UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Civil, Arquitetura e Urbanismo

GUILHERME DAYRELL MENDONÇA

ALGORITMOS INTELIGENTES APLICADOS À PREDIÇÃO DE ATRASOS EM TRANSPORTE AÉREO

INTELLIGENT ALGORITHMS APPLIED TO THE PREDICTION OF AIR FREIGHT TRANSPORTATION DELAY

CAMPINAS
2021

GUILHERME DAYRELL MENDONÇA

ALGORITMOS INTELIGENTES APLICADOS À PREDIÇÃO DE ATRASOS EM TRANSPORTE AÉREO

INTELLIGENT ALGORITHMS APPLIED TO THE PREDICTION OF AIR FREIGHT TRANSPORTATION DELAY

Dissertação de Mestrado apresentada a Faculdade de Engenharia Civil, Arquitetura e Urbanismo da Unicamp, para obtenção do título de Mestre em Engenharia Civil, na área de Transportes.

Master Dissertation presented to the Faculty of Civil Engineering, Architecture and Urban Planning of the University of Campinas in partial fulfillment of the requirements for the degree of Master in Civil Engineering, in the area of Transportation.

Orientador: Prof. Dr. Orlando Fontes Lima Junior

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO DEFENDIDA PELO ALUNO GUILHERME DAYRELL MENDONÇA E ORIENTADO PELO PROF. DR. ORLANDO FONTES LIMA JUNIOR.

CAMPINAS

2021

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Rose Meire da Silva - CRB 8/5974

M523i	Mendonça, Guilherme Dayrell, 1984- Intelligent algorithms applied to the prediction of air freight transportation delay / Guilherme Dayrell Mendonça. – Campinas, SP : [s.n.], 2021. Orientador: Orlando Fontes Lima Junior. Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia Civil, Arquitetura e Urbanismo. 1. Mineração de dados (Computação). 2. Aprendizagem de máquina. 3. Cadeia de suprimentos. I. Lima Junior, Orlando Fontes, 1958-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Civil, Arquitetura e Urbanismo. III. Título.

<u>Informações para Biblioteca Digital</u>

Título em outro idioma: Algoritmos inteligentes aplicados à predição de atrasos em transporte aéreo
Palavras-chave em inglês:
Data mining
Machine learning
Supply chains
Área de concentração: Transportes
Titulação: Mestre em Engenharia Civil
Banca examinadora:
Orlando Fontes Lima Junior [Orientador]
Stanley Robson de Medeiros Oliveira
Andrea Leda Ramos de Oliveira
Data de defesa: 27-08-2021
Programa de Pós-Graduação: Engenharia Civil

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0001-5612-5038
- Currículo Lattes do autor: http://lattes.cnpq.br/4223748970745787

UNIVERSIDADE ESTADUAL DE CAMPINAS

FACULDADE DE ENGENHARIA CIVIL, ARQUITETURA E URBANISMO

ALGORITMOS INTELIGENTES APLICADOS À PREDIÇÃO DE ATRASOS EM TRANSPORTE AÉREO

Guilherme Dayrell Mendonça

Dissertação de Mestrado aprovada pela Banca Examinadora, constituída por:

Prof. Dr. Orlando Fontes Lima Junior
Presidente e Orientador/FEC UNICAMP

Prof. Dr. Stanley Robson de Medeiros Oliveira
EMBRAPA-CNPTIA

Prof. Dra. Andrea Leda Ramos de Oliveira
FEAGRI/ UNICAMP

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação e na Secretaria do Programa da Unidade.

Campinas, 27 de Agosto de 2021

DEDICATÓRIA

Aos meus pais Dilberto (*in memoriam*) e Maria Angela

À minha esposa Laura e minha filha Beatriz

Aos meus queridos Tia Diva, Tio Ciro e minha prima Marcioni que partiram desse mundo vítimas da Covid-19

AGRADECIMENTOS

Agradeço a Deus por me ensinar e me guiar todos os dias com sua Palavra de esperança e fé.

Aos meus pais por sempre me apoiarem nos desafios e conquistas da minha vida de forma incondicional.

A minha esposa Laura, por ser minha rocha e porto seguro nas tempestades da vida e ser minha companheira acadêmica.

A minha filha Beatriz que nasce junto com esse mestrado para me alegrar incondicionalmente em um amor Ágape.

Ao meu irmão por estar sempre ao meu lado e ser exemplo em que me inspiro.

Ao meu orientador, Professor Orlando que me acolheu como seu orientando, me desenvolveu academicamente e abriu as portas do LALT (Laboratório de Aprendizagem em Logística e Transportes) para mim.

Aos colegas do LALT por todo o conhecimento e convivência que enriqueceram minha vida profissional.

A todos os meus amigos e parentes que vibram com as minhas conquistas acadêmicas.

RESUMO

A indústria de carga aérea está enfrentando cada vez mais desafios operacionais devido à competição global mais acirrada e aos requisitos de nível de serviço mais elevados dos clientes. As técnicas de aprendizado de máquina estão sendo consequentemente aplicadas como uma abordagem de gerenciamento de risco da cadeia de suprimentos da aviação (SCRM) a fim de prever atrasos, reduzir a incerteza operacional e reduzir custos. O objetivo da pesquisa é avaliar se o uso de técnicas de aprendizado de máquina contribui para uma melhor previsão de atrasos de remessas por transporte aéreo de forma a otimizar o desempenho da capacidade de controle da cadeia de abastecimento internacional. Para tanto são testadas diferentes classes de algoritmos na fase de mineração de dados do KDD (Knowledge Discovery in Databases): Support Vector Machine, Random Forest, Artificial Neural Networks e K-Nearest Neighbors. O algoritmo Random Forest obteve o melhor resultado de acuracidade com 86% para opção de teste de validação cruzada após um procedimento de balanceamento de classe combinada. No geral, a pesquisa também acrescenta à literatura atual, uma vez que os dados de transporte e de fornecedor são usados em uma aplicação específica de aprendizado de máquina.

Palavras Chave: Mineração de Dados, Aprendizagem de Máquina, Cadeia de Suprimentos

ABSTRACT

The air freight cargo industry is increasingly facing operational challenges due to tougher global competition and higher service level requirements from customers. Machine Learning techniques are consequently being applied as an aviation supply chain risk management (SCRM) approach in order to predict delays, reduce operational uncertainty and reduce costs. The objective of the research is to evaluate whether the use of machine learning techniques contributes to a better prediction of air transport shipment delays in order to optimize the performance of the international supply chain control capacity. For that, different classes of algorithms are tested in the data mining phase of KDD (Knowledge Discovery in Databases): Support Vector Machine, Random Forest, Artificial Neural Networks and K-Nearest Neighbors. The Random Forest algorithm achieved the best result with accuracy of 86% in the cross validation test scenario after a combined class balancing procedure. Overall, the research also adds to the current literature as both transport and supplier data are used in a specific machine learning application.

Keywords: Data Mining, Machine Learning, Supply Chains

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

1 INTRODUCTION

Supply chain management can be defined as a “network of organizations that are involved, through upstream and downstream linkages, in the different processes and activities that produce value in the form of products and services in the hands of the ultimate consumer” (CHRISTOPHER, 1992, p. 320). The entire process of logistics can be described as below:

> Deals with moving of materials into, through, and out of a firm, can be divided into three parts: (1) *inbound logistics*, which represents the movement and storage of materials received from suppliers; (2) *materials management*, which covers the storage and flows of materials within a firm; and (3) *outbound logistics* or *physical distribution*, which describes the movement and storage of products from the final production point to the customer (FARAHANI; REZAPOUR; KARDAR, 2011, p.11).

The inbound and outbound logistics flows are mainly supported by transportation activities. Transport refers to moving product from one location to another as it proceeds upstream to the customer and represents a significant component of the costs incurred by most chains. Supply chains use a combination of the following modes of transport: air, express parcel carriers, truck, railroad, maritime, pipeline and intermodal (SUNIL; CHOPRA, 2016). The globalization of the economy, which has been confirmed by the trade barriers decline and the falling transport, communication and coordination costs, has changed the role of the manufacture procedures from being determined in few plants to being fragmented in different facilities and in different countries and has considerably increased the trade rate. This has fostered a tight raise of trade flows, affecting the logistics activities, particularly transport (KHERBASH; MOCANA 2015).

In particular, the air cargo industry plays a key role in global transportation due to its expedite service characteristics. The air freight market was estimated to be worth $92.81 billion in 2018 and is expected to increase to $183.16 billion by 2025 (SANDERS, 2020). Freight is a direct representation of the health of the global economy and while airfreight may be a tiny proportion of all freight by tonnage (2–3%), nonetheless it can represent a significant amount of countries’ total imports and exports by value, typically between 35–40% in many advanced economies (GSF, 2015).

Many industries benefit from the faster alternative of this transportation mode on their daily business operations. Airfreight is ideally suited to the Just in Time (JIT) manufacturing processes favored by automotive manufacturers. Just in time manufacturing can be defined as “an approach to achieving excellence in a manufacturing company based on the continuing elimination of waste (waste being considered as those things which do not add value to the product such as supply chain deviations)” (WALLACE; DOUGHERTY, 1987). In the case of a manufacturing crisis, such as a parts recall, the rapid reactivity of air cargo can get manufacturing back on track in a short time, saving automotive companies hundreds of thousands – if not millions – of dollars in lost time and productivity. Air cargo is therefore an important link in a JIT supply chain, enabling automotive manufacturers to reduce downtime on the assembly line and thereby maintain profits.

The air freight cargo industry is increasingly facing operational challenges due to tougher global competition and higher service level requirements from customers. Transportation delays may represent considerable additional costs to the supply chain. According to the Zhang and Figliozzi (2009) survey, transport delays were cited by 42,9% of respondents to impact the administration workload and costs and by approximately 28% to increase transportation and inventory costs. Higher administrative costs related to customer service, communication, documentation and tracking were highlighted by wholesalers. Wholesalers and manufacturers that import supplies and raw material were also concerned about higher inventory costs caused by longer lead times. In general, importers also showed a higher concern about the impact of delays on promotions and sales plans as well as costs associated to custom procedures and inspections (ZHANG; FIGLIOZZI, 2009).

From the use of data analysis techniques to interpret a growing database, much can be done to identify trends and anticipate changes such as transportation delays that impact the business as a whole (WALLER; FAWCETT, 2013a). The analytics capability has a direct effect on supply chain agility and competitive advantage. Organizational flexibility also plays an important moderation role on the path unifying the agility and competitive dimensions (DUBEY *et al.*, 2019d). The ability to predict delays and act in advance to avoid the abovementioned impacts is bound to become a competitive advantage in supply chain management as it contributes to cost and stock out occurrence reductions (specially for JIT production environments of the automotive industry). However, the full application of artificial intelligence and data analytics is still to be achieved as companies progressively implement projects in this area. The next section further describes the research problem definition and objectives to tackle the current challenges in delay prediction within the Brazilian context of supply chain operations.

2 PROBLEM DEFINITION AND OBJECTIVES

Artificial intelligence and data analytics technologies are driving the development of transformative business models with new platforms that automate processes, match demand and supply, dynamically define pricing and make real-time decisions (AKTER *et al*., 2020). These methodologies have an intrinsic correlation with the enhanced use of information datasets to gain further insights into daily decision making and increase Supply Chain Risk Management capabilities. Supply Chain Risk Management (SCRM) encompasses a wide variety of strategies to identify, assess, mitigate and monitor events or conditions which might have an impact, mostly adverse, on any part of a supply chain (BARYANNIS *et al*., 2019b). Sanchez-Rodrigues, Potter and Naim (2010) indicate that the main drivers impacting the sustainability of transport operations are delays, variable demand/poor information, delivery constraints and insufficient supply chain integration.

These drivers are directly related to the logistics triad concept: the set of relationships between the supplier of the goods, the customer for the goods and the logistics provider (or carrier) (SANCHEZ-RODRIGUES *et al*., 2008). The consequence of issues in these relationships is the reduction of the efficiency of transport operations. In general, the current concept of inbound logistics management in Brazil is not based on the adoption of technologies that monitor the supply chain in real time and propose prescriptive or predictive solutions to mitigate risks and control operational instabilities (QUEIROZ; TELLES, 2018). Additional costs are incurred from this inefficiency to predict delays and react on a timely manner.

The main research question which will be answered can be described as follows: **The improvement of predictability of international air shipment deliveries increase the ability to control supply chains risks and improve their performance?**

Based on that question, the research hypothesis is formalized below:

Research Hypothesis: The use of supervised learning algorithms can contribute to improve predictability of international shipment delays and improve the supply chain performance.

The general objective of the research is to evaluate whether the use of machine learning techniques contributes to a better prediction of air transport shipment delays in order to optimize the performance of the international supply chain control capacity. For that, different classes of algorithms are tested, namely: Support Vector Machine, Random Forest, Artificial Neural Networks and K-Nearest Neighbor algorithms. By anticipating the supply chain delay before picking up the cargo with the supplier, there is an improvement in supply chain efficiency and cost reduction.

In addition, this research aims to contribute to the expansion of studies related to the application of intelligent algorithms in the Brazilian context of logistics operations. Due to the growing importance of research in this area, it is necessary to expand the application of these concepts in the Brazilian reality as a way to contribute to the evolution of knowledge of the theme in the region.

In summary, this research aims to provide supply chain practitioners and scholars a new approach regarding air freight delay management. The key element is to combine supplier, customer and transportation operational data to identify patterns which indicate higher probability of delay occurrence. Based on this iterative analytical assessment, transportation performance is bound to improve as the main bottlenecks are identified and solved proactively before they might occur. In the next section, the main concepts that underpin this research investigation are presented.

3 LITERATURE REVIEW

Artificial Intelligence is a broad area of research and investigation in the contemporary business world. This research aims to provide state-of-the-art literature review of artificial intelligence application to Supply Chain Management. Main artificial intelligence and big data analytics works applied to the context of supply chain operations using the SCOR®[1] model as reference are reviewed in Section 3.1. Data mining and intelligent algorithms concepts utilized in this research methodology are described in Section 3.2. Further discussion is carried out on how intelligent algorithms can support Supply Chain Risk Management (SCRM) practices in order to improve operations results with special focus on transportation and air freight management in Section 3.3. To sum up, the literature review is structured from a broader to a more specific perspective in order to provide the background associated with the research topic: intelligent algorithms applied to delay prediction and risk mitigation.

3.1. MAIN AREAS OF ARTIFICIAL INTELLIGENCE APPLICATION IN SCM ACCORDING TO THE SCOR REFERENCE MODEL

3.1.1. General Context of AI application in SCM

Research in Artificial Intelligence applied to the Supply Chain area is rapidly growing not only in supply and demand management but also in other application areas such as operations optimization. According to Russell and Norvig (1995), Artificial Intelligence (AI) is known for its ability to think like humans, act like humans, think rationally, and act rationally. Thus, with respect to these distinctive features, AI can be further classified into a number of sub-fields: (1) artificial neural networks (ANN) and rough set theory ("thinking humanly"); (2) machine learning, expert systems, and Genetic Algorithms ("acting humanly"); (3) fuzzy logic ("thinking rationally"); and (4) agent-based systems ("acting rationally"). In specific, the Machine Learning functionality is primarily intended to enable computers to learn without necessarily being programmed for such activity. Its application has already been used to predict collaborative behavior in supply chain management (MIN, 2010).

"The explosively growing, widely available, and gigantic body of data makes our time truly the data age" (HAN; KAMBER; PEI, 2011, p. 13). The development of Internet of

[1]This research considered the AI literature review in regards also to the six SCOR® (Supply Chain Operations Reference) management fields, namely: Plan, Source, Make, Deliver, Return and Enable. The business processes proposed by the SCOR® model embraces various tiers along the supply chain and include a set of management practices recognized by companies in many industries (LIMA-JUNIOR; CARPINETTI, 2016; DIDEHKHANI; JASSBI; PILEVARI, 2009).

Things (IoT) devices combined with greater data storage capacity expanded data analytics and artificial intelligence applications. The term Big Data was first coined by Cox and Ellsworth (1997) in an article that indicated the eminent limitation of information storage on hardware resources caused by the exponential growth of information available in computer systems. In addition, the term Big Data Analytics (BDA) was defined as the application of advanced techniques of data mining, statistical analysis and predictive analysis of very large databases aiming at generating value to the organizational decision-making process (TIWARI; WEE; DARYANTO, 2018).

Artificial intelligence methodologies have intrinsic correlation mainly on the enhanced utilization of datasets of information to gain further insights into the daily decision making. A study conducted by DHL in conjunction with IBM identified the latest applications and best practices (DHL, 2021). Applications include end user support solutions, voice interaction solutions with end consumers, machine learning applied to social networks, creation of expert content, identification of information standards, robots in the retail operation, autonomous vehicles, assistance robots in manufacturing, predictive management of demand, among others. The DHL Logistics Trend Radar is currently considered one the main industry benchmarks regarding future technological trends in logistics and supply chain management. It is noteworthy to mention that according to this index, Artificial Intelligence will have high impact within 5 years time range in the logistics industry. The main important output from this study is the confirmation that predictive logistics stands out as one of the most promising areas within the AI scope of development in Supply Chain Management, as described below:

> Predictive logistics remains the most important AI application for industry professionals, given the abundance of supply chain data from which to draw predictive insights. For instance, with double-digit e-commerce growth increasing last-mile diversity and complexity, AI is making strides in dynamic route optimization, managing numerous variables such as delivery time windows, ad hoc pickups and traffic patterns to generate accurate time-window predictions for customers. As AI becomes more intelligent, predictive technology could take logistics players a step further into the territory of anticipatory delivery models, supplying goods to customers before they even realize what is needed (DHL, 2021).

The adoption of such tools presupposes a more advanced stage of technological development and an entrepreneurial culture of investment in innovation. In general, the ability of a company to promote innovation in logistics is positively correlated with the generation of competitive advantage in the market in which it operates (GRAWE, 2009). According to Tiwari, Wee and Daryanto (2018), the biggest challenge for Supply Chain professionals today is to find the best way to deal with the growing availability of large information bases. Among the possible benefits of the use of Big Data Analytics tools is the construction of an agile supply chain with greater capacity to monitor social media, events, static and dynamic information points, thus increasing the possibility of performing actions to adapt the operation. Hofmann and Rutschmann (2018) researched large retailers such as Amazon, which have implemented the technique of advance shipment of products based on prescriptive models of demand forecasting. One of the main benefits is to avoid the recurrence of finished products stock peaks through proactive mitigation actions. Furthermore, important information can be provided on the traffic conditions that would reduce the fuel consumption of delivery vehicles, thereby increasing the logistical sustainability character. Finally, it can be of great importance the identification, management and mitigation of risks caused by externalities such as natural disasters or supply disruptions caused by unstable social situations (TIWARI; WEE; DARYANTO, 2018).

Bowers, Petrie and Holcomb (2017) argued that for a company to benefit from the practice of data analytics, it is necessary to reduce the reaction time after receiving the information. As an example, Hanesbrands Inc., a US capital goods company, aimed at adjusting its Machine Learning algorithms to better react when a supplier rescheduling occurs. The other possible reactions also range from the adjustment of the freight rate to the re-sequencing of production schedules to avoid a line stop. However, these cases are exceptions, since few companies today, according to the authors, are able to transform the high availability of information into competitive advantage and value for the end customer.

Within this context of value creation, according to Brinch (2018), the use of Big Data tools offers three possible dimensions of analysis: (1) Discovery Value that describes the company ability to structure a reliable database (2) Creation Value which represents the capacity to transform information into a source of decision-making and, (3) Capture Value in which the company achieves an improvement in the operational or financial results through the use of Big Data. Based on these concepts it is possible to create an evaluation model of how the value is being managed in the Big Data Analytics process of a given company.

In addition to companies, BDA can also be used for operational decisions related to the humanitarian supply chain, increasing coordination and integration by providing greater visibility of the capabilities of each agent in the temporary flow of supplies (DUBEY *et al.*, 2018; PAPADOPOULOS *et al.*, 2017). Another possible positive impact advocated by Hazen *et al.* (2018), is the possibility of transforming the supply chain into a more sustainable organization by broadening the field of analysis for environmental and social issues of the process.

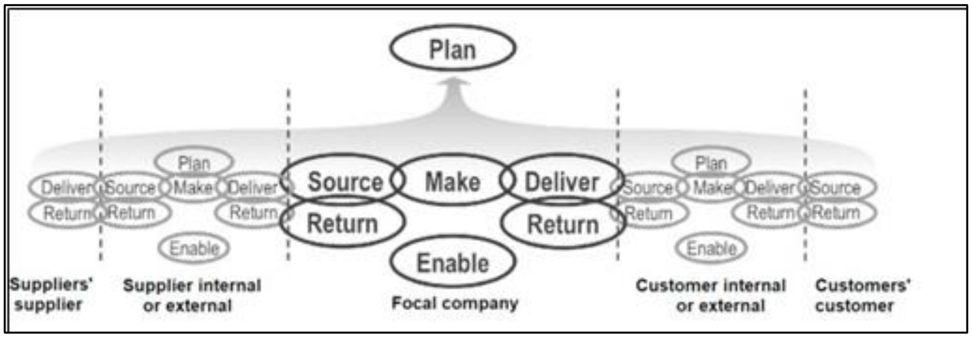
The formation of the big data database can be obtained from different sources, such as Internet of Things (IoT) products or machines. The use of these sources has increased the companies' ability to measure operational performance since it allows real-time analysis along the supply chain (DWEEKAT; HWANG; PARK, 2017). For example, IoT applied to cargo vehicles is becoming a key source of information on drivers' conduct and the relationship thereof with fuel consumption and vehicle depreciation (HOPKINS; HAWKING, 2018).

AI can create value in areas such as consumer behavior, supply chain visibility and transparency, operational and maintenance efficiency, information management, responsiveness, and the generation of new business opportunities based on market trends. Conversely, its development limitations are usually related to Information Technology infrastructure, human resources and knowledge, and openness to information exchange in the supply chain (KACHE; SEURING, 2017). The next section provides more specific information on which areas of Supply Chain Management AI and BDA are mostly applied according to the SCOR model process reference.

3.1.2. AI / BDA in the context of SCOR model

This research has analyzed the application of AI and BDA in the context of the SCOR operation model. According to Figure 1, the SCOR model main areas are depicted:

Figure 1: Major management processes proposed by the SCOR® model



Source: (SUPPLY CHAIN COUNCIL, 2012)

> The Plan processes aggregate demand and supply to develop a course of action which best meets sourcing, production, and delivery requirements. Source processes procure goods and services to meet planned or actual demand. Make processes transform product to a finished state to meet planned or actual demand. Deliver processes provide finished goods and services to meet planned or actual demand, typically including order management, transportation management, and distribution management. Return processes are associated with returning or receiving returned products for any reason. These processes extend into post-delivery customer support. Finally, Enable processes are associated with the management of the supply chain. These processes include management of: business rules, performance, data, resources, facilities, contracts, supply chain network management, managing regulatory compliance and risk management (SUPPLY CHAIN COUNCIL, 2012).

Tables 2 and 3 illustrate, respectively, all AI and BDA papers divided by research methodology and area of classification regarding the SCOR model, namely: Enable, Source, Make, Deliver, Plan and Return. While AI papers are mostly concentrated on Model application, BDA research spans over other areas such as Survey, Empirical and Theoretical investigation and is more equally distributed. This may well show that BDA is still developing its conceptual and practical background as a recent area of study in supply chain.

Regarding SCOR classification, Plan and Enable areas are the most representative. Enable ranks first in BDA papers (Table 2) and second for AI publications (Table 1). Conversely, Plan ranks first in AI papers (Table 1) and second in BDA only studies (Table 2). Overall, the SCOR Plan area is the most representative if we consider both BDA and AI papers. The focus of this research is on the Enable area which encompasses the risk management activity within the SCOR model.

Table 1 Most relevant research fields (AI) regarding SCOR model categories

Research Method / Authors	Enable	Source	Make	Deliver	Plan	Return
Case Study	**5**	**2**	**1**	**3**	**4**	**0**
Borade, A. B. and Sweeney, E., 2015	0	1	0	1	1	0
Ma, H. *et al.* , 2018	1	0	1	0	0	0
Mahroof, K. , 2019	1	0	0	1	1	0
Orji, I. J. and Wei, S., 2015	1	1	0	0	0	0
Slimani, I. , 2017	0	0	0	1	1	0
Tsang, Y. P. *et al.*, 2018	1	0	0	0	1	0
Urciuoli, L., and Hintsa, J., 2018	1	0	0	0	0	0
Framework	**7**	**2**	**1**	**0**	**6**	**1**
Garg, V. K. and Viswanadham, N., 2010	1	0	0	0	0	1
Kartal, H. *et al.*, 2016	0	1	0	0	1	0
Piramuthu, S., 2005a	0	0	0	0	1	0
Piramuthu, S., 2005b	1	1	0	0	1	0
Piramuthu, S., 2005c	1	0	0	0	0	0
Pontrandolfo, P. *et al.*, 2002	0	0	1	0	1	0
Rampersad, G., 2020	1	0	0	0	0	0
Siurdyban, A. and Moller, C., 2012	1	0	0	0	1	0
Tripathi, S. and Gupta, M., 2020	1	0	0	0	0	0
Xu, Z. Y. *et al.*, 2006	1	0	0	0	1	0
Model	**26**	**16**	**6**	**22**	**42**	**0**
Aggarwal, A. K. and Dave, D. S., 2018	0	0	0	0	1	0
Amirkolaii, K. N. , 2017	0	0	0	1	1	0
Carbonneau, R. *et al.*, 2007	0	0	0	1	1	0
Carbonneau, R. *et al.*, 2008	0	0	0	1	1	0
Carbonneau, R. *et al.*, 2012	0	0	0	1	1	0
Castillo-Villar, K. K. and Herbert-Acero, J. F., 2013	1	0	1	1	1	0
Cavalcante, I. M. *et al.*, 2019	1	1	0	0	1	0
Chaharsooghi, S. K. *et al.*, 2008	0	0	0	0	1	0
Chen, C. and Xu, C., 2018	1	1	0	0	0	0
Chi, H. M. *et al.*, 2007	0	1	0	1	1	0
Curcio, D. *et al.*, 2007	0	0	0	1	1	0
De Santis, R. B. *et al.* , 2017	0	1	0	0	1	0
Efendigil, T. *et al.*, 2009	0	0	0	1	1	0
Fu, J. and Fu, Y., 2015	0	1	0	0	1	0
Giannakis, M. and Louis, M, 2011	1	1	1	1	1	0
Giannoccaro, I. and Pontrandolfo, P., 2002	0	1	1	1	1	0
Guosheng, H. and Guohong, Z., 2008	1	1	0	0	0	0
Gyulai, D. *et al.*, 2018	0	0	1	0	1	0
Hiromoto, R. E. *et al.*, 2017	1	0	0	0	0	0
Hong, G. H. and Ha, S. H, 2008	1	1	0	0	1	0
Ilie-Zudor, E. *et al.* , 2015	1	0	0	1	1	0
Jafarzadeh-Ghoushchi, S. and Rahman, M.N.A. , 2016	1	0	0	1	1	0
Kar, A. K., 2015	1	1	0	0	0	0
Kazemi, A.; Fazel Zarandi, M. H., 2008	1	0	1	0	1	0
Kiekintveld, C. *et al.* , 2009	0	0	0	0	1	0
Kong, F. and LI, J., 2018	0	0	0	0	1	0
Kumar, D. *et al.*, 2013	0	1	0	0	1	0
Mojaveri, H. R. S. *et al.*, 2009	0	0	0	0	1	0
Mokhtarinejad, M. *et al.*, 2015	0	0	0	1	1	0
Moraga, R. *et al.*, 2011	1	0	0	0	1	0
Park, Y. B. el al., 2018	1	0	0	0	1	0
Pereira, M. M. *et al.*, 2018	1	1	0	1	1	0
Raut, R.D. *et al.*, 2017	1	1	0	0	1	0
Shahrabi, J. *et al.*, 2009	0	0	0	1	1	0
Shokouhyar, S. *et al.*, 2019	1	0	0	0	0	0
Singh, L. P. and Challa, R. T, 2016	1	0	0	1	1	0
Slimani, I. *et al.*, 2015	0	0	0	0	1	0
Sun, Z.-L. *et al.*, 2008	0	0	0	1	1	0
Tse, Y. K.*et al.*, 2009	1	0	0	0	1	0
Vahdani, B. *et al.*, 2014	0	0	0	0	1	0
Valluri, A. *et al.* , 2009	1	0	0	0	1	0
Wanke, P. *et al.*, 2017	0	0	0	1	1	0
Wieczorek, L. and Ignaciuk, P., 2018	1	1	0	1	1	0
Wong, J. T. *et al.*, 2012	0	0	1	0	1	0
Wu, P.J. *et al.*, 2018	1	0	0	1	1	0
Yuen, J. S. M. *et al.* , 2018	1	0	0	1	1	0

Zhang, H. *et al.*, 2004	1	1	0	0	0	0
Zhang, R. *et al.* , 2016	1	1	0	0	0	0
Zhu, Y. *et al.*, 2017	1	0	0	0	0	0
Total of Papers	**37**	**20**	**8**	**24**	**51**	**1**

Source: Own Elaboration

Table 2 Most relevant research fields (BDA) regarding SCOR model categories

Research Method / Author	Enable	Source	Make	Deliver	Plan	Return
Case Study	**6**	**3**	**0**	**8**	**9**	**1**
Andersson, J.; Jonsson, P. , 2018	0	0	0	1	1	0
Boldt, L. C. et al, 2016	0	0	0	1	1	0
Engelseth, P. and Wang, H., 2018	1	1	0	0	1	0
Gravili, G. *et al.*, 2018	0	0	0	0	1	0
Hopkins, J. and Hawking, P., 2018	1	0	0	1	0	1
Matthias, O. *et al.*, 2017	1	1	0	1	1	0
Moktadir, M. A. *et al.* , 2019	1	0	0	0	0	0
Moretto, A. *et al.*, 2017	1	1	0	0	0	0
Nita, S., 2015	0	0	0	1	1	0
Singh, A. *et al.*, 2018	0	0	0	1	1	0
Yu, L. *et al.* , 2019	0	0	0	1	1	0
Zhan, Y. *et al.*, 2018	1	0	0	1	1	0
Empirical	**7**	**2**	**2**	**4**	**4**	**0**
Ittmann, H. W. , 2015	1	1	0	1	1	0
Niu, B. *et al.* , 2019	1	0	0	1	1	0
Richey Jr, R. G. *et al.*, 2016	1	0	0	0	0	0
Sanders, N. R., 2016	1	1	1	1	1	0
Sodero, A. *et al.*, 2019	1	0	0	0	0	0
Tsao, Y. C. , 2017	1	0	0	1	1	0
Zhong, R. Y. *et al.*, 2015	1	0	1	0	0	0
Framework	**9**	**4**	**2**	**4**	**9**	**5**
Arya, V., 2017	0	1	0	1	1	0
Chavez, R. *et al.*, 2017	0	1	0	0	1	0
Cheng, O. K. M. *et al.*, 2016	1	0	0	0	1	0
Dubey, R., A. *et al.*, 2016	0	1	0	0	0	1
Hu, H. *et al.*, 2014	1	0	0	0	0	0
Ivanov, D. *et al.*, 2019	1	0	0	0	0	0
Jeble, S. *et al.*, 2018	1	0	0	1	1	0
Papadopoulos, T.,*et al.*, 2017	1	0	0	1	1	1
Rehman, M. H. U. *et al.*, 2016	1	0	0	0	1	0
Ren, S. *et al.* , 2019	1	0	1	0	1	1
Rodriguez, L.*et al.*; 2018	1	0	0	0	0	1
Shukla M. and Tiwari, M. K., 2017	0	1	1	1	1	1
Wang, G. *et al.*, 2016	1	0	0	0	1	0
Model	**12**	**5**	**3**	**6**	**12**	**5**
Choi, T.-M., 2018	1	0	0	1	1	0
Côrte-Real, N. *et al.*, 2017	0	1	0	0	1	1
Ehret, M. and Wirtz, J., 2017	1	0	1	0	1	0
Giannakis, M. and Louis, M., 2016	1	1	1	1	1	0
Hofmann, E. , 2017	1	1	0	0	1	0
Jiang, C. and Sheng, Z., 2009	1	0	0	1	1	0
Kaur, H. and Singh S. P. , 2018	0	1	0	0	1	1
Lau, R. Y. K. *et al.*, 2018	0	0	0	1	1	0
Lee, C. K. H., 2017	1	0	0	1	1	0
Prasad, S.; *et al.*, 2018	1	0	0	0	0	1
Simchi-Levi, D. and Wu, M. X., 2018	1	0	0	1	1	0
Bumblauskas, D. *et al.*, 2017	0	0	1	0	0	0
Waller, M. A.and Fawcett, S. E., 2013b	1	0	0	0	0	0
Wamba, S. F. *et al.*, 2017	0	0	0	0	1	0
Wu, K. J. *et al.*, 2017	1	0	0	0	1	1
Wu, P. J. and Lin, K. C., 2018	1	0	0	0	0	0
Zhao, R., Y. *et al.*, 2017	1	1	0	0	0	1
Survey	**15**	**0**	**1**	**0**	**6**	**2**
Chen, D. Q. *et al.*, 2015	0	0	0	0	1	0
Dubey, R. *et al.*, 2019a	1	0	0	0	0	0
Dubey, R. *et al.*, 2019b	1	0	0	0	0	0
Dubey, R. *et al.*, 2019c	1	0	0	0	0	1
Dubey, R. *et al.*, 2019d	1	0	0	0	0	0
Fernando, Y. *et al.*, 2018	1	0	0	0	1	0
Gunasekaran, A. *et al.*, 2017	1	0	0	0	1	0
Gupta, S. *et al.*	1	0	0	0	0	0
Lai, Y. *et al.*, 2018	1	0	0	0	0	0
Mandal, S, 2018	1	0	0	0	1	0
Mandal, S, 2019	1	0	0	0	1	0

Mani, V. *et al.* , 2017	1	0	1	0	1	0
Mikalef, P. *et al.*, 2019	1	0	0	0	0	0
Raut, R. D. *et al.*, 2019	1	0	0	0	0	1
Rossmann, B. *et al.*, 2018	1	0	0	0	0	0
Schoenherr, T. and Speier-Pero, C., 2015	1	0	0	0	0	0
Wamba, S. F. et al, 2020	1	0	0	0	0	0
Theoretical	**6**	**0**	**1**	**1**	**2**	**2**
Albergaria, M., and Jabbour, C. J. C., 2020	1	0	0	0	0	0
Hazen, B. T. *et al.* , 2014	1	0	1	0	0	0
Hazen, B. T. *et al.* , 2016	1	0	0	0	0	0
Hazen, B. T. *et al.*, 2018	0	0	0	0	1	0
Hofmann, E. and Rutschmann, E., 2018	0	0	0	1	1	0
Li, J., *et al.*, 2015	0	0	0	0	0	1
Singh S. K. and El-Kassar, A. N., 2019	1	0	0	0	0	1
Smyth, K. B. *et al.*, 2018	1	0	0	0	0	0
Zhong, R. Y. *et al.*, 2016	1	0	0	0	0	0
Total	**55**	**14**	**9**	**22**	**41**	**15**

Source: Own Elaboration

The Plan category consists of papers essentially focused on further understanding and applying Artificial Intelligence and Big Data Analytics on demand management. The main Plan papers in Table 1 highlight: predictive approaches (PEREIRA *et al.*, 2018), Bullwhip Effect mitigation (MOJAVERI *et al.*, 2009; SINGH; CHALLA, 2016), demand forecasting improvement (AMIRKOLAII *et al.*, 2017; CARBONNEAU *et al.*, 2008; EFENDIGIL *et al.* 2009), development of Vendor Management Inventory technique (CHI *et al.*, 2007) and inventory management optimization (GIANNOCCARO; PONTRANDOLFO, 2002). Furthermore, the majority of Big Data Analytics papers in Table 2 are related to demand forecasting (HOFMANN, RUTSCHMANN, 2018; LAU *et al.*, 2018; LEE, 2017; NITA, 2015; YU *et al.*, 2019). The papers revised on this SCOR plan section have demonstrated that demand forecasting is a fast growing area of machine learning application within the supply chain management field of study. Classic statistical techniques such as exponential smoothing and time series analysis are increasingly being replaced by data mining methods which assess the variance of variables which influence customer behavior or demand patterns. However, further studies should be conducted as the prediction performance of machine learning models require additional computational effort and do not always guarantee better prediction results in comparison with statistical methods. Comparative studies identified that gap and suggested ways to increase the accuracy levels of machine learning models (MAKRIDAKIS; SPILIOTIS; ASSIMAKOPOULOS, 2018).

In contrast to the previous SCOR category, Big Data Analytics is more prominent in the Enable research field. Enable papers in Table 2 are mostly surveys on service supply chains and on the development of capabilities such as agility and preparedness (Fernando *et al.*, 2018;

MANDAL, 2018; ROßMANN *et al.*, 2018), risk management (ENGELSETH *et al.* 2018; IVANOV *et al.*, 2019; MANI *et al.* 2017; ZHAO *et al.*, 2017; WU *et al.*, 2017) and theoretical construction aimed at future applications (HAZEN *et al.*, 2014; 2016; SINGH; EL-KASSAR, 2019; SMYTH, *et al.*, 2018; ZHONG *et al.*, 2016). In specific, the focus of this master research is on supply chain risk management associated with data analytics to improve performance. The studies abovementioned of Mani *et al.* (2017) on the sustainability sphere and social impact assessment, Engelseth *et al.* (2018) on the international supply chain import process optimization and Ivanov et. al. (2019) on the ripple effect control, demonstrate how effective data analytics is becoming on a wider framework of risk control. The access to data as a source of predictive actions to mitigate undesirable effects has proven to be a consistent path to improve performance. Last but not least, Artificial Intelligence Enable papers in Table 1 have a distinct framework focused on development of business process design (PIRAMUTHU, 2005c; PIRAMUTHU, 2005b; SIURDYBAN; MØLLER, 2012; XU *et al.*, 2006). As main technological enabler, IoT (Internet of Things) applications in supply chain management are also highlighted in Table 2 (EHRET; WIRTZ, 2017; HIROMOTO *et al.*, 2017; MA *et al.*, 2018; TSANG, 2018; YUEN *et al.*, 2018).

In third place, Deliver SCOR category in Table 2 consist of papers that cover topics such as analytics to improve distribution practices (SINGH *et al.*, 2018; SIMCHI-LEVI; WU, 2018) and operations management (GIANNAKIS; LOUIS, 2018) using BDA methodologies. Conversely, Deliver papers in Table 1 are more focused on route and resource optimization (CURCIO *et al.*, 2007; MOKHTARINEJAD *et al.*, 2015; WIECZOREK; IGNACIUK, 2018), finished goods warehouse management (MAHROOF, 2019), inventory allocation (WANKE *et al.*, 2017) and network design (ILIE-ZUDOR, *et al.*, 2015).

Finally, Source and Return SCOR areas were the least explored by the papers selected in this section. Source papers concentrated mainly on optimizing supplier selection (CAVALCANTE *et al.*, 2019; CHEN XU, 2018; GUOSHENG; GUOHONG, 2008; KAR, 2015; MORETTO, 2017; ORJI ; WEI, 2015; RAUT, *et al.*, 2017; ZHANG, *et al.*, 2016; ZHANG, *et al.*, 2004). As demonstrated by Engelseth *et al.* (2018), the international procurement process may benefit not only from better supplier selection prediction tools but by improving overall process control. As intended to be demonstrated by this research, the sole focus on supplier selection indicates that the global optimization effort may not be fully achieved if the subsequent logistics processes are not comprised within the wider framework of data analysis. To conclude, selected Return papers consisted of works focused on supply chain sustainability (GARG and VISWANADHAM, 2010; HOPKINS; HAWKING, 2018; KAUR; SINGH, 2018; RAUT *et al.*, 2019; REN *et al.*, 2019; RODRIGUEZ, 2018; SHUKLA; TIWARI, 2017; SINGH; EL-KASSAR, 2019)

Taking everything into consideration, AI and BDA application in sourcing and procurement is mostly related to improving supplier selection processes and can be clearly expanded to other areas of inbound operations (e.g. transportation). In addition, there is a concentration of works related to demand forecasting improvement in the Plan SCOR area which are still to be confirmed by future studies. Last but not least, the Enable dimension analyses showed how critical is to consider this dimension when assessing the application of AI in SCM, especially for risk management capabilities development. Topics such as resources (including Human Resources), contracts, managing regulatory compliance and risk management have played a pivotal role in facilitating the implementation of models in real business situation. As discussed before, data driven supply chain and predictive logistics are key to achieve higher standards of results and costs in the future. Even though current literature review has shown that some progress has been made, there are areas that can be further explored such as delay prediction and mitigation.

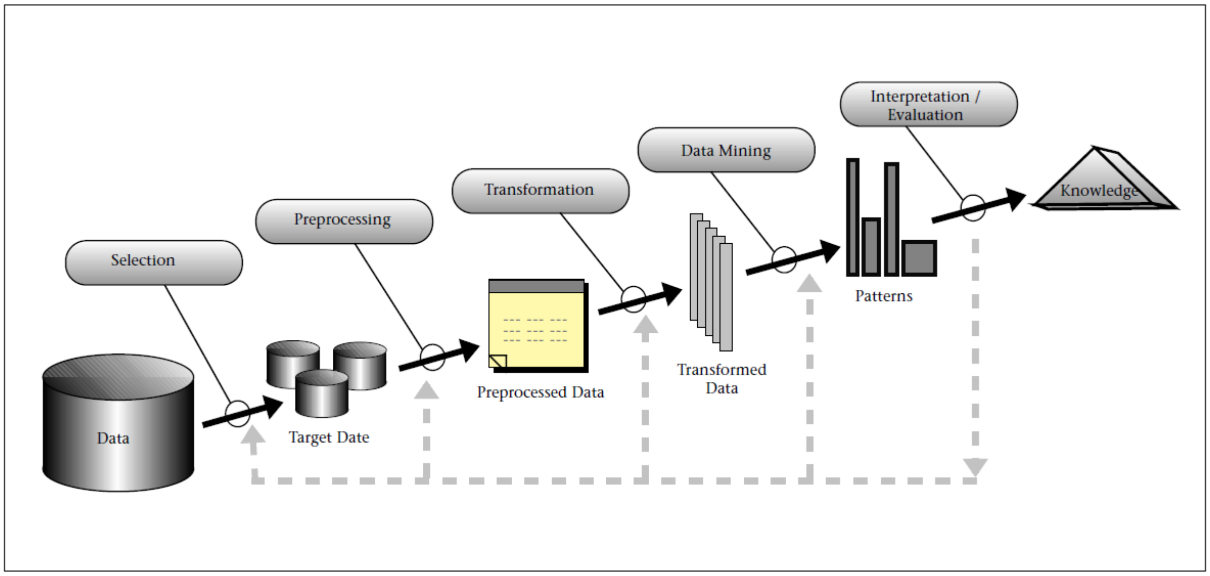
In order to further explore that investigation path, the concepts of data mining and intelligent algorithms are described in the next section. This theoretical reference is key as it will be the basis of this research discussion of air freight delay prediction as a risk management tool within the SCOR Enable dimension of operations.

3.2. CONCEPTS OF DATA MINING AND INTELLIGENT ALGORITHMS

According to Han, Kamber and Pei (2011, p. 13), “powerful and versatile tools are badly needed to automatically uncover valuable information from the tremendous amounts of data and to transform such data into organized knowledge. This necessity has led to the birth of data mining”. Amo (2010) considers data mining as a step within the KDD methodology

(Knowledge Discovery from Data). The KDD can be described by the following sequence: the processes of (i) data cleaning, (ii) data integration, (iii) selection, (iv) data transformation, (v) data mining, (vi) pattern evaluation or post- processing and (vii) visualization of results or knowledge presentation (SHAFIQUE; QAISER, 2014). The process model of the KDD is shown in Figure 2 below:

Figure 2: KDD Model Steps



Source: Fayyad, Piatetsky-Shapiro e Smyth (1996)

The data cleaning step targets the removal of noise and inconsistent data. Data integration objective is to consolidate data sources into a single data warehouse. Data Selection process retrieves relevant data for the analysis from the data warehouse. Data transformation consists of transformation and consolidation of data to allow appropriate mining operations. This step may include data reduction in order to obtain a smaller representation of the original dataset without losing integrity. Data mining is where intelligent methods are applied to extract data patterns. Pattern evaluation comprises of the identification of truly interesting patterns that represents knowledge. Finally, the knowledge presentation step encompasses techniques that provide visualization of present mined knowledge to users (HAN; KAMBER; PEI, 2011).

The data mining models can be divided into two main categories, namely predictive and descriptive:

> The purpose of a data mining effort is normally either to create a descriptive model or a predictive model. A descriptive model presents, in concise form, the main characteristics of the data set. It is essentially a summary of the data points, making it possible to study important aspects of the data set. The purpose of a predictive model is to allow the data miner to predict an unknown (often future) value of a specific variable; the target variable. If the target value is one of a predefined number of discrete (class) labels, the data mining task is called classification. If the target variable is a real number, the task is regression (JAIN; SRIVASTAVA, 2013, p. 116).

The data mining classification task can be supported by different data mining techniques. A data mining technique consists of the definition of methods that aim at achieving the pattern and knowledge discovery goals (JAIN; SRIVASTAVA, 2013). Machine learning algorithms are one of the main current data mining techniques that can be used, for example, as classifiers for discrete target attributes based on an existing labeled training dataset (supervised learning). In specific, the binary classification task is one of the most common types of predictive problems in which the target attribute has only two possible classes as possible outputs (CANBEK *et al.*, 2017).

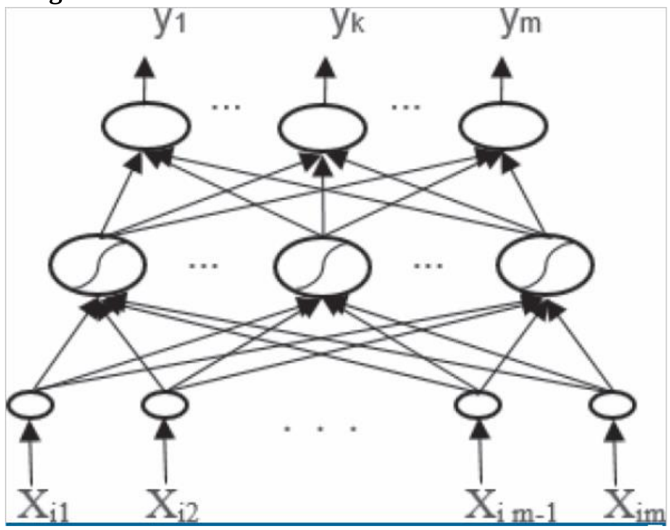
Different training and test options may be applied associated with those algorithms. One may mention the Cross-Validation test option that divides the database in different number of folds and alternates records between the test and training sets on each iteration to check the consistency of the prediction force. Conversely, in the Hold Out test strategy a fixed test and training group is defined without alternating data subsets between them (SCHAFER, 1993). Cross-validation is usually the preferred method for smaller datasets because it gives your model the opportunity to train on multiple train-test defined by the number of K-fold partitions. In spite of the bigger computational effort, this yields better indication of how well your model will perform on unseen data. Hold-out, on the other hand, is dependent on just one train-test split. That makes the hold-out method score dependent on how the data is split into train and test sets and is generally useful to segregate raw data for future validations of the training model (YADAV; SHUKLA, 2016).

The Random Forest (RF), Support Vector Machines (SVM), Artificial Neural Networks (ANN) and K-Nearest Neighbors (KNN) algorithms are examples of machine learning techniques that can be applied in data mining prediction problems of this kind. The Random Forest algorithm consists of a set of decision trees combined to solve a problem. According to Quinlan (1986), a decision tree is a hierarchical classification system based on the partition of a universe of objects into classes. Each decision tree can be built based on a random sample of the data and in each node the best split attribute (biggest gain of information) is chosen (DIETTERICH, 1998). According to Breiman (2001, p.6), random forest can be defined as "a classifier consisting of a collection of tree-structured classifiers {h(x,k), k = 1, . . .} where the {k} are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x." After a large number of trees are generated, the most popular class is selected.

In turn, Support Vector Machine uses a non-linear mapping to transform the original data (training) into a larger dimension. In the new dimension, SVM seeks a hyperplane that generates an optimal linear separation or with maximum margin (VAPNIK; CHERVONENKIS, 1964; VAPNIK, 1989). Platt (1998) proposed an algorithm called SMO (Sequential Minimal Optimization) for training support vector machines in order to avoid the very large quadratic programming (QP) optimization problem. Basically, the SMO breaks the QP problem into a series of smallest problems which are solved analytically avoiding the matrix computation and decreasing substantially computation time.

Artificial Neural Networks emulates the neural interconnections in the brain which are abstracted and implemented on digital computers. The multilayer perceptron (MLP) algorithm is an artificial neural network structure and is a nonparametric estimator that can be used for classification tasks (MUBAREK; ADALI, 2017). A typical multilayer perceptron (MLP) neural network and a hidden neuron in the hidden layer are depicted in Figure 3. A hidden layer is required for MLPs to classify linearly inseparable data sets.

Figure 3: Artificial Neural Networks structure



Source: (MUBAREK; ADALI, 2017)

Gardner and Dorling (1998) further describe the MLP functioning:

> The nodes are connected by weights and output signals which are a function of the sum of the inputs to the node modified by a simple nonlinear transfer, or activation, function. It is the superposition of many simple nonlinear transfer functions that enables the multilayer perceptron to approximate extremely non-linear functions. The output of a node is scaled by the connecting weight and fed forward to be an input to the nodes in the next layer of the network. This implies a direction of information processing; hence the multilayer perceptron is known as a feed-forward neural network. The architecture of a multilayer perceptron is variable but in general will consist of several layers of neurons. The input layer plays no computational role but merely serves to pass the input vector to the network. The terms input and output vectors refer to the inputs and outputs of the multilayer perceptron and can be represented as single vectors (GARDNER; DORLING, 1998, p. 2).

The KNN (K-Nearest Neighbors) utilizes the Euclidean distance between instances to classify attributes as demonstrated in the formula below (FIX; HODGES, 1951).

$$d(E_i, E_j) = \sqrt{\sum_{r=1}^{M} (x_{ir} - x_{jr})^2}$$

The K parameter of the K-Neighbor algorithm is the number of neighbors to be considered in the classification iteration. Due to the fact that processing is delayed until a new element is classified, this method can also be characterized as "lazy" learning and requires additional storage and computational processing capacity. The number of K neighbors can be defined by hyperparameter adjustment techniques in which the best accuracy levels are identified in successive simulations using cross validation test option. (AHA; KIBLER; ALBERT, 1991).

In addition to the abovementioned algorithms, other data mining techniques such as attribute selection can be used to gain processing performance, simplify prediction models and provide a better study on the relationship between attributes. The statistical Chi-Square test evaluates the worth of an attribute by computing the value of the chi-squared statistic with respect to the class. Basically, it performs a hypothesis test to validate whether there is correlation between the attributes and target attribute. Variables with zero significance can be eliminated (LIU; SETIONO, 1995). The Wrapper method searches for an optimal feature subset tailored to a particular algorithm and a domain (KOHAVI; SOMMERFIELD, 1995). The best first search and hill climbing algorithms can be used within the Wrapper method to obtain optimal results. According to Kohavi (1994), best first search leads to better accuracy as it expands the search for better subsets. According to Kohavi and John (1997), "The idea behind the wrapper approach, is simple: the induction algorithm is considered as a black box. The feature subset with the highest evaluation is chosen as the final set on which to run the induction algorithm". In short, the best attributes to be chosen may vary according to the machine learning technique to be employed on the dataset.

Finally, CFS (Correlation-based Feature Subset Selection) calculates a correlation matrix of attribute-class and attribute-attribute. Instead of focusing on individual variables, it seeks to find a subset of attributes that are highly correlated with the target attribute and which do not have strong correlation in between them. Thus, subsets of attributes are formed and through merit (S) the attributes with the greatest contribution are defined to describe the target attribute. It initializes with an empty subset and utilizes the best-first-search heuristics up to the halt criteria of 5 consecutive subsets which do not improve the merit level (HALL, 1998).

In spite of the application of predictive and attribution selection techniques, results may not reach satisfactory levels. One of the reasons behind could be that the target attribute has unbalanced representation between the positive and negative classes. The positive class is the variable that has less representativeness in the dataset and is generally the goal of prediction analysis. Methods of undersampling and oversampling can be applied. The NCL (Neighborhood

Cleaning Rule) is based on the KNN and is an undersampling technique (BATISTA; PRATI; MONARD, 2004). The K parameter of the K-Neighbor algorithm is the number of neighbors to be considered in the classification iteration. Based on the closest neighbors of the positive class, the NCL algorithm removes the records of the majority class as a way to balance the total sampling. On the other hand, Smote algorithm performs an oversampling by interpolating new records with respect to the positive or minority class (CHAWLA *et al.*, 2002). Random undersampling may eliminate important instances causing loss of information. On the other hand, Random oversampling can lead to overfitting. Overfitting happens when

> …a learning algorithm fits the training data set so well that noise and the peculiarities of the training data are memorized. According to the result of learning algorithms performance drops when it is tested in an unknown data set…On the other hand, underfitting occurs when the model is incapable of capturing the variability of the data (ALLAMY, 2015).

Finally, the use of intelligent algorithms can be evaluated by a wide range of performance indicators:

> In predictive analytics, a table of confusion (sometimes also called a confusion matrix) is a table with two rows and two columns that reports the number of false positives, false negatives, true positives, and true negatives. This allows more detailed analysis than mere proportion of correct classifications (accuracy). Accuracy will yield misleading results if the data set is unbalanced; that is, when the numbers of observations in different classes vary greatly. The precision is one important metric to be considered. In pattern recognition, information retrieval and classification (machine learning), precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances, while recall (also known as sensitivity) is the fraction of relevant instances that were retrieved. Both precision and recall are therefore based on relevance. (GERON, 2019, p. 90).

Finally, the ROC Curve is another common tool used with binary classifiers (PRATI *et al.*, 2008). The ROC curve plots the True Positive Rate (TPR or recall) against the False Positive Rate (FPR). The FPR is the ratio of negative instances that are incorrectly classified as positive while the True Positive Rate is the correctly classified instances as shown in the formulas below:

$$TPR=TP/(TP+FN)$$

$$FPR=FP/(FP+TN)$$

One additional way to compare classifiers is to measure the area under curve (AUC). A perfect classifier will have a ROC AUC equal to 1, whereas a purely random classifier will have a ROC AUC equal to 0.5 (GERON, 2019). Finally, the Kappa concordance test was proposed by Cohen (1960) in order to measure the degree of agreement between variables. Values above 0.8 are considered optimal, between 0,6 and 0,8 as good, between 0,4 and 0,6 as regular and below 0,4 as poor result. The next section presents how those intelligent algorithms are currently being applied to improve risk management capabilities mainly on supply chain operations.

3.3. INTELLIGENT ALGORITHMS APPLIED TO SUPPLY CHAIN RISK MANAGEMENT

Overall, Supply Chain Risk Management approaches can be classified by four main categories: disruption risk management (DRM), operational risk control (ORC), disaster and emergency management (DEM), and logistics service risk analysis (LSRA) (CHOI; CHIU; CHAN, 2016). In the LSRA field of research, “the arm length relationship between customers and Third-Party Logistics providers have positive influence on the supply chain performance” (GOVINDAN; CHAUDHURI, 2016). Freight forwarders are implementing Artificial Intelligence and Big Data Analytics solutions in order to mitigate risks for their customers (DHL, 2021). The SCRM strategies also benefits directly from this collaborative approach as considerable amounts of operational data can be shared in order to predict possible supply chain glitches. Data mining techniques such as Machine Learning can be thus used to identify those potential threats by analyzing real-time information and proposing mitigation actions (HASSAN, 2019).

The benefits of adopting data analysis include the reduction of the bullwhip effect (negative effect of increasing the levels of average upstream stock), less frequent stock outs and a higher level of service offered to the final customer (CHOPRA; MENDL, 2016). The supply chain becomes more integrated, reducing uncertainties related to the transportation, storage or distribution of products, contributing to the reduction of the total logistical cost. (TIWARI; WEE; DARYANTO, 2018; WANG *et al.*, 2016; WALLER; FAWCETT, 2013b). As an example, Ben-Daya, Hassini and Bahroun (2017) showed that the use of information sent by products throughout the stages of the production process using RFID positively impact the accuracy of the inventory and performance of the production plan execution.

Machine Learning (ML) research has already a wide field of application in Supply Chain Risk Management (SCRM). The overall objective is to boost the supply chain capacity to handle and control risks before they impact daily business (BARYANNIS et. al., 2019b). Shahrabi, Mousavia and Heydar (2009) argues that Artificial Neural Networks and Support Vector Machines presented better results in terms of demand planning accuracy than moving average and exponential smoothing (with and without trend) laying the ground for better production performance and less deviations. Likewise, Mojaveri *et al.* (2009) applied the SVM and ANN algorithms to predict demand levels. ANN outperformed all the classical statistical methods and also the SVM methodology as more accurate forecasting tool.

Cavalcante *et al.* (2019) used KNN and Logistic Regression algorithms to predict better supplier delivery performance based on a binary classification problem of two main dependent variable classes: deliveries on time and late deliveries. The main output of this study was that it was possible to create a risk profile that represents the probability of success in predicting the supplier behavior in the system regarding the target feature, which is the OTD (on time delivery) in this model. Based on the risk profile, it would be possible to structure continuous improvement strategies for supplier development (Cavalcante *et al.*, 2019). Similarly, Guosheng and Guohong (2008) proposed a methodology to predict supplier performance and enhance the selection process using Artificial Neural Networks and Support Vector Machines. Based on expert ranking data on 22 variables ranging from production capacity to level of service, it was possible to predict best vendors to be contracted. The SVM algorithm has outperformed the ANN both in training and test scenarios.

De Santis, De Aguiar and Goliatt (2017) proposed a different approach to deal with inventory stock out by applying Logistic Regression and Classification Tree (Cart) machine learning algorithms to identify materials at risk of backorder before the event occurs. One of the main conclusions of the paper was that the proposed predictive methodology exhibited a real potential of increasing service level in real inventory management systems. Artificial Neural Networks are also one of the major techniques currently used to increase demand forecasting accuracy thus decreasing disruption risks (PEREIRA *et al.*, 2018; SLIMANI; EL FARISSI; ACHCHAB, 2017). Lee (2017) proposed a model which combined clustering analysis from consumer behavior with GA (Genetic Algorithms) leading to potential transportation cost and risk reduction.

Park, Yoon and Yoo (2018) developed a framework to assess supply chain risk and define main processes which could lead to disruption using PCA (Principal Component Analysis) based on simulated and real data. Giannakis and Louis (2011) proposed a multi-agent framework to statistically evaluate risks by assessing key performance indicators of the supply chain process such as On Time Delivery. Associates expertise over problematic suppliers can also be translated into Clustering algorithms in order to take appropriate mitigation decisions in the supply chain (ER KARA; OKTAY FIRAT; GHADGE, 2020).

Further applications of ML using transportation data to improve Supply Chain Risk Management (SCRM) and avoid failures are also becoming increasingly valid (HO *et al.*, 2015). Baryannis and Dani (2019) tested a combined framework of AI (Support Vector Machine and Decision Tree) with supply chain experts information to predict delays (binary classification problem) within a real world multi-tier aerospace manufacturing supply chain with focus on supplier attributes. Viellechner and Spinler (2020) used origin port, destination port and vessel data to test and validate whether there would happen delay in intercontinental container shipments using mainly Neural Networks. Wu *et al.* (2017) structured a decision tree analysis model to predict global supply chain cargo loss severity in terms of financial impact based on input attributes related to product, geographical information and transport data. The abovementioned studies have proved that data mining techniques are a key enabler for improving supply chain performance. Specifically in the automotive industry, there has been also research using machine learning techniques such as Support Vector Machine to identify supply chain risk using textual information from the internet (HASSAN, 2019).

The application of machine learning techniques in the air freight cargo industry is growing as a whole (Chung et. al., 2020). One may mention Liu et. al. (2019) paper that proposed a model to predict how “regional convective weather” affects ground delay program (GDP). In order to build this useful technique for flight operators and other stakeholders, they applied SVM

(Support Vector Machine), logistics regression and RF (Random Forest) models. The Random Forest model outperformed the other algorithms and confirmed its high suitability for nominal variable as data inputs. Etani (2019) studied the application of Random Forest to predict delays based on Weather conditions in Japanese airports. As a result, on-time arrival fight is predicted at 77% of the accuracy with using Random Forest Classifier of machine learning. Herrema *et al.* (2019) also used the Random Forest prediction algorithm for assessing runway capacity and utilization to avoid delays. The machine learning method obtained an accuracy of 79% and was used to observe key related precursors of unique data patterns.

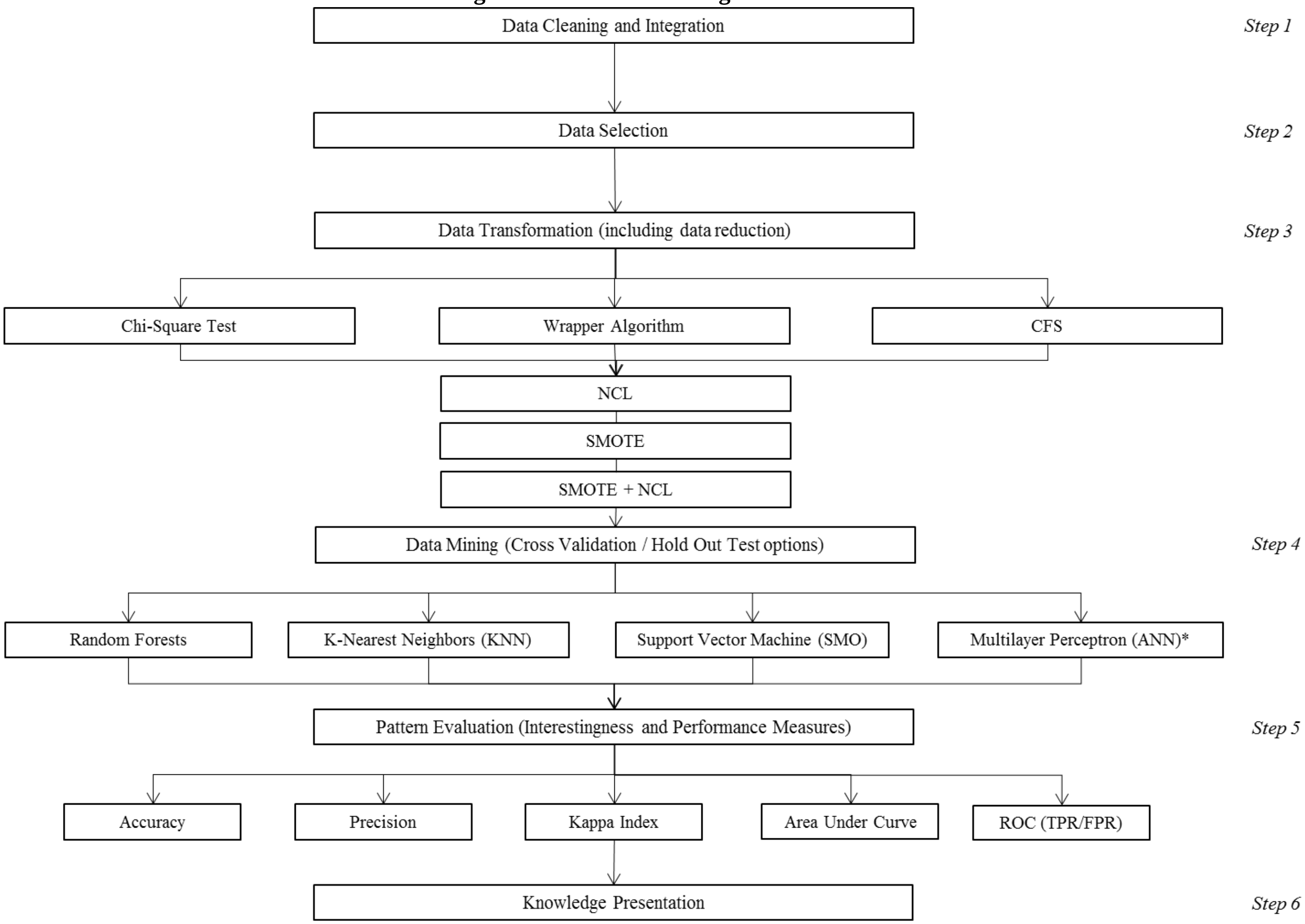
Yu B. et. al. (2019) researched commercial air transport micro influential factors (e.g. air route situation and crowdedness degree of airport) that influence flight delays using Neural Networks. The proposed method has proven to be highly capable of handling the challenges of large datasets and capture the key factors influencing delays. Congestion analysis is also a major field of study regarding delay causality in transport systems. Diana (2018), used different model approaches such as ensemble learning models (Random Forest) to confirm that machine learning can support on predictive operations control. Gui et. al. (2020) proposed a combined Random Forest model application based on Big Data related to flight delay factors such as Airport, Flight, Air Route and other operational information. Compared with the previous schemes, the proposed random forest-based model obtained higher prediction accuracy (90.2% for the binary classification) and could overcome the over-fitting problem.

Taking everything into consideration, the application of intelligent algorithms has proven to be a powerful tool in predictive logistics. In specific, the Random Forest algorithm in data driven air freight transportation management and analyses, has generally outperformed the other techniques as the most efficient classifier. The algorithms from the literature review were considered as methodological reference for building and testing a new approach for predicting delays in international air freight supply chains in this research. The objective is to expand the current predictive logistics literature by applying a thorough data mining methodology that combines different sampling (Cross-Validation and Hold Out), attribute selection (Chi-Square, CFS, Wrapper), dimensionality reduction (NCL and SMOTE) and algorithm (SVM, KNN, MLP and RF) techniques. In short, this framework could be used as reference in future research or practical work not only in transportation but also on other supply chain risk management fields. Research methodology is described in the next section.

4 MATERIALS AND METHODS

The methodology applied on this research was based on the KDD process described in Section 3.2. Figure 4 illustrates the main methodological workflow adopted. Main research steps are described in the following subsections.

Figure 4: KDD Methodological Workflow



* ANN Algorithm used additional criteria to select variables due to computational and results improvement assessment (Further details in 4.1.2. Section)

Source: Own Elaboration

4.1. MATERIALS

4.1.1. Data Preparation (Cleaning / Integration / Selection)

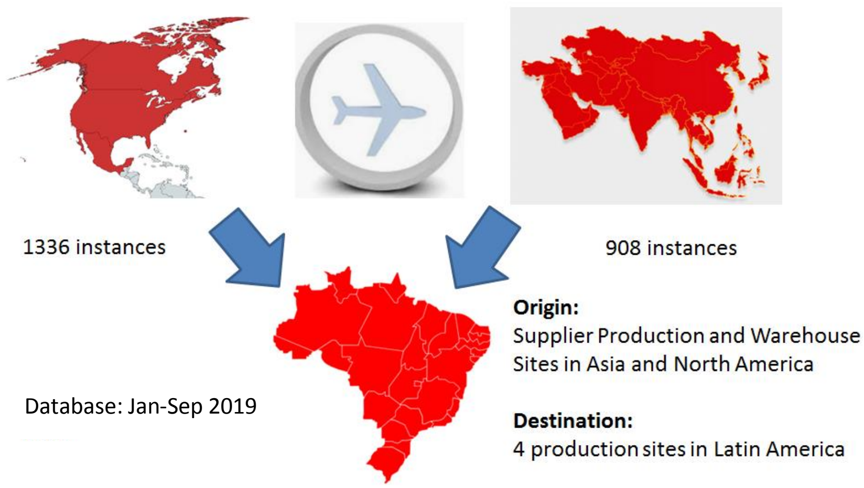
The research database was comprised of air freight intercontinental shipments to supply four automotive supply parts production plants in Latin America. The multinational company of the case study operates in Latin America and has approximately 80% of its revenue generated in Brazil. Main customers range from automakers to auto parts dealers from the aftermarket segment spread mostly in the Southern and Southeastern regions of the country. The major part of production components is imported leading to high operational complexity to meet revenue targets and market demand. The air freight modal is usually utilized to compensate delayed sea freight shipments and avoid production line stoppage and stock outs.

In specific, the international air logistics pre-carriage process starts when the cargo is picked up at the supplier's factory or warehouse in the country of origin. The consolidated cargo is sent to the airport of origin and the main route between continents is carried out by airlines subcontracted by the air freight forwarder. Finally, in the destination country, the cargo is customs cleared by the broker and transported by road to the destination industrial plant.

The initial air shipment historical database had 153 variables and one dependent variable. The target attribute contained the information of whether there was a delay for each shipment (binary classification problem). Amongst the main operational explanatory attributes one can mention: Country of Origin, Route, City of Origin, Region of Origin, Airport of Destination, Airport of Origin, Date of Shipment, Type of Service (Standard, Emergency), Supplier Name, Pick-up Address, Customer Name, Customer Address, Airline, Operational Dates (Authorization, Pick-up, Airport Departure, Airport Arrival, Plant Arrival) and Cargo Weight.

All variables could be categorized as nominal with the exception of Weight, Customer Number, Supplier Number and Operational Dates. The target attribute was also nominal (Delayed / Not Delayed). Finally, the air shipment file had 2244 instances (flight data) related to the period from January to September 2019 considering flights that arrived from North America and Asia to Latin America in that period. Data was gathered from the Logistics Service Provider database which controls and monitors import shipments. Figure 5 below represents overview information of the original dataset:

Figure 5: Overview Dataset Information



Source: Own Elaboration

An initial analysis of the database was performed in the cleaning and integration phase of the KDD (Step 1) according to Figure 4. A few preprocessing tasks were necessary. A total of 108 redundant and not relevant attributes were removed. The final number of 45 attributes was obtained plus one target attribute with the information of occurrence or not of delay. Appendix 1 summarizes the attributes chosen for the future analyses.

The data was extracted from the freight forwarder database and converted to the format ARFF datafile of the Weka® software in the selection phase of the KDD (Step 2). According to Frank, Hall and Witten (2016), the main functions and applications of the Weka Software can be described as follows:

> The WEKA workbench is a collection of machine learning algorithms and data preprocessing tools that includes virtually all the algorithms described in our book. It is designed so that you can quickly try out existing methods on new datasets in flexible ways. It provides extensive support for the whole process of experimental data mining, including preparing the input data, evaluating learning schemes statistically, and visualizing the input data and the result of learning. As well as a wide variety of learning algorithms, it includes a wide range of preprocessing tools. This diverse and comprehensive toolkit is accessed through a common interface so that its users can compare different methods and identify those that are most appropriate for the problem

> at hand. WEKA was developed at the University of Waikato in New Zealand; the name stands for Waikato Environment for Knowledge Analysis… The system is written in Java and distributed under the terms of the GNU General Public License. It runs on almost any platform and has been tested under Linux, Windows, and Macintosh operating systems (FRANK; HALL; WITTEN, 2016).

4.1.2. Data Transformation (w/ Feature Selection)

In the data transformation phase (Step 3), it was decided not to replace the missing values using the Mode technique or KNN algorithm since they were mostly related to shipment dates. This change could significantly affect the accuracy of the model. In addition, the Random Forest algorithm used is not impacted by attributes with missing values.

As a way to increase the performance of the MLP and SMO algorithms, the attribute selection approach was used (KUMAR; MINZ, 2014). It is worth mentioning that the Random Forest algorithm has this solution included in its logic and it is not necessary to perform the dimensionality reduction procedure. Specifically, Chi-Square, Wrapper and CFS (Correlation Feature Selection) methods were used to select attributes. The attributes chosen in each variable reduction method are described in the Appendix 2 for the SMO algorithm, Appendix 3 for the MLP and Appendix 4 for the KNN. In addition to the attribute selection techniques abovementioned, it is noteworthy to highlight that an additional reduction criteria was performed within each Chi-Square, CFS and Wrapper techniques for the MLP algorithm. The nominal attributes with a high number of categories were eliminated in order to improve computational performance. This approach has proven to improve the results in spite of the loss of information. Appendix 3 depicts the final list of attributes used for the MLP algorithm training and testing.

In addition to the attribute selection procedure, The NCL (Neighbor Cleaning Rule) and Smote (Synthetic Minority Oversampling Rule) algorithms were also applied to improve class balance for the target attribute and thereby overall results (TORGO *et al.*, 2013). The positive or minority class represented 25% of the total number of instances (delayed shipments).

4.2. DATA MINING AND KNOWLEDGE EVALUATION METHODS

4.2.1. Data Mining

The algorithms of Random Forest, K-Nearest Neighbors, Support Vector Machine (SVM) and Multilayer Perceptron (MLP) were applied to the database (Step 4) in a binary classification problem with two classes: delayed or not delayed shipments. The main objective is to predict the positive class which corresponds to the minority class of delayed shipments. The KNN number K of neighbors was defined as 9 neighbors after Hyperparameter adjustment using cross validation technique. In other words, the higher accuracy level for this dataset was achieved using 9 neighbors as reference. In addition to the 10-Fold Cross-Validation test option, the Hold Out methodology was used in a complementary manner. The main objective was to ensure the application of the classifiers in a more real training and testing situation as well. Class balancing (SMOTE and NCL) was performed only for the training sample while the test subset remained unbalanced. Both methods were kept as reference for final analysis and model selection definition.

4.2.2. Results Evaluation

Based on the observation of the classifier global efficiency indicators (Kappa and Accuracy), it was initially defined if the classifier was sufficiently trained to be applied with new real data in order to support the improvement of future logistic performance (Step 5). The precision of the positive class was also used as a specific indicator of the performance of the classifier considering the delay occurrence prediction capability. In addition to those metrics, TPR, FRP and AUC were applied to the dataset. The higher the TPR and AUC and the lower the FPR the better the performance of the classifier. The ROC (Receiver Operator Characteristic Curve) was used as final analysis criteria to define the best classifier. The Rocon® software was the tool applied for calculating the ROC curve. Basically, based on the inputs of TPR and FPR of the positive class, it plotted the results to identify the best classifier.

4.2.3. Knowledge Validation

The final results were presented in order to answer the research hypothesis of whether the use of supervised learning algorithms can contribute to improve predictability of international shipment delays and improve the supply chain performance. The main variables affecting air freight performance and how managers can act to mitigate supply chain risks were discussed on the Conclusion section.

Next section describes the main results achieved through the KDD methodological application of Data Mining (Step 4 and 5).

5 RESULTS AND DISCUSSION

The results shown in Table 3 below were obtained using the training option of 10-Fold Cross-Validation for the MLP, KNN, Random Forest and SVM algorithms. The objective of this initial testing was to provide reference results to be compared with the output of the proposed KDD methodology steps with attribute selection and class balancing.

Table 3: Application of Machine Learning Algorithms (without attribute selection and class balancing)

Classifier	Results		
	Accuracy	Precision	Kappa
Random Forest	81.149	0,79	0,37
SVM	76.871	0,53	0,36
KNN	80.169	0,665	0,39
MLP	77.133	0	0

The best result was achieved by Random Forest with approximately 81% accuracy and 0.37 Kappa. Although the accuracy was greater for the Random Forest, all models presented a low result for the Kappa index. In specific, the MLP did not perform well mainly because of the complexity of the attributes which were mostly nominal variables (29 out of 45 attributes). The computational effort to construct the binary input layer for the neural network has proven to be an obstacle to the overall convergence of the algorithm. The 45 attributes summed up to 5255 categories of values which demanded a high number of neurons in the hidden layer leading up to poor results on Precision metric and the Kappa index.

Thus, it was decided to carry out the process of variable selection (Chi-Square, CFS and Wrapper) for SVM and MLP and also the class balancing procedure

(SMOTE, NCL and NCL + SMOTE) for all algorithms. For the SMO, the following attribute selection were performed: (i) chi-Square method, three attributes with zero statistical significance were removed, (ii) Wrapper the attributes with no representativeness were deleted and finally (iii) the CFS algorithm