foresight professionals, Gary & von der Gracht (2015) determined that it is most expected that analytical competencies will be integrated with other professional skills in the Assimilation scenario. In other words, a futurist needs to seek data and analytical expertise, as well as strategic design, innovation, and competitive intelligence.

In foresight supported by BDML tools, techniques such as deep learning, AI, and NLP, can automatically report and share results of the analyzes performed, constantly monitoring emerging topics and complementing future scenarios already developed. In this way, BDML tools will no longer be just passive tools in executing and implementing foresight. They can become virtual assistants who collect data, analyze, and even make decisions (KELLER & VON DER GRACHT, 2014; BELKOM, 2020). In general terms, short-term automated decision-making is already a reality in electricity prices due to Smart Grids technologies (ROOZBEHANI ET AL., 2010) and financial markets (CHABOUD ET AL., 2014). However, for long-term decisions, automated tools' reliability seems more complicated, and it will demand much more technical and technological developments in Machine Learning and AI to become a reality.

In a certain way, some tools have been used to automatically deriving consequences after analyzing foresight results and transfer them to actual actions. Rinne (2004) describes his approach of automating the transfer of foresight to action through understanding the dynamics of detecting opportunities, innovation, market limitation, and the interaction of elements in a roadmap exercise. Gausemeier et al. (1998) present a scenario transfer process, analyzing the derivation of threats and opportunities from scenario analysis to formulate policies, strategies, and actions. A widely known approach used to employ foresight outcomes is based on multi-criteria decision-making tools (MCDM). Campanella & Ribeiro, (2011) conceived a multi-criteria tool adapted to dynamic circumstances, aiming to solve the problem of many different fields in decision-making under a dynamic environment.

All the mentioned experiences show the influence of BDML tools in deriving futures studies outcomes into decision-making. Survey respondents considered that BDML will support transferring foresight results to action in a near-future.

Data vulnerability is already an issue regarding the reliance on distributed information systems. Bertino & Sandhu (2005) reinforce that the popularization of cloud computing, grid-based computing, and on-demand business brings new dynamic challenges to data protection, confidentiality, integrity, and availability. An example of data security concerns is the hack of a massive amount of data from the Climatic Research Unit at the University of East Anglia in November 2009, also known as "Climategate." The hack attack intended to disqualify the scientists' arguments on climate change by scrutinizing data quality and manipulation. Such controversy has significantly affected public beliefs about global warming.

All these elements can cause a significant impact data reliability, and survey respondents considered that it is expected to be a relevant and undesirable issue in the short future for futures studies.

4.3.2 Foresight Methods and Data Analytics

The projections in this group have shown a medium-high expectation of occurrence, high desirability, and also a high impact on the foresight industry if it occurs. These four projections discuss data aspects and how data can drive qualitative and creative steps in foresight methodologies.

The use of data analytics is already a reality in the development of futures studies. Many authors (KAYSER ET AL., 2014; KAYSER & SHALA, 2020; SANTO ET AL., 2006; YUFEI ET AL., 2016) apply big data analytics and machine learning techniques for future-oriented activities consistently and in line with the best-known foresight methodologies, such as Scenarios and Roadmap. The integration of ICT, Big Data, and Machine Learning tools in futures studies is, until now, very applicable to many of the traditional foresight methodologies. Such tools can act in different moments of foresight process, such as in desk research, literature, and consistency analysis through information search strategy, data extraction, pre-processing, analysis, and validation techniques. The bibliometric analysis developed in this work identified the foresight methodologies that are most currently integrated with BDML tools: Patent Analysis, Time Series Analysis, Scenarios, Modeling, and Bibliometrics. Such methodologies are quite suitable for integration with new methods to collect and analyze massive data since most of them are quantitative methodologies. However, semi-quantitative and qualitative methodologies also benefit from using these techniques as a way to guide, based on data and evidence, insights, discussions and decision making.

4.3.3 Decision Making and Foresight Evaluation

Group 3 includes projections that have a medium-low expected probability of occurrence. However, these projections’ desirability is classified as high and the impact on foresight industry as medium. The projections correspond to the beginning and the end of the foresight process. The projections are P2 (support methodological choice), P10 (improve decision accuracy), and P11 (support study evaluation).

According to Popper (2008), the methodological choice for a foresight study is multi-factor and predominantly based on intuition, insight, impulsiveness, and, sometimes, inexperience. The author also concludes that elements such as the method's nature (qualitative, semi-quantitative, and quantitative) and the mix of methods (how methods can be framed together in a single study) are the most influential factors in methodological decisions. However, experts consulted in the survey indicate that access to BDML tools may have little influence on the study's methodological choice.

Along the same lines, confidence and accuracy in decision-making based on a futures study supported by BDML is also limited. Confidence is associated with subjective uncertainty in decision making. In the field of behavioral decision research, confidence can be measured as an indicator of the level of belief in the quality of the decision taken (SNIEZEK, 1992). In general, experts consider that there is not expected that using BDML tools can impact the level of belief in the quality of their decisions.

Although foresight has been used for decades, the evaluation and monitoring of studies' effectiveness and efficiency are still scarce in the literature. This deficiency of evaluation is caused by the lack of integrated and complex approaches to foresight, the frequency of execution and repeatability of foresight projects, and the long-term horizon of future-oriented research (MAZURKIEWICZ

ET AL., 2013). Georghiou & Keenan (2006) talks about the different types of foresight assessment according to the different rationales used in the study. The generic motivations for evaluation and monitoring are based on accountability, justification, and consolidation of the lessons learned from a foresight study. In general, the authors present three examples of rationales for foresight: providing police advice, building advocacy coalitions, and providing social forums. Based on each rationale's expected outcomes, the evaluation can focus on the changes in decision-making processes, the number of actors included in discussions, the nature of networks generated, the future actions, the quality of future-oriented debates, and the benefits to participants. Thus, experts considered that monitoring and evaluation of futures studies suffer a medium-low impact from massive data.

4.3.4 Futures Studies Objective

The projections in group 4 have a low value of the expected probability of occurrence, a medium value of desirability, and a low impact on the foresight industry. In general, experts considered that BDML tools would not significantly impact the foresight process in supporting objectives definition (P1) and objectives' reach (P12).

Foresight objectives definition is the initial stage of foresight development. Even before defining methodologies, teams, and specialists to be consulted, the rationales and objectives of a future-oriented activity must be well defined and consistent. Such definition is directly related to the field of knowledge, the sponsors' motivation, and the type of decision that foresight study will inform. Thus, it is understandable that data collection and analysis tools have little influence on defining key objectives.

In general, the use of BDML in futures studies catalyzes discussions, insights, and decisions in the foresight development process. However, in addition to the uncertainty elements related to data (bias, quality, analysis), external factors such as power, politics, culture, and change management impact achieving foresight objectives.

4.4 Conclusion

Chapter 4 of this work, the future of futures studies, set out to understand the extent to which the adoption of Big Data and Machine Learning tools impact futures studies, conceptually, methodologically, and in the perception of uncertainty in decision making. This discussion is accompanied by the growing use of massive data and robust analysis in futures studies by practitioners and foresight experts, as pointed out in chapter 3 of this work. It was collected opinions from 479 foresight experts about projections related to foresight and BDML use. Such projections were built from the literature review and bibliometric data analysis, exploring the five stages of the foresight process (MILES, 2002).

In connection with the first stage of the foresight process, called Pre-Foresight, experts were asked about the impacts of BDML on the definition of objectives and methodology. In Recruitment, the projections involved implications on data collection and quality, in addition to required analytical skills for futures studies supported by BDML. Regarding the Generation phase, the projections reflected the impacts of the data (analysis and possible biases) and the integration of BDML tools and foresight methodologies. In Action, the projections assessed the effects on transferring results to support strategy or policymaking and the general accuracy in the decision-making. In the last stage of the foresight process, the projections focused on BDML tools' impact in monitoring, evaluating, and achieving futures study objectives.

The results obtained through the web survey showed that the projections could be grouped into four different groups according to the expected occurrence probability. The groups are divided into projections with 1) a high probability of occurrence, 2) medium-high probability of occurrence, 3) medium-low probability of occurrence, and 4) low probability of occurrence. In addition to the projections, the experts were asked about their experience using BDML tools to reduce the study's uncertainty and to manage demographic issues.

The projections with a high probability of occurrence were P3 (require analytical skills), P9 (support results transfer), and P8 (increase data manipulation). Except for data manipulation, such projections were also classified as high desirability and high impact on the foresight industry. The adoption of

BDML tools in futures studies strongly impacts uncertainty regarding massive data integrity and possible biases. Such a result is shown as an obstacle to adopting these tools, classified as low desirability by the experts.

The foresight team will demand a data scientist responsible for managing the study data. Such competence must meet foresight's objectives in collecting, pre-analysis, analysis, and visualization of the data, supporting the entire team in the process of joint interpretation of the obtained information. Experts also believe that BDML tools may help transfer the results for adequate support in decision-making. The literature explores automatic decision-making systems, which the use of massive data could strongly impact.

The projections with a medium-high probability of occurrence were P4 (support data collection), P5 (improve data quality), P6 (enhance data analysis), and P7 (drive qualitative methods). In general, despite not reaching the highest level of expected probability in the survey, such projections have shown that they are highly desired and have a high impact on the foresight industry. Many authors use BDML tools complementarity with one or more foresight methodologies, assisting them in collecting, analyzing, visualization, and quality of the data used in the futures study.

The projections with a medium-low probability of occurrence were P2 (guide methodological choice), P10 (increase decision accuracy), and P11 (support study evaluation). Such projections showed high desirability by the respondents but a medium impact in the foresight industry. Among the criteria for a futures study's methodological definition, BDML tools are not as relevant as the study's budget, sponsor motivation, and scientific field. However, the use of such tools is shown to have a slight influence concerning accuracy in decision-making based on foresight, mainly in forecast studies.

The low probability projections were P1 (support objectives definition) and P12 (support reach the objectives). Such projections showed medium desirability and low impact on the foresight industry. It can be concluded that the use of BDML tools does not influence the definition of the objective or its accomplishment since many factors that influence the reach are beyond the scope of general studies.

Some methodological limitations of the survey analysis must be mentioned. First, we could not run a complete comparison between the characteristics of the original sample (7.753 foresight experts) to those who gave complete and valid answers to the survey. However, the minimum sample size was achieved for 95% of confidence level and 5% of margin of error (FLOREY, 1993). Second, the experts e-mails addresses were collected through academic publication and 66% of valid answers point to university affiliations. Therefore an academic bias could influence on the study.

FINAL REMARKS

Futures studies have played an essential role in supporting the design of public policies, corporate strategies, and decision-making (GEORGHIOU ET AL., 2008). The concepts and methods have evolved and adapted to deal with the highly uncertain and complex environment we live in (MILES ET AL., 2008). Technological developments, mainly related to information and communication technologies (ICT), directly impact the evolution of studies and how the future is thought, being one of the responsible for its greater efficiency and dissemination (KELLER & VON DER GRACHT, 2014). The digital revolution introduced and popularized new paradigms related to data production, collection, storage, and analysis. Big Data and Machine Learning (BDML) are technologies currently employed in several contexts, impacting many aspects of social and economic life. However, few researchers study the impacts that such tools produce in futures studies, both conceptually and methodologically. Thus, this work intends to elucidate such questions.

This project's overall objective is to identify and analyze trends of futures studies in science, technology, and innovation from the perspective of Big Data and Machine Learning, to understand how these techniques are changing conceptual and methodological approaches for futures studies. For specific objectives, it proposes 1) analyze the status and potential of Big Data and Machine Learning tools applied to prospective studies in STI, 2) Discuss how BDML affect the current understanding of uncertainty in futures studies, and 3) Identify impacts of BDML in the methodological approaches for futures studies. A central research question is: Given the pieces of evidence that Big Data and Machine Learning have obvious effects on the availability and processing of large amounts of data, to what extent these effects will affect the capabilities, the methodological approaches, and the tools employed for futures studies? Which roles Big Data and Machine Learning may play in the permanent endeavor of dealing with uncertainty?

Thus, this work was built in two parts, with two chapters each, and this final chapter of concluding remarks. The first part of the work aims to present the concepts that underlie the thesis' entire discussion. The second part focuses on practical experiences and analyzes involving futures studies and BDML.

The first chapter presents the theoretical bases for futures studies and foresight, including concepts related to uncertainty and decision-making. The second chapter presents concepts related to Big Data and Machine Learning, including the discussion about the origin, development, applications, and limitations of these technologies. Both chapters of the first part use literature review as the main methodological approach. The third chapter aims to understand the current panorama of futures studies supported by BDML tools. In the fourth chapter, the objective is to prospect the near future impact of BDML on conceptual and methodological elements of foresight. The complete methodology of part 2 is strongly based on Keller & von der Gracht (2014) and on Mota et al. (2020). A bibliometric analysis was conducted to collect and analyze data regarding futures studies supported by BDML tools in the past decades. Text Mining and Social Network Analysis were used to analyze the available data. The fourth chapter collected opinions about the impacts of BDML in futures studies through a survey sent to foresight specialists (list of experts was obtained through bibliometric data). Experts evaluated 12 projections about futures studies and BDML. Statistical and data visualization techniques were used to analyze the survey results. The methodology of chapters 3 and 4 of this work are complementary approaches and similar to the framework of many technological foresight methodologies.

Futures studies are directly connected to uncertainty. Researchers explore several approaches of uncertainty to base their investigation and theoretical frameworks. In futures studies, uncertainty is understood as time-dependent (SARITAS & ONER, 2004), which means that it is more intense when projected in a more distant future. Uncertainty is also classified in many ways, but its objective and subjective dimensions are further explored in this work. Several methodologies are employed to deal with the uncertainty of the future and the complex environment of society. A generic foresight process, presented by Miles (2002) and Popper (2008), are used as a framework to understand future-oriented activities. It includes 5 phases: Pre-Foresight, Recruitment, Generation, Action, and Renewal. Foresight methodologies are also divided into qualitative, semi-quantitative, and quantitative approaches.

Technological development offers several new tools for creating, collecting, storing, and analyzing data, such as Big Data and Machine Learning.

Big Data is commonly conceptualized from the 3 Vs: large data volume, high data variability, and fast speed of creation, collection, and analysis of data. Machine Learning is understood as algorithms that, based on training data, replicate the found patterns to predict future data behavior. These two technologies are complementary since Machine Learning algorithms are often used in the analysis of massive data. The greater the volume of data, the more accurate the learning acquired by machine learning algorithms. Several analyses can be obtained from BDML, such as text analytics, audio analytics, video analytics, social media analytics, and predictive analysis (forecast). However, BDML has some limitations. The creation and sharing of misinformation, fake news, and deep fakes can compromise data quality and confidence. Issues related to interpretability, ethics, and data privacy also generate discussions among experts. The use of BDML tools in futures studies is already a reality. Practitioners uses BDML analysis and techniques to generate insights, encourage discussions, and support future decisions. In general, such tools are being methodologically incorporated into futures studies and can directly impact perspectives on the future.

Before deepening the discussion on the impact of BDML in futures studies, this work sought to bring an overview of the futures studies' development based on BDML tools. Authors (MILES ET AL., 2008, KELLER & VON DER GRACHT, 2014) already pointed to the greater use of ICT tools and their impact on collaboration, data management, dissemination, and effects. The bibliometric analysis of this work showed that BDML tools' use is still quite restricted, at least in the academic literature. Less than 2% of foresight articles in STI applies BDML techniques as a part of future-oriented activities. However, its growth has occurred consistently in recent years. The United States and South Korea are countries that stand out for their contribution in this area, as well as the journal "Technological Forecasting and Social Change", with almost 20% of the papers of futures studies with BDML. It is essential to highlight universities' role in the development of BDML applications in futures studies, mainly weaving collaborations with research institutes, companies, and governments.

Regarding the methodologies, it is observed the frequent use of BDML techniques supporting foresight methodologies, such as patent analysis, time series analysis, scenario, modeling, and bibliometrics. The most used techniques are text mining (as used in this work), data mining, and machine learning algorithms. Thus, from these data, it is possible to understand how BDML has been incorporated to futures studies methodologies.

Researchers in futures studies were consulted through a web survey about the future impacts of the application of BDML tools. The survey contained 12 projections that dealt with the possible impacts of these tools on the process of futures studies in 2025. The short time horizon was defined because technological advances in the digital area grow exponentially, and the purpose of the survey is to understand also current impacts on futures studies that have already adapted to the new technological environment. The projections were divided in 5 foresight phases as constructed by Miles (2002) and Popper (2008) and addressed questions about 1) Pre-Foresight: definition of objectives and methodology, 2) Recruitment: recruitment of data and skills for future analysis, 3) Generation: analysis, manipulation, and integration of massive data with other methodologies, 4) Action: use of the study's results in accurate decision-making and 5) Renewal: evaluation of foresight achievement.

In general, the results show that it is highly expected that futures studies will require new analytical skills to deal with massive data and analysis. Transferring foresight results to decision making is also positively affected by BDML due to the data support, prioritization and automated decision-making. However, frequently data manipulation issues will bring a new element of uncertainty to this scenario. Experts also expect that the effectiveness of collecting, pre-processing, analyzing, and integrating data with foresight methodologies will be improved through the use of BDML tools. In general, the projections related to the initial phase of foresight development (Pre-Foresight) and the final phase (Renewal) were considered unlikely to occur. The experts do not foresee a BDML support in reaching overall objectives in futures studies. However, decision-making accuracy is likely to be improved.

Uncertainty and BDML-based futures studies

One of the objectives of this thesis was to understand the role of uncertainty in the foresight process. Thus, this session discusses the uncertainty frameworks that best suit the interaction of futures studies with BDML.

Uncertainty proved to be a topic little explored by researchers using BDML-based futures studies. Among the papers analyzed in chapter 3 (FSTI+BDML database), only 10 mentioned uncertainty in their abstracts or keywords.

Based on the literature review, bibliometric analysis, and analysis of the survey about futures studies and BDML, it is possible to add some evidences about the effects of BDML over uncertainty in foresight studies. Three dimensions may be highlighted: 1) data reliance, 2) data-method integration, and 3) decision-making.

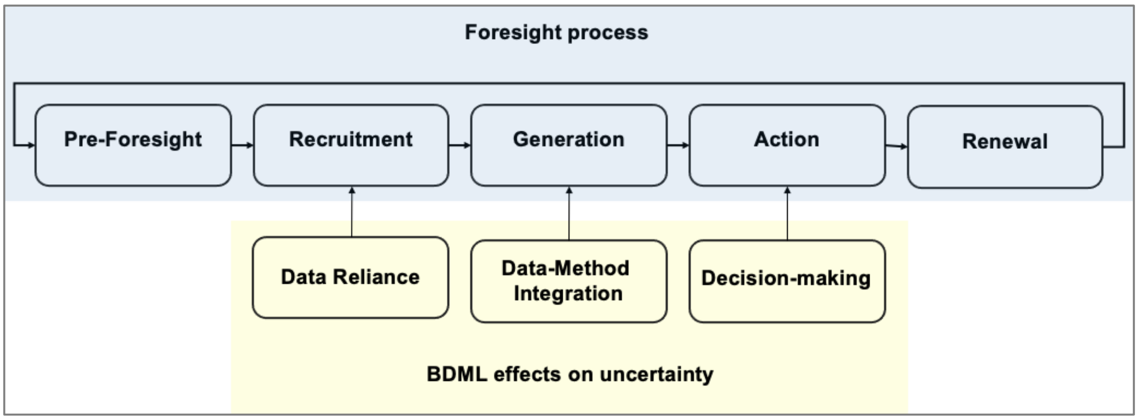


Figure 50 - BDML effects on uncertainty in foresight activities

Source: Author's elaboration based on Miles (2002); Popper (2008)

- Data reliance

Reimsbach-kounatze (2015) point out three reasons for Big Data analysis errors: low data quality, improper use of data, and a change in the data environment. Data reliance is directly linked with possible low quality and manipulation of future-relevant data. OECD (2011) states that data quality is related to its "fitness for use" and should be seen as a multi-faceted concept that includes relevance, accuracy, credibility, timeliness, accessibility, interpretability, and coherence. According to Reimsbach-kounatze (2015), 50% to 80% of a data analyst's time is spent cleaning up low-quality data. The quality of the data is directly associated with the quality of the results.

In addition to quality, the concern pointed out by the experts in the survey was the data manipulation, which involves partial data, biased algorithms, and data misuse. Such misuse can occur at various moments in the data analytics process, such as through dimensionality reduction and data cleaning techniques.

The results presented in chapter 4 point to a lack of data reliance due to the perspective of data manipulation. Data reliance can be seen as a foresight element impacted by BDML, slightly increasing the substantive uncertainty (DOSI & EGIDI, 1991) related to the futures study.

- Data-method integration

Keller and von der Gracht (2014) discuss the necessary infrastructure for the development of foresight and highlight the interdisciplinary nature of this field of futures studies. Practitioners seek information and support methodologies from other disciplines, such as statistics, econometrics, strategy, and innovation, which need to be systematically integrated into foresight. The integration of Big Data and Machine Learning tools and techniques in futures studies follows this same trend, being continuously integrated, adapted, and used by practitioners in their futures studies.

More effective data collection and analysis help practitioners understand the data patterns and, consequently, the research domain dynamics.

A greater understanding of the data directly affects what Dosi & Egidi (1991) defines as procedural uncertainty, uncertainty related to data analysis capacity. In this case, the impact is a significant decrease in uncertainty through the use of massive data and complex analyses. Data-driven forecast studies, which are characterized as data-centered and fewer complex problems, tend to be impacted by the use of BDML.

- Decision-making

Many factors influence decision-making in environments where foresight is employed. Some of these factors go beyond futures studies' scope to support decision-making, such as power, politics, and cultural relations. However, the results indicate that BDML may help transfer futures studies' results to support the development of public policies, business strategies, and decision making. The support of massive data and the application of techniques, such as Multi-criteria Decision-Making (MCDM), may directly assist the transfer of insights and visions developed in foresight to action (CAMPANELLA & RIBEIRO, 2011).

The accuracy of decision-making based on futures studies that rely on BDML is relatively impacted, since 60% of the consulted experts support the statement that BDML can reduce uncertainty related to foresight. Confidence in decision-making supported by foresight is a subjective perception, which measures the level of belief in the quality of the decision. The quality of the decision involves factors other than just the data and its analysis. In general, experts believe that prospective capabilities will not be significantly increased, but the accuracy of decision-making through massive data analysis will be positively impacted. There are evidences that short-term, data-driven forecast studies will be the most impacted by the adoption of BDML tools, which is in line with the empirical findings of Tetlock and Gardner (2015).

The conclusion is that subjective uncertainty will always be present in decision making, being slightly mitigated by using BDML tools and techniques, particularly for short time-horizon prospective which is strongly based on existing data.

However, taking the definition of true uncertainty proposed by many authors (as for Frank Knight; George Shackle; and Ralph Stacey), it is not possible to say BDML is a way to eliminate uncertainty, and this is true for at least one particular reason: there is no available data about the future.

Below, a summary table (Table 35) of the main elements of foresight impacted by using BDML tools, the impact on uncertainty, and a brief description are presented. The symbol (↑) indicates a slight increase in the level of uncertainty related to the foresight element. On the other hand, the sign (↓) indicates a slight decrease, and the symbol (↓ ↓) indicates a significant reduction in the level of uncertainty in the foresight. The last column summarizes the expected impacts related to each element.

Table 35 - Uncertainty impacts in BDML-based foresight

Dimensions of foresight	BDML Impact on uncertainty	Dimensions of uncertainty	Description
Data reliance	↑	Substantive uncertainty (DOSI & EGIDI, 1991)	The major limitation of BDML in futures studies is the reliability of the data. Experts agree that the possibility of manipulating and biasing data and disseminating misinformation can increase the uncertainty of the foresight exercise
Data-Method integration	↓↓	Procedural uncertainty (DOSI & EGIDI, 1991)	Data collection, analysis, interpretation, and integration in the foresight methodologies support practitioners in developing the study intensely and beneficially. A better understanding of the dynamics of the data helps to reduce the uncertainty associated with the future.
Decision-making	↓	Subjective uncertainty (TANNERT ET AL., 1996)	BDML-based foresight results supporting decision-making is a factor that points to a slight decrease in uncertainty. On the other hand, inevitably, the subjective uncertainty associated with the decision-making process will still be present due to the world's non-deterministic nature.

Source: Author‘s elaboration

Limitations of the study

Some limitations of this work in analyzing BDML tools applied to foresight studies must be mentioned. As a first point, the definitions of uncertainty present in the literature are diverse. This work discussed an outline of the concepts of uncertainty. However, other definitions, evaluation of its role in futures studies, and the manner of dealing with uncertainty may vary according to the theoretical framework used in different studies.

Methodological limitations may be mentioned, mainly in chapters 3 and 4. Many futures studies are not published in scientific journals and, therefore, were not included in this bibliometric analysis. In survey analysis, it was not possible to verify survey sample statistical significance. Population and sample characteristics were not compared due to a limitation of population data, although the minimum sample size was achieved for 95% of confidence level and 5% of margin of error.

Another limitation is the exponential evolution of digital technologies. The technological development of new techniques for collecting, cleaning, and analyzing massive data can change methodological approaches for futures studies. Standardization, access, and improvement in data quality can directly affect data reliance. New digital technologies may prove to be more integrated with the foresight methods presented in this work, and the automation of decision-making tools can affect the perception of uncertainty.

Finally, one more word is worth mentioning for the research agenda in the domain of foresight and futures studies. Whatever the evolution of BDML (and other similar approaches yet to come), uncertainty as an absolute concept will always prevail. Data may improve qualified information but will never substitute imagination, creativity, and expectation. If it does so, decision-making will not be a matter of human affairs anymore. Instead, what may be the most important effect of BDML over foresight is precisely its influence on imagination, creativity, and expectation. If this is true, uncertainty may even be increased because large amounts of data and a large capacity of processing will open new frontiers of knowledge.

BIBLIOGRAPHY

ADNAN, K.; AKBAR, R. **Limitations of information extraction methods and techniques for heterogeneous unstructured big data.** International Journal of Engineering Business Management. SAGE Publications Inc. 2019.

ALBERT, T.; MOEHRLE, M. G.; MEYER, S. **Technology maturity assessment based on blog analysis.** Technological Forecasting and Social Change, v. 92, p. 196–209, 1 mar. 2015.

ALBORNOZ, M. **Applying FTA methods in Latin American Countries.** *In*: Future-Oriented Technology Analysis: Strategic Intelligence for an Innovative Economy. Springer Berlin Heidelberg, 2008.

ALLCOTT, H.; GENTZKOW, M. **Social media and fake news in the 2016 election**. Journal of Economic Perspectives, v. 31, n. 2, p. 211–236, 1 mar. 2017.

AMANATIDOU, E. **“Joint foresight”: Towards a mechanism for joint programming in Europe?** Foresight, v. 10, n. 6, p. 103–117, 2008.

AMANATIDOU, E.; BUTTER, M.; CARABIAS, V.; KONNOLA, T.; LEIS, M.; SARITAS, O.; SCHAPER-RINKEL, P.; RIJ, V. VAN. **On concepts and methods in horizon scanning: Lessons from initiating policy dialogues on emerging issues**. Science and Public Policy, v. 39, n. 2, p. 208–221, 1 mar. 2012.

ANDERSON, J. **Technology foresight for competitive advantage.** Long Range Planning, v. 30, n. 5, p. 665–677, 1 out. 1997.

ANDRADE, R. P. DE. **A construção do conceito de incerteza**. Nova Economia, v. 21, n. 2, p. 171–195, 2011.

APREDA, R.; BONACCORSI, A.; DELL’ORLETTA, F.; FANTONI, G. **Functional technology foresight. A novel methodology to identify emerging technologies.** European Journal of Futures Research, v. 4, n. 1, p. 13, 2016.

ASHTON, K. **That “Internet of Things” Thing.** RFID Journal, p. 97–114, 22 jun. 2009.

ASHTON, W. B.; JOHNSON, A. H.; STACEY, G. S. **Monitoring science and technology for competitive advantage**. Competitive Intelligence Review, v. 5, n. 1, p. 5–16, 1994.

BALACHANDRA, R.; FRIAR, J. H. **Factors for success in r&d projects and new product innovation: a contextual framework.** IEEE Transactions on Engineering Management, 1997.

BARBIER, G.; LIU, H. **Data Mining in Social Media**. *In*: Social Network Data Analytics. Springer US, 2011.

BARNEY, J. B.; NELSON, R. R.; WINTER, S. G. **An Evolutionary Theory of Economic Change**. Administrative Science Quarterly, 1987.

BEAVER, D.; KUMAR, S.; LI, H. C.; SOBEL, J.; VAJGEL, P. **Finding a needle in Haystack: Facebook’s photo storage**. Proceedings of the 9th USENIX Symposium on Operating Systems Design and Implementation, 2019.

BELKOM, R. VAN. **The Impact of Artificial Intelligence on the Activities of a Futurist.** World Futures Review, v. 12, n. 2, p. 156–168, 20 jun. 2020.

BERTINO, E.; SANDHU, R. **Database security - concepts, approaches, and challenges**. IEEE Transactions on Dependable and Secure Computing**.** Institute of Electrical and Electronics Engineers Inc, 2005.

BLAIR, P. **Technology Assessment: Current Trends and the Myth of a Formula**. Available at: <https://ota.fas.org/technology_assessment_and_congress/blair/>, May, 1994.

BOSTROM, N. **Superintelligence: Path, Dangers, Strategies**. Oxford: Oxford University Press, 2016.

BOYD, D.; CRAWFORD, K. **Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon.** Information Communication and Society, v. 15, n. 5, p. 662–679, jun. 2012.

CABRAL, B.; FONSECA, M.; MOTA, F. **What is the future of cancer care? A technology foresight assessment of experts' expectations**. Economics of Innovation and New Technology, v. 28, n. 6, p. 635–652, 2019.

CAGNIN, C. **Estudos de futuro de CT&I no Brasil.** Parcerias Estratégicas, v. 19, n. 39, p. 49–64, 2014.

CALLEBAUT, W. **Scientific perspectivism: A philosopher of science's response to the challenge of big data biology**. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, v. 43, n. 1, p. 69–80, mar. 2012.

CALOF, J.; SMITH, J. E. **Foresight impacts from around the world: A special issue.** Foresight, v. 14, n. 1, p. 5–14, 2012.

CAMERER, C.; WEBER, M. **Recent developments in modeling preferences: Uncertainty and ambiguity.** Journal of Risk and Uncertainty, v. 5, n. 4, p. 325–370, out. 1992.

CAMPANELLA, G.; RIBEIRO, R. A. **A framework for dynamic multiple-criteria decision making**. Decision Support Systems, v. 52, n. 1, p. 52–60, 1 dez. 2011.

CARIBONI, D. **Big data for climate-smart agriculture – enhancing and sustaining rice systems for Latin America and the world**, CIAT, available at: <https://ciat.cgiar.org/ciat-projects/big-data-for-climate-smart-agriculture-—enhancing-and-sustaining-rice-systems-for-latin-america-and-the-world/>, 2014.

CASTELLS, M. **Grassrooting the space of flows**. Urban Geography, 1999.

CCAFS. **Colombia rice growers saved from ruin after being told not to plant their crop (CIAT).** Available at: <https://ccafs.cgiar.org/media/news/colombia-rice-growers-saved-ruin-after-being-told-not-plant-their-crop-guardian>, Sept. 2015.

CHABOUD, A. P.; CHIQUOINE, B.; HJALMARSSON, E.; VEGA, C. **Rise of the machines: Algorithmic trading in the foreign exchange market.** Journal of Finance, v. 69, n. 5, p. 2045–2084, 1 out. 2014.

CHEN, A. **The Human Toll of Protecting the Internet from the Worst of Humanity**. The New Yorker, 28 jan. 2017.

CHEN, Y.; ARGENTINIS, E.; WEBER, G. **IBM Watson: How Cognitive Computing Can Be Applied to Big Data Challenges in Life Sciences Research.** Clinical Therapeutics, 2016.

CHESNEY, R.; CITRON, D. K. **Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security**. SSRN Electronic Journal, 6 ago. 2018.

CHOO, C. **Information management for the intelligent organization: the art of scanning the environment**. 3rd ed. ed. Medford NJ: Information Today, 2002.

CHOUDHURY, S.; FISHMAN, J. R.; MCGOWAN, M. L.; JUENGST, E. T. **Big data, open science and the brain: Lessons learned from genomics**. Frontiers in Human Neuroscience Frontiers Media S. A., may 2014.

COATES, J. F. **The future of foresight-A US perspective**. Technological Forecasting and Social Change, v. 77, n. 9, p. 1428–1437, 1 nov. 2010.

COLOMBO, A. W.; SCHLEUTER, D.; KIRCHER, M. **An approach to qualify human resources supporting the migration of SMEs into an Industrie4.0-compliant company infrastructure.** IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society. Institute of Electrical and Electronics Engineers Inc., 2015

COOKE-DAVIES, T.; CICMIL, S.; CRAWFORD, L.; RICHARDSON, K. **We're not in Kansas Anymore, Toto: Mapping the Strange Landscape of Complexity Theory, and Its Relationship to Project Management.** Project Management Journal, v. 38, n. 2, p. 50–61, 1 jun. 2007.

COURTNEY, H. **20/20 Foresight: Crafting Strategy in an Uncertain World**. Boston, Mass: Harvard Business School Press, 2001.

CUHLS, K.; JOHNSTON, R. **Corporate foresight**. *In*: Future-Oriented Technology Analysis: Strategic Intelligence for an Innovative Economy. Springer Berlin Heidelberg, p. 103–114, 2008.

CUKIER, K. **Data, data everywhere: A special report on managing information**. The Economist, available at: <https://www.economist.com/special-report/2010/02/27/data-data-everywhere>, Feb. 2010.

CURNOW, R. C.; MORING, G. G. **"Project sappho". A study in industrial innovation.** Futures, v. 1, n. 2, p. 82–90, 1 dez. 1968.

DAVIDSON, P. **Is Probability Theory Relevant for Uncertainty? A Post Keynesian Perspective.** The Journal of Economic Perspectives, v. 5, n. 1, p. 129–143, 1991.

DAVIDSON, P. **Uncertainty in Economics**. *In*: Encyclopedia of Statistical Sciences. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2006.

DAVISON, L. **Measuring competitive intelligence effectiveness: Insights from the advertising industry.** Competitive Intelligence Review, 2001.

DELLOITE. **Industry 4.0 - Challenges and Solutions for the Digital Transformation and Use of Exponential Technologies.** Available at: <https://www2.deloitte.com/content/dam/Deloitte/ch/Documents/manufacturing/ch-en-manufacturing-industry-4-0-24102014.pdf>, 2015.

DEQUECH, D. **Fundamental Uncertanty and Ambiguity.** Eastern Economic Journal, v. 26, n. 1, p. 19, 2000.

DIEBOLD, F. X. **A Personal Perspective on the Origin(s) and Development of "Big Data": The Phenomenon, the Term, and the Discipline, Second Version**. SSRN Electronic Journal, 23 jan. 2013.

DOSI, G.; EGIDI, M. **Substantive and procedural uncertainty - An exploration of economic behaviours in changing environments**. Journal of Evolutionary Economics, v. 1, n. 2, p. 145–168, jun. 1991.

DOTSIKA, F.; WATKINS, A. **Identifying potentially disruptive trends by means of keyword network analysis.** Technological Forecasting and Social Change, v. 119, p. 114–127, 1 jun. 2017.

DURKIN, J. **Expert systems: A view of the field.** IEEE Expert-Intelligent Systems and their Applications, 1996.

EEROLA, A.; MILES, I. **Methods and tools contributing to FTA: A knowledge-based perspective**. Futures, v. 43, n. 3, p. 265–278, 1 abr. 2011.

ELITH, J.; LEATHWICK, J. R.; HASTIE, T. **A working guide to boosted regression trees**, Journal of Animal Ecology, jul. 2008.

ELLIS, E. G. **People Can Put Your Face on Porn—and the Law Can't Help You**, Wired, available at: <https://www.wired.com/story/face-swap-porn-legal-limbo/>, 2018.

ELLSBERG, D. **Risk, Ambiguity, and the Savage Axioms**. The Quarterly Journal of Economics, v. 75, n. 4, p. 643, nov. 1961.

EUROPEAN COMMISSION. **The European Foresight Monitoring Network: Collection of EFMN Briefs**. Luxembourg: Office for Official Publications of the European Communities, 2009

FLICK, C. **Informed consent and the Facebook emotional manipulation study.** Research Ethics, v. 12, n. 1, p. 14–28, 11 jan. 2016.

FLOREY, V. C. DU. **Sample size for beginners**. British Medical Journal, v. 306, n. 6886, p. 1181–1184, 1993.

FLORIDI, L.; STRAIT, A. **Ethical Foresight Analysis: What it is and Why it is Needed?** Minds and Machines, v. 30, n. 1, p. 77–97, 1 mar. 2020.

FORD, M. **Architects of Intelligence: The truth about AI from the people building it**. Birmingham: Packt Publishing, 2018.

FRIEDMAN, N.; GEIGER, D.; GOLDSZMIDT, M. **Bayesian Network Classifiers**. Machine Learning, v. 29, n. 2–3, p. 131–163, 1997.

GANDOMI, A.; HAIDER, M. **Beyond the hype: Big data concepts, methods, and analytics**. International Journal of Information Management, v. 35, n. 2, p. 137–144, 2015.

GARLAND, D. **The rise of risk.** Risk and Morality. University of Toronto Press,1 jan. 2003.

GARY, J. E.; GRACHT, H. A. VON DER. **The future of foresight professionals: Results from a global Delphi study**. Futures, v. 71, p. 132–145, 1 ago. 2015.

GAUSEMEIER, J.; FINK, A.; SCHLAKE, O. **Scenario Management: An Approach to Develop Future Potentials.** Technological Forecasting and Social Change, v. 59, n. 2, p. 111–130, 1 out. 1998.

GEORGHIOU, L.; HARPER, J. C.; KEENAN, M.; MILES, I.; POPPER, R.; ELGAR, E. **The Handbook of Technology Foresight**. Cheltenham: Edward Elgar Publishing, 2008.

GEORGHIOU, L.; KEENAN, M. **Evaluation of national foresight activities: Assessing rationale, process and impact.** Technological Forecasting and Social Change, v. 73, n. 7, p. 761–777, 1 set. 2006.

GINSBERG, J.; MOHEBBI, M. H.; PATEL, R. S.; BRAMMER, L.; SMOLINSKI, M. S.; BRILLIANT, L. **Detecting influenza epidemics using search engine query data**. Nature, v. 457, n. 7232, p. 1012–1014, 19 fev. 2009.

GLASSEY, O. **Folksonomies: Spontaneous crowd sourcing with online early detection potential?** Futures, v. 44, n. 3, p. 257–264, 1 abr. 2012.

GODET, M. **From forecasting to 'la prospective' a new way of looking at futures.** Journal of Forecasting, v. 1, n. 3, p. 293–301, 1 jul. 1982.

GRACHT, H. A; BANULS, V. A.; TUROFF, M.; SKULIMOWSKI, A. M. J.; GORDON, T. J. **Foresight support systems: The future role of ICT for foresight.** Technological Forecasting and Social Change, v. 97, p. 1–6, 2015.

GTAI. **Industries 4.0 - Smart Manufacturing for the Future.** Berlin: German Trade and Invest, 2014.

GUNDECHA, P.; LIU, H. **Mining Social Media: A Brief Introduction**. *In*: 2012 Tutorials in Operations Research. INFORMS, 2012.

HAARHAUS, T.; LIENING, A. **Building dynamic capabilities to cope with environmental uncertainty: The role of strategic foresight.** Technological Forecasting and Social Change, v. 155, 1 jun. 2020.

HAFEZI, R. **How Artificial Intelligence Can Improve Understanding in Challenging Chaotic Environments**. World Futures Review, v. 12, n. 2, p. 219–228, 23 jun. 2020.

HAFEZI, R.; AKHAVAN, A. N.; ZAMANI, M.; PAKSERESHT, S.; SHAMSHIRBAND, S. **Developing a data mining based model to extract predictor factors in energy systems: Application of global natural gas demand.** Energies, v. 12, n. 21, 2019.

HAGEN, L.; YI, H. S.; PIETRI, S.; KELLER, T. E. **Processes, potential benefits, and limitations of big data analytics: A case analysis of 311 data from city of Miami.** ACM International Conference Proceeding Series. Association for Computing Machinery, 18 jun. 2019

HARFORD, T. **Big data: A big mistake?** Significance, v. 11, n. 5, p. 14–19, 1 dez. 2014.

HASSAN, S. U.; IMRAN, M.; IQBAL, S.; ALJOHANI, N. R.; NAWAZ, R. **Deep context of citations using machine-learning models in scholarly full-text articles**. Scientometrics, v. 117, n. 3, p. 1645–1662, 1 dez. 2018.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. Springer Series in Statistics, 2009.

HAWKING, S. W. **A Brief History of Time - From the Big Bang to Black Holes**. Cambridge: Cambridge University Press, 1988.

HAZEN, B. T.; BOONE, C. A.; EZELL, J. D.; JONES-FARMER, L. A. **Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications**. International Journal of Production Economics, v. 154, p. 72–80, 2014.

HELMSTAEDTER, M.; BRIGGMAN, K. L.; TURAGA, S. C.; JAIN, V.; SEUNG, H. S.; DENK, W. **Connectomic reconstruction of the inner plexiform layer in the mouse retina.** Nature, v. 500, n. 7461, p. 168–174, 7 ago. 2013.

HERMANN, M.; PENTEK, T.; OTTO, B. **Design principles for industrie 4.0 scenarios.** Proceedings of the Annual Hawaii International Conference on System Sciences. IEEE Computer Society, 7 mar. 2016.

HERRERA, A. **Modelo mundial latinoamericano**. Nueva Sociedad, 1976.

HESS, D. J. **Science Studies: An Advanced Introduction**. NYU Press, 1997.

HILBERT, M. **Big Data for Development: A Review of Promises and Challenges**. Development Policy Review, v. 34, n. 1, p. 135–174, 1 jan. 2016.

HINES, A.; GARY, J.; DAHEIM, C.; LAAN, L. VAN DER. **Building Foresight Capacity: Toward a Foresight Competency Model.** World Futures Review, v. 9, n. 3, p. 123–141, 10 set. 2017.

HIRSCHBERG, J.; HJALMARSSON, A.; ELHADAD, N. **You're as sick as you sound: Using computational approaches for modeling speaker state to gauge illness and recovery.** *In*: Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics. Springer US, 2010.

IDEN, J.; METHLIE, L. B.; CHRISTENSEN, G. E. **The nature of strategic foresight research: A systematic literature review.** Technological Forecasting and Social Change, v. 116, p. 87–97, 2017.

JACK, C. **Lexicon of Lies: Terms for Problematic Information.** Data & Society Research Institute, p. 1–20, 2017.

JIANG, J. **Information extraction from text**. *In*: Mining Text Data. Springer US, 2013.

JORDAN, M. I.; MITCHELL, T. M. **Machine learning: Trends, perspectives, and prospectsScience.** American Association for the Advancement of Science, 17 jul. 2015.

JUN, S.; PARK, S. S. **Examining technological innovation of Apple using patent analysis**. Industrial Management and Data Systems, v. 113, n. 6, p. 890–907, 2013.

KAGERMANN, H.; WAHLSTER, W.; HELBIG, J. **Recommendations for implementing the strategic initiative INDUSTRIE 4.0**. Final report of the Industrie 4.0 WG, April, 2013.

KARACA, F.; ÖNER, M. A. **Scenarios of nanotechnology development and usage in Turkey.** Technological Forecasting and Social Change, 2015.

KASTRINOS, N. **Do statistics need foresight?** Foresight, v. 20, n. 2, p. 137–149, 2018.

KAUFMAN, L. M. **Data security in the world of cloud computing**. IEEE Security and Privacy, v. 7, n. 4, p. 61–64, 2009.

KAYNAK, O. **The exhilarating journey from industrial electronics to industrial informatics.** IEEE Transactions on Industrial Informatics, May, 2005.

KAYSER, V.; BIERWISCH, A. **Using Twitter for foresight: An opportunity?** Futures, v. 84, p. 50–63, 1 nov. 2016.

KAYSER, V.; BLIND, K. **Extending the knowledge base of foresight: The contribution of text mining**. Technological Forecasting and Social Change, v. 116, p. 208–215, 1 mar. 2017.

KAYSER, V.; BLIND, K. **Extending the knowledge base of foresight: The contribution of text mining.** Technological Forecasting and Social Change, v. 116, p. 208–215, 1 mar. 2016.

KAYSER, V.; GOLUCHOWICZ, K.; BIERWISCH, A. **Text Mining for Technology Roadmapping - The Strategic Value of Information.** International Journal of Innovation Management, v. 18, n. 03, 19 jun. 2014.

KAYSER, V.; SHALA, E. **Scenario development using web mining for outlining technology futures.** Technological Forecasting and Social Change, v. 156, p. 120086, 1 jul. 2020.

KELLER, J.; GRACHT, H. A. VON DER. **ICT and the Foresight Infrastructure of the Future.** World Futures Review, v. 6, n. 1, p. 40–47, 10 mar. 2014.

KELLER, J.; GRACHT, H. A. VON DER. **The influence of information and communication technology (ICT) on future foresight processes - Results from a Delphi survey.** Technological Forecasting and Social Change, v. 85, n. 2014, p. 81–92, 2014.

KEYNES, J. M. **A Treatise on Probability.** Dover Publications, 1921.

KILDIENE, S.; KAKLAUSKAS, A.; ZAVADSKAS, E. K. **COPRAS based comparative analysis of the European country management capabilities within the construction sector in the time of crisis.** Journal of Business Economics and Management, v. 12, n. 2, p. 417–434, 9 jun. 2011.

KIM, L.; JU, J. **Can media forecast technological progress?: A text-mining approach to the on-line newspaper and blog's representation of prospective industrial technologies.** Information Processing and Management, v. 56, n. 4, p. 1506–1525, 1 jul. 2019.

KING, J. **A review of bibliometric and other science indicators and their role in research evaluation**. Journal of Information Science, v. 13, n. 5, p. 261–276, 1987.

KNIGHT, F. **Risk, uncertainty and profit**. Boston; New York: Houghton Mifflin Company, 1921.

KRANENBURG, R. VAN; ANZELMO, E.; BASSI, A.; CAPRIO, D.; DODSON, S.; RATTO, M. **The Internet of Things.** Proceedings of 1st Berlin Symposium on Internet and Society, Berlin. 2011.

KRITTANAWONG, C.; KUKIN, M. L. **Current Management and Future Directions of Heart Failure With Preserved Ejection Fraction: a Contemporary Review**. Current Treatment Options in Cardiovascular Medicine. Springer Healthcare, , 20 abr. 2018.

KUOSA, T. **Evolution of futures studies**. Futures, v. 43, n. 3, p. 327–336, 2011.

L'HEUREUX, A.; GROLINGER, K.; ELYAMANY, H. F.; CAPRETZ, M. A. M. **Machine Learning with Big Data: Challenges and Approaches.** IEEE Access, v. 5, p. 7776–7797, 2017.

LABRINIDIS, A.; JAGADISH, H. V. **Challenges and opportunities with big data**. Proceedings of the VLDB Endowment, v. 5, n. 12, p. 2032–2033, ago. 2012.

LASI, H.; FETTKE, P.; KEMPER, H. G.; FELD, T.; HOFFMANN, M. **Industry 4.0**. Business and Information Systems Engineering, v. 6, n. 4, p. 239–242, 1 ago. 2014.

LAZER, D. M. J. *et al.* **The science of fake news: Addressing fake news requires a multidisciplinary effort.** Science, v. 359, n. 6380, p. 1094–1096, 9 mar. 2018.

LAZER, D.; BAUM, M.; GRINBERG, N.; FRIEDLAND, L.; JOSEPH, K.; HOBBS, W.; MATTSSON, C.; BENKLER HARVARD, Y.; WATTS, D. **Combating Fake News: An Agenda for Research and Action**. Harvard Kennedy School and Northeast University, May 2017.

LECUN, Y.; BENGIO, Y.; HINTON, G. **Deep learning.** Nature Publishing Group. 27 may 2015.

LEE, Y. W.; STRONG, D. M.; KAHN, B. K.; WANG, R. Y. **AIMQ: A methodology for information quality assessment**. Information and Management, v. 40, n. 2, p. 133–146, 1 dez. 2002.

LEE, D.; KANG, S.; SHIN, J. **Using deep learning techniques to forecast environmental consumption level.** Sustainability (Switzerland), v. 9, n. 10, 20 out. 2017.

LEE, J.; KIM, G.; JANG, D.; PARK, S. **A novel method for technology forecasting based on patent documents.** Advances in Intelligent Systems and Computing, v. 271, p. 81–90, 2014.

LI, S.; HU, JIE; CUI, Y.; HU, JIANJUN. **DeepPatent: patent classification with convolutional neural networks and word embedding**. Scientometrics, v. 117, n. 2, p. 721–744, 1 nov. 2018.

LIAO, Y.; DESCHAMPS, F.; LOURES, E. DE F. R.; RAMOS, L. F. P. **Past, present and future of Industry 4.0 - a systematic literature review and research agenda proposal**. International Journal of Production Research, v. 55, n. 12, p. 3609–3629, 18 jun. 2017.

LIKERT, R. **A technique for the measurement of attitudes**. Archives of Psychology, v. 22, n. 140, p. 55, 1932.

LIN, Y.; DUAN, X.; ZHAO, C.; XU, L. **Systems science: Methodological approaches**. 1st Edition ed. Boca Raton, FL: CRC Press, 2013.

LIU, R.; YANG, B.; ZIO, E.; CHEN, X. **Artificial intelligence for fault diagnosis of rotating machinery: A review.** Mechanical Systems and Signal Processing, Academic Press, , 1 ago. 2018.

LIU, B. **Sentiment analysis and opinion mining**. Synthesis Lectures on Human Language Technologies, v. 5, n. 1, p. 1–184, 23 may, 2012.

LOUKIDES, M. **What is data science? The future belongs to the companies and people that turn data into products**. O'Reilly, Radar/Data Science, Available at: <https://www.oreilly.com/radar/what-is-data-science/>, June, 2010.

LOVERIDGE, D. **Foresight: The Art and Science of Anticipating the Future**. Routledge, 2009.

LU, Y. **Industry 4.0: A survey on technologies, applications and open research issues.** Journal of Industrial Information Integration**.** Elsevier B.V.1 jun. 2017.

LUKAC, D. **The fourth ICT-based industrial revolution "industry 4.0" - HMI and the case of CAE/CAD innovation with EPLAN P8.** 2015 23rd Telecommunications Forum, TELFOR 2015. Institute of Electrical and Electronics Engineers Inc., 8 jan. 2016

MA, J.; SHERIDAN, R. P.; LIAW, A.; DAHL, G. E.; SVETNIK, V. **Deep neural nets as a method for quantitative structure-activity relationships.** Journal of Chemical Information and Modeling, v. 55, n. 2, p. 263–274, 23 feb. 2015.

MAGRUK, A. **Concept of uncertainty in relation to the foresight research**. Engineering Management in Production and Services, v. 9, n. 1, p. 46–55, 1 mar. 2017.

MARKOWETZ, A.; BŁASZKIEWICZ, K.; MONTAG, C.; SWITALA, C.; SCHLAEPFER, T. E. **Psycho-Informatics: Big Data shaping modern psychometrics.** Medical Hypotheses, v. 82, n. 4, p. 405–411, 1 abr. 2014.

MARTIN, B. R. **The origins of the concept of “foresight” in science and technology: An insider’s perspective.** Technological Forecasting and Social Change, v. 77, n. 9, p. 1438–1447, 1 nov. 2010.

MAZURKIEWICZ, A.; POTERALSKA, B.; SACIO-SZYMAŃSKA, A. **Implementation and evaluation of foresight results**. XXIV ISPIM Conference – Innovation in Global Markets: Challenges for Sustainable Growth. Helsinki, Finland, June, 2013.

MCCARTHY, J.; MINSKY, M. L.; ROCHESTER, N.; SHANNON, C. E. **A proposal for the Dartmouth summer research project on artificial intelligence.** AI Magazine, v. 27, n. 4, p. 12–14, 15 dez. 2006.

MCGUIRE, T.; MANYIKA, J.; CHUI, M. **Why Big Data is the new competitive advantage**. Ivey Business Journal. July/August 2012.

MEADOWS, D. H.; MEADOWS, D. L.; RANDERS, J.; BEHRENS III, W. W. **Limites do Crescimento: um relatório para o Projeto do Clube de Roma sobre o Dilema da Humanidade**. São Paulo: Perspectiva, 1973.

METCALFE, S.; SALLES-FILHO, S.; DUARTE, L. T.; BIN, A.; AZEVEDO, A. T. **Retaking Shackle ’ s model towards R & D and innovation planning and management**. 15th International Conference of the International Joseph A. Schumpeter Society (ISS),Vol. 1, pp.1-3, Jena, Germany, 2014.

MGI. **Big data: The next frontier for innovation, competition, and productivity**. McKinsey Global Institute. Available at: <https://www.mckinsey.com/~/media/McKinsey/BusinessFunctions/McKinsey Digital/OurInsights/Big data The next frontier for innovation/MGI_big_data_full_report.pdf>. May, 2011.

MILES, I. **Appraisal of Alternative Methods and Procedures for Producing Regional Foresight.** Conference: European Commission’s DG Research funded STRATA – ETAN Expert Group Action, p. 21, 2002.

MILES, I. **The development of technology foresight: A review**. Technological Forecasting and Social Change, v. 77, n. 9, p. 1448–1456, 2010.

MILES, I.; HARPER, J. C.; GEORGHIOU, L.; KEENAN, M.; POPPER, R. **The Many Faces of Foresight.** *In*: The Handbook of Technology Foresight- concepts and practices**,** 2008.

MILES, I.; SARITAS, O. **The depth of the horizon: Searching, scanning and widening horizons.** Foresight. Emerald Group Publishing Limited, nov. 2012.

MILES, I.; SARITAS, O.; SOKOLOV, A. **Foresight for Science, Technology and Innovation**. Science, Technology and Innovation Studies. Springer, 2016.

MILES, I.; MEISSNER, D.; VONORTAS, N. S.; CARAYANNIS, E. **Technology foresight in transition.** Technological Forecasting and Social Change, v. 119, n. April, p. 211–218, 2017.

MINGERS, J.; LEYDESDORFF, L. **A review of theory and practice in scientometrics.** European Journal of Operational Research**.** Elsevier,1 out. 2015.

MIŠKUF, M.; ZOLOTOVÁ, I. **Comparison between multi-class classifiers and deep learning with focus on industry 4.0.** 2016 Cybernetics and Informatics, K and I 2016 - Proceedings of the 28th International Conference. Institute of Electrical and Electronics Engineers Inc., 21 mar. 2016

MITCHELL, T. **Machine Learning**. New York: McGraw-Hill, 1997.

MITTELSTADT, B. D.; FLORIDI, L. **The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts**. Science and Engineering Ethics, v. 22, n. 2, p. 303–341, 23 abr. 2016.

MOEUF, A.; PELLERIN, R.; LAMOURI, S.; TAMAYO-GIRALDO, S.; BARBARAY, R. **The industrial management of SMEs in the era of Industry 4.0**. International Journal of Production Research, v. 56, n. 3, p. 1118–1136, 1 fev. 2018.

MOORE, P.; XHAFA, F.; BAROLLI, L.; THOMAS, A. **Monitoring and Detection of Agitation in Dementia: Towards Real-Time and Big-Data Solutions.** Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing. IEEE, oct. 2013.

MOTA, F.; BRAGA, L.; ROCHA, L.; CABRAL, B. **3D and 4D bioprinted human model patenting and the future of drug development.** Nature Biotechnology, v. 38, n. 6, p. 689–694, 2020.

MULLER, C. **Artificial Intelligence - The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society (own-initiative opinion).** European Economic and Social Committee. Available at: <https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/artificial-intelligence-consequences-artificial-intelligence-digital-single-market-production-consumption-employment-and>. 31 May 2017.

NEWMAN, M. E. J. **Resource Letter CS-1: Complex Systems**. Citation: American Journal of Physics, v. 79, p. 800, 2011.

OECD. **Quality Framework for OECD Statistical Activities**. OCDE Statistics Directorate, Nov. 2011.

OECD. **Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by "Big Data"**. OCDE Digital Economy Papers, n. No 222, 2013.

OECD. **Data-driven Innovation for Growth and Well-being: Interim Synthesis Report**. Oecd, 2014.

OLIVEIRA, L. E. S.; ÁLVARES, A. J. **Axiomatic Design Applied to the Development of a System for Monitoring and Teleoperation of a CNC Machine through the Internet.** Procedia CIRP.Elsevier B.V., 1 jan. 2016.

PANG, A. S. K. **Social scanning: Improving futures through Web 2.0; or, finally a use for twitter.** Futures, v. 42, n. 10, p. 1222–1230, dez. 2010.

PANKRATOVA, N.; SAVASTIYANOV, V. **Foresight Process Based On Text Analytics**. Information Content and Processing, v. 1, n. 1, p. 54–65, 2014.

POPPER, R. **Mapping foresight revealing how Europe and other world regions navigate into the future.** EFMN, Luxembourg: Publications Office of the European Union, European Commission**.** 2009.

POPPER, R. **How are foresight methods selected?** Foresight, v. 10, n. 6, p. 62–89, 2008.

PORTER, A. L. *et al.* **Technology futures analysis: Toward integration of the field and new methods.** Technological Forecasting and Social Change, v. 71, n. 3, p. 287–303, 1 mar. 2004.

PRANIGRAHI, B. K.; ABRAHAM, A.; DAS, S. **Computational intelligence in power engineering**. Berlin, Heidelberg: Springer, 2010.

PRIEM, J.; PIWOWAR, H. A.; HEMMINGER, B. M. **Altmetrics in the wild: Using social media to explore scholarly impact**. Computer Science. 20 mar. 2012.

PRIMI, A.; TOSELLI, M. A global perspective on industry 4.0 and development: new gaps or opportunities to leapfrog? Journal of Economic Policy Reform, v. 23, n. 4, p. 371–389, 1 out. 2020.

PUSCHMANN, C.; BURGESS, J. **Metaphors of big data**. International Journal of Communication, v. 8, n. 1, p. 1690–1709, jun. 2014.

QUIGGIN, J. **Economists and Uncertainty**: Risk and Sustainable Management Group Working Papers. University of Queensland, School of Economics, abr. 2005.

REGER, G. **Technology Foresight in Companies: From an Indicator to a Network and Process Perspective**. Technology Analysis & Strategic Management, v. 13, n. 4, p. 533–553, 2001.

REIMSBACH-KOUNATZE, C. **The Proliferation of " Big Data " and Implications for Official Statistics and Statistical Agencies**. OECD Digital Economy Papers, n. 245, p. 3–39, 2015.

RINNE, M. **Technology roadmaps: Infrastructure for innovation**. Technological Forecasting and Social Change, v. 71, n. 1–2, p. 67–80, 1 jan. 2004.

ROHRBECK, R. **Corporate foresight: Towards a maturity model for the future orientation of a firm**. *Springer, Nov*. 2010

ROHRBECK, R.; BATTISTELLA, C.; HUIZINGH, E. **Corporate foresight: An emerging field with a rich tradition**. Technological Forecasting and Social Change, v. 101, p. 1–9, 1 dez. 2015.

ROHRBECK, R.; KUM, M. E. **Corporate foresight and its impact on firm performance: A longitudinal analysis**. Technological Forecasting and Social Change, v. 129, n. February 2017, p. 105–116, 1 abr. 2018.

ROJKO, A. **Industry 4.0 Concept: Background and Overview**. International Journal of Interactive Mobile Technologies (iJIM), v. 11, n. 5, p. 77–90, 24 jul. 2017.

ROOZBEHANI, M.; DAHLEH, M.; MITTER, S. **Dynamic pricing and stabilization of supply and demand in modern electric power grids.** 2010 1st IEEE International Conference on Smart Grid Communications, SmartGridComm, 2010.

SAINATH, T. N.; MOHAMED, A. R.; KINGSBURY, B.; RAMABHADRAN, B. **Deep convolutional neural networks for LVCSR.** ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings. 18 out. 2013

SALO, A.; GUSTAFSSON, T.; RAMANATHAN, R. **Multicriteria methods for technology foresight**. Journal of Forecasting, v. 22, n. 2–3, p. 235–255, 1 mar. 2003.

SAMUEL, A. L. **Some Studies in Machine Learning Using the Game of Checkers**. v. 3, n. 3, p. 535–554, 1959.

SANTO, M.; COELHO, G. M.; SANTOS, D. M. DOS; FILHO, L. F. **Text mining as a valuable tool in foresight exercises: A study on nanotechnology**. Technological Forecasting and Social Change, v. 73, n. 8, p. 1013–1027, 2006.

SARITAS, O. **Systemic foresight methodology**. *In*: Science, Technology and Innovation Policy for the Future: Potentials and Limits of Foresight Studies. Springer, 2013.

SARITAS, O.; BURMAOGLU, S. **The evolution of the use of Foresight methods : a scientometric analysis of global FTA research output.** Scientometrics, v. 105, n. 1, p. 513–524, 2015.

SARITAS, O.; ONER, M. A. **Systemic analysis of UK foresight results Joint application of integrated management model and roadmapping.** Technological Forecasting and Social Change, v. 71, n. 1–2, p. 27–65, 2004.

SAUERMANN, H.; ROACH, M. **Increasing web survey response rates in innovation research: An experimental study of static and dynamic contact design features.** Research Policy, v. 42, n. 1, p. 273–286, 2013.

SCHENATTO, F. J. A.; POLACINSKI, É.; ABREU, A. F. DE; ABREU, P. F. DE. **Análise crítica dos estudos do futuro: Uma abordagem a partir do resgate histórico e conceitual do tema.** Gestao e Producao, v. 18, n. 4, p. 739–754, 2011.

SHACKLE, G. L. S. **Decision Order and Time in Human Affairs**. *In*: Cambridge: Cambridge University Press, p. 3–113, 1969.

SHACKLE, G. L. **Uncertainty in Economics**. Cambridge: Cambridge University Press, 1955.

SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding Machine Learning: From Theory to Algorithms**. Cambridge University Press, 2014.

SHILTON, K. **Participatory personal data: An emerging research challenge for the information sciences.** Journal of the American Society for Information Science and Technology, v. 63, n. 10, p. 1905–1915, 1 out. 2012.

SHIRLEY, D. A.; LANGAN-FOX, J. **Intuition: A Review Of The Literature.** Psychological Reports, p. 563–584, 1996.

SIMON, H. A. **Theories of Decision-Making in Economics and Behavioral Science**. The American Economic Review, v. 49, n. 3, p. 253–283, 1959.

SIMON, H. A. **Why Should Machines Learn?** *In*: Machine Learning. [s.l.] Springer Berlin Heidelberg, 1983.

SIMON, H. A. **Making Management Decisions: the Role of Intuition and Emotion**. Academy of Management Executive, Vol. 1, No. 1, pp. 57-64, 1987.

SIMON, H. A. **Bounded rationality in social science: Today and tomorrow**. Mind & Society, v. 1, n. 1, p. 25–39, mar. 2000.

SLAUGHTER, R. A. **Foresight Beyond Strategy: Social Initiatives by Business and Government.** Long Range Planning, 1996.

SNIEZEK, J. A. **Groups under uncertainty: An examination of confidence in group decision making**. Organizational Behavior and Human Decision Processes, v. 52, n. 1, p. 124–155, 1 jun. 1992.

STACEY, R. **Complexity and the Group Matrix**. Group Analysis, v. 34, n. 2, p. 221–239, 30 jun. 2001.

STAHL, B. C.; WRIGHT, D. **Ethics and Privacy in AI and Big Data: Implementing Responsible Research and Innovation**. IEEE Security and Privacy, v. 16, n. 3, p. 26–33, 2018.

TANNERT, C.; ELVERS, H.; JANDRIG, B. **The ethics of uncertainty**. EMBO reports, v. 8, n. 10, p. 892–896, 7 out. 2007.

TENE, O.; POLONETSKY, J. **Big Data for All: Privacy and User Control in the Age of Analytics.** Northwestern Journal of Technology and Intellectual Property, v11, issue 5, 2013.

TETLOCK, P. E.; GARDNER, D. **Superforecasting: The art and science of prediction.** New York, NY, US: Crown Publishers/Random House, 2015.

THOM, N. **Measuring the Value Contribution of Corporate Foresight.** SSRN Electronic Journal, n. November 2010, 2012.

TOMPSON, J.; JAIN, A.; LECUN, Y.; BREGLER, C. **Joint Training of a Convolutional Network and a Graphical Model for Human Pose Estimation**. Advances in Neural Information Processing Systems, v. 2, n. January, p. 1799–1807, 11 jun. 2014.

TORABI, F.; TABOADA, M. **Big Data and quality data for fake news and misinformation detection**. Big Data and Society, v. 6, n. 1, 2019.

TRAPPEY, A. J. C.; TRAPPEY, C. V.; GOVINDARAJAN, U. H.; SUN, J. J. H. **Patent Value Analysis Using Deep Learning Models - The Case of IoT Technology Mining for the Manufacturing Industry.** IEEE Transactions on Engineering Management, p. 1–13, 2019.

UHM, D.; RYU, J. B.; JUN, S. **An interval estimation method of patent keyword data for sustainable technology forecasting**. Sustainability (Switzerland), v. 9, n. 11, 2017.

VANTAGE POINT. **Products – The Vantage Point**. Available in <https://www.thevantagepoint.com/products.html>; Accessed in: 20 January, 2021.

VARDE, P. V.; PECHT, M. G. **Risk-Based Engineering**. Singapore: Springer Singapore, 2018.

VAUGHAN, F. E. **Awakening Intuition**. Anchor Press, 1979.

VOROS, J. **A generic foresight process framework.** Foresight. Emerald Group Publishing Ltd. 2003.

VOSOUGHI, S.; ROY, D.; ARAL, S. **The spread of true and false news online**. Science, v. 359, n. 6380, p. 1146–1151, 9 mar. 2018.

WAKEHAM, J. **Uncertainty: History of the Concept**. Second Edition. Elsevier, 2015.

WANG, R. Y.; STOREY, V. C.; FIRTH, C. P. **A Framework for Analysis of Data Quality Research**. IEEE Transactions on Knowledge and Data Engineering, v. 7, n. 4, p. 623–640, 1995.

WANG, Y.; KUNG, L. A.; BYRD, T. A. **Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations**. Technological Forecasting and Social Change, v. 126, p. 3–13, 2018.

WASSERMAN, S.; FAUST, K. **Social Network Analysis: Methods and Applications**. Cambridge: Cambridge University Press, 1994.

WHITE, P.; BRECKENRIDGE, R. S. **Trade-offs, limitations, and promises of big data in social science research.** Review of Policy Research, v. 31, n. 4, p. 331–338, 2014.

WU, X.; ZHU, X.; WU, G. Q.; DING, W. **Data mining with big data**. IEEE Transactions on Knowledge and Data Engineering, v. 26, n. 1, p. 97–107, jan. 2014.

XIONG, H. Y. *et al.* **The human splicing code reveals new insights into the genetic determinants of disease**. Science, v. 347, n. 6218, p. 1254806, 9 jan. 2015.

XU, L. DA; XU, E. L.; LI, L. **Industry 4.0: State of the art and future trends**. International Journal of Production Research, v. 56, n. 8, p. 2941–2962, 2018.

XU, L. DA; HE, W.; LI, S. **Internet of things in industries: A survey.** IEEE Transactions on Industrial Informatics. IEEE Computer Society,1 nov. 2014.

YING, Z.; SUI PHENG, L. **Complexity Theory Review**. *In*: Project Communication Management in Complex Environments. Springer Singapore. p. 13–31. 2014.

YODA, T. **Perceptions of domain experts on impact of foresight on policy making: The case of Japan**. Technological Forecasting and Social Change, 2011.

YOON, J. **Detecting weak signals for long-term business opportunities using text mining of Web news**. Expert Systems with Applications, v. 39, n. 16, 15 nov. 2012.

YOUTIE, J.; HICKS, D.; SHAPIRA, P.; HORSLEY, T. **Pathways from discovery to commercialisation: Using web sources to track small and medium-sized enterprise strategies in emerging nanotechnologies**. Technology Analysis and Strategic Management, v. 24, n. 10, p. 981–995, nov. 2012.

YUFEI, L.; YUAN, Z.; LING, L. **Application of Big Data Analysis Method in Technology Foresight for Strategic Emerging Industries**. Chinese Journal of Engineering Science, v. 18, n. 4, p. 121, 2016.

ZAVADSKAS, E. K.; TURSKIS, Z. **Multiple criteria decision making (MCDM) methods in economics: An overview**. Technological and Economic Development of Economy, v. 17, n. 2, p. 397–427, 2011.

ZHANG, Y.; LU, J.; LIU, F.; LIU, Q.; PORTER, A.; CHEN, H.; ZHANG, G. **Does deep learning help topic extraction? A kernel k-means clustering method with word embedding.** Journal of Informetrics, v. 12, n. 4, p. 1099–1117, 1 nov. 2018.

ZHANG, Y.; ZHANG, G.; CHEN, H.; PORTER, A. L.; ZHU, D.; LU, J. **Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research.** Technological Forecasting and Social Change, v. 105, p. 179–191, 2016.

ZHEN, Z.; YAO, Y. **Optimizing deep learning and neural network to explore enterprise technology innovation model**. Neural Computing and Applications, p. 1–17, 23 jun. 2020.

ZHOU, Y.; DONG, F.; LIU, Y.; LI, Z.; DU, J. F.; ZHANG, L. **Forecasting emerging technologies using data augmentation and deep learning**. Scientometrics, v. 123, n. 1, p. 1–29, 1 abr. 2020.

ZINN, J. O. **Introduction: The Contribution of Sociology to the Discourse on Risk and Uncertainty**. *In*: Social Theories of Risk and Uncertainty: An Introduction. Oxford, UK: Blackwell Publishing Ltd., p. 1–17, 2009.

ANEXXES

ANNEX 1 - Free and Informed Consent Form

FREE AND INFORMED CONSENT TERM

Research Title: Trends in Foresight for Science, Technology and Innovation from Big Data and Machine Learning Perspectives
Researchers: Vinícius Muraro da Silva (viniciusmuraro@ige.unicamp.br), Professor Sérgio Salles-Filho (sallesfi@unicamp.br)
Ethics Committee Approval Number: 25592219.6.0000.8142

You are being invited to participate as a volunteer in this survey. This document, called the Free and Informed Consent Term, aims to ensure your rights as a participant. Please read carefully and calmly, taking the opportunity to clarify your doubts. If there are questions before or even after signing it, you can clarify them with the researcher. If you prefer, you can take this term and consult your family or others before deciding to participate. There will be no penalty or loss if you do not agree to participate or withdraw your authorization at any time.

Background and Objectives:

This study is relevant to the field of scientific and technological policy as it aims to elucidate the influence of tools arising from the information technology revolution (Big Data and Machine Learning) applied to future studies, which are important tools for generating policies. Thus, a greater understanding of the methodological limitations of these tools will be possible and gives us the possibility to understand in which contexts their use becomes more interesting, generating more complex approaches and adapted to the final objective of the developer. Thus, with the results of this research, it will be possible to indicate more assertive methodologies adapted to the current context of massive data and, thus, generate more effective policies. The general objective of this project is to study and analyze the trends of future studies in science, technology and innovation in the perspective of machine learning and big data, to elucidate if and how these techniques are changing conceptual and methodological approaches for future studies.

Procedures:

By participating in this survey you are being asked to answer some questions, in an online questionnaire, about your experience and the experience of the institution to which you belong with future studies. The questions revolve around the results and support of future studies in decision making combined with the methodologies, tools and techniques used in its development. The estimated response time is approximately 5 minutes. In some cases, a new questionnaire will be required in case the participant needs to complete the answers. The results of the work may be in meetings or scientific journals. The data of this research will be stored digitally, for a period of 5 years after the end of the research, according to Res. CNS 510/16

Discomforts and Risks:

You should not participate in this study if you are inexperienced with developing future studies or if you are not comfortable with answering questions about the topic. The survey does not present predictable risks due to the responses of the survey participants.

Benefits:

The research foresees as indirect benefits the sharing of the knowledge acquired from the collected data, offering the participants an overview of the benefits and limitations of the use of Big Data and Machine Learning tools in future studies.

Monitoring and assistance:

At any time, before, during or until the end of the research, participants may contact the researchers for clarification and assistance on any aspect of the research.

Confidentiality and privacy:

You have the guarantee that your identity will be kept confidential and no identified information will be given to other people who are not part of the research team. When disclosing the results of this study, your name will not be mentioned.

Reimbursement and Indemnity:

You will have no financial compensation and no expense for this research.You will be guaranteed the right to indemnity in the event of any damages resulting from the research when proven under the terms of the current legislation.

Contact:

In case of doubts about the research, if you need to consult this consent record or any other questions, you can contact the researcher Vinicius Muraro da Silva, Institute of Geosciences - State University of Campinas, located at Rua Carlos Gomes, 250, CEP 13083-855 in Campinas-SP, or by e-mail murarosilva@gmail.com or by phone +55(19) 98317 8080.

In case of complaints or complaints about your participation and about ethical issues of the study, you can contact the secretariat of the Research Ethics Committee in Human and Social Sciences (CEP-CHS) of UNICAMP from 8:30 am to 11:30 am and from 1:00 pm to 5:00 pm at Rua Bertrand Russell, 801, Block C, 2nd floor, room 05, CEP 13083-865, Campinas - SP; phone +55(19) 3521-6836; e-mail: cepchs@unicamp.br.

The Research Ethics Committee (CEP)

CEP's role is to evaluate and monitor the ethical aspects of all research involving human beings in Brazil. The National Research Ethics Commission (CONEP) aims to develop regulations on the protection of human beings involved in research. It plays a coordinating role in the network of Research Ethics Committees (CEPs) of the institutions, in addition to assuming the role of an advisory body in the area of research ethics

Consentimento livre e esclarecido:

After receiving clarifications on the nature of the research, its objectives, methods, expected benefits, potential risks and the inconvenience that this may cause, the volunteer accepts to participate by initiating answering the presented form.

Responsabilidade do Pesquisador:

I assure you that I have complied with the requirements of resolution 510/2016 CNS / MS and complementary in the preparation of the protocol and in obtaining this Informed Consent Form. I also assure you that I have explained and provided a copy of this document to the participant. I inform you that the study was approved by the CEP before which the project was presented and by CONEP, when pertinent. I undertake to use the material and data obtained in this research exclusively for the purposes set out in this document or according to the consent given by the participant.

Campinas, September 10th, 2020.

Vinícius Muraro da Silva

Vinicius Muraro da Silva
PhD Candidate

ANNEX 2 - Survey Questionnaire

Future of Future Studies

Introduction

Future studies have a key role in generating policies, strategies, and decision-making in many research fields, in an academic, corporate, and public sector. Methodological trends in future studies, especially foresight, have pointed to IT tools, Big Data, and Machine Learning to support practitioners in carrying out their studies. However, the benefits and limitations of its use are still not clear enough.

The central question about the use of Big Data and Machine Learning tools in future studies is to understand the impact of these mechanisms on bringing a new perspective of conceptual and methodological approaches. This survey aims to envision the future of Future Studies regarding the application of Big Data and Machine Learning tools and techniques and inform the scientific, policy, and practitioners communities by assessing the opinions of researchers from all over the world.

Your contribution

We are asking you to share your perceptions regarding the future of Future Studies regarding the application of Big Data and Machine Learning tools.
The questionnaire will take only 5 minutes of your time to complete.
The questionnaire will be available for completion up to October 26th, 2020.

Confidentiality and Privacy

This survey is for research purposes only. By answering the questionnaire, you give us your consent for the use of any data you provide. Any individual responses will not be identified. The survey results will be submitted to a peer-reviewed journal, and we will send the article to you once published. If you have any queries about this survey, please email viniciusmuraro@ige.unicamp.br. You can access here our consent term.

This work is financed by FAPESP - The São Paulo Research Foundation and conducted by the researcher Vinicius Muraro da Silva. It is part of the InSySPo Project (System Innovation: Organizational Strategy, Research, and Innovation Policy Governance) at the Science and Technology Policy Department (DPCT) at the University of Campinas (Unicamp), Brazil.

UNICAMP DPCT InSySPo FAPESP

Future of Future Studies

Big Data refers to using a massive set of data in terms of acquiring, managing, and processing information. In this work, Big Data tools and techniques are defined as computational tools that work on data mining, data collection, data storing, data cleaning, data analysis, or data consumption.
Machine learning is defined "field of study that gives computers the ability to learn without being explicitly programmed (Samuel, 1959), finding patterns in sets of training data and applying the learned pattern in classifications of new events.

* 1. Do you have experience as a futurist, practitioner, or future studies supported by ICT, Big Data, or Machine Learning techniques?

◯ Yes

◯ No

UNICAMP DPCT InSySPo FAPESP

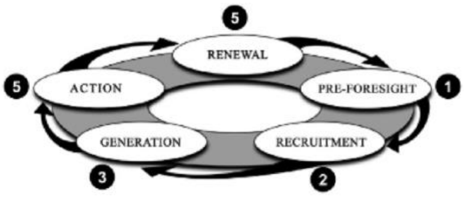
Future of Future Studies

Instructions

You are being asked to answer 12 projections about the future impacts of Big Data and Machine Learning tools and techniques for future studies in 2025. For each projection you should evaluate:
- The **Expected Probability** that the projection occurs by 2025;
- The **Desirability** of that projection occurring by 2025 and;
- The **Impact on Foresight Industry** that the projection will have if it occurs by 2025 (as Foresight Industry, we refer to all the stakeholders involved in the conceptualization, development, and implementation of future studies for science, technology, and innovation)

The projections are divided according to the foresight stages (Miles, 2002; Popper, 2008) presented in the figure below: Pre-Foresight, Recruitment, Generation, Action, and Renewal.
At each stage, it is possible to leave comments on the projections presented on that page.

Figure 1: Foresight Process by Miles (2002)



To start the survey, please answer the following questions:

2. Do you believe that Big Data and Machine Learning tools and techniques will reduce uncertainty related to future studies in 2025?

- ◯ Yes
- ◯ No

Future of Future Studies

Pre-Foresight

Please, evaluate the projections below in terms of Expected Probability of occurring in 2025, Desirability to occur in 2025, and Impact on Foresight Industry if it occurs in 2025.

3. **Projection 1:** Future-oriented activities' objectives will be easily defined using Big Data and Machine Learning tools for its development in 2025.

	1 - very low	2- low	3 - neither low or high	4 - high	5 - very high
Expected Probability	◯	◯	◯	◯	◯
Desirability	◯	◯	◯	◯	◯
Impact on Foresight Indutry	◯	◯	◯	◯	◯

4. **Projection 2:** The possibility of using Big Data and Machine Learning tools will make the methodological choice for future studies easier in 2025.

	1 - very low	2 - low	3 - neither low or high	4 - high	5 - very high
Expected Probability	◯	◯	◯	◯	◯
Desirability	◯	◯	◯	◯	◯
Impact on Foresight Industry	◯	◯	◯	◯	◯

5. Comments (optional)

UNICAMP DPCT InSySPo FAPESP

Future of Future Studies

Recruitment

Please, evaluate the projections below in terms of Expected Probability of occurring in 2025, Desirability to occur in 2025, and Impact on Foresight Industry if it occurs in 2025.

6. **Projection 3:** Data Scientists or coding, analyzing, and data visualization skills will be required to develop consistent future studies in 2025.

	1 - very low	2 - low	3 - neither low or high	4 - high	5 - very high
Expected Probability	◯	◯	◯	◯	◯
Desirability	◯	◯	◯	◯	◯
Impact on Foresight Industry	◯	◯	◯	◯	◯

7. **Projection 4:** The relevant data for future-oriented activities will be less time-consuming and easily accessed using Big Data tools in 2025.

	1 - very low	2 - low	3 - neither low or high	4 - high	5 - very high
Expected Probability	◯	◯	◯	◯	◯
Desirability	◯	◯	◯	◯	◯
Impact on Foresight Industry	◯	◯	◯	◯	◯

8. **Projection 5:** The quality of future-relevant data will be significantly enhanced by the application of Big Data and Machine Learning tools or techniques in foresight studies in 2025.

	1 - very low	2 - low	3 - neither low or high	4 - high	5 - very high
Expected Probability	◯	◯	◯	◯	◯
Desirability	◯	◯	◯	◯	◯
Impact on Foresight Industry	◯	◯	◯	◯	◯

9. Comments (optional)

UNICAMP DPCT InSySPo FAPESP

Future of Future Studies

Generation

Please, evaluate the projections below in terms of Expected Probability of occurring in 2025, Desirability to occur in 2025, and Impact on Foresight Industry if it occurs in 2025.

10. **Projection 6:** The quality of future-relevant data analysis will be significantly enhanced by the adoption of Big Data and Machine Learning tools and techniques in foresight studies in 2025.

	1 - very low	2 - low	3 - neither low or high	4 - high	5 - very high
Expected Probability	○	○	○	○	○
Desirability	○	○	○	○	○
Impact on Foresight Industry	○	○	○	○	○

11. **Projection 7:** Big Data and Machine Learning tools and techniques will be critical to support qualitative analysis in foresight activity in 2025.

	1 - very low	2 - low	3 - neither low or high	4 - high	5 - very high
Expected Probability	○	○	○	○	○
Desirability	○	○	○	○	○
Impact on Foresight Industry	○	○	○	○	○

12. **Projection 8:** The use of Big Data and Machine Learning solutions for future studies will increase the frequency of manipulated (biased) data in 2025.

	1 - very low	2 - low	3 - neither low or high	4 - high	5 - very high
Expected Probability	○	○	○	○	○
Desirability	○	○	○	○	○
Impact on Foresight Industry	○	○	○	○	○

13. Comments (optional)

UNICAMP DPCT InSySPo FAPESP

Future of Future Studies

Action

Please, evaluate the projections below in terms of Expected Probability of occurring in 2025, Desirability to occur in 2025, and Impact on Foresight Industry if it occurs in 2025.

14. **Projection 9:** Big Data and Machine Learning tools and techniques will increase future studies' embeddedness to strategic planning or decision-making in 2025.

	1 - very low	2 - low	3 - neither low or high	4 - high	5 - very high
Expected Probability	○	○	○	○	○
Desirability	○	○	○	○	○
Impact on Foresight Industry	○	○	○	○	○

15. **Projection 10:** Big Data and Machine Learning tools and techniques will increase the accuracy of decision-making based on foresight in 2025.

	1 - very low	2 - low	3 - neither low or high	4 - high	5 - very high
Expected Probability	○	○	○	○	○
Desirability	○	○	○	○	○
Impact on Foresight Industry	○	○	○	○	○

16. Comments (optional)

UNICAMP DPCT InSySPo FAPESP

Future of Future Studies

Renewal

Please, evaluate the projections below in terms of Expected Probability of occurring in 2025, Desirability to occur in 2025, and Impact on Foresight Industry if it occurs in 2025.

17. **Projection 11:** Evaluating and monitoring future-oriented activities will be easily reached when using Big Data and Machine Learning tools in 2025.

	1 - very low	2 - low	3 - neither low or high	4 - high	5 - very high
Expected Probability	◯	◯	◯	◯	◯
Desirability	◯	◯	◯	◯	◯
Impact on Foresight Industry	◯	◯	◯	◯	◯

18. **Projection 12:** Future studies' objectives will be easily reached when Big Data and Machine Learning tools are used for its development in 2025.

	1 - very low	2 - low	3 - neither low or high	4 - high	5 - very high
Expected Probability	◯	◯	◯	◯	◯
Desirability	◯	◯	◯	◯	◯
Impact on Foresight Industry	◯	◯	◯	◯	◯

19. Comments (optional)

Future of Future Studies

Demographic Questions

Thank you for taking part in this survey.

Below there are some demographic questions. Please feel free to answer or not.

20. What kind of institution are you affiliated with?

- [ ] University
- [ ] Consultancy
- [ ] Research Center or Institution
- [ ] Company
- [ ] Government
- [ ] Other (please specify)

21. How many years of experience do you have in future-oriented activities?

- ○ Under 1 year
- ○ 1-5 years
- ○ 5-10 years
- ○ 10-15 years
- ○ 15+ years

22. What is your region?

- ○ Africa
- ○ Asia
- ○ Europe
- ○ Latin America and the Caribbean
- ○ Northern America
- ○ Oceania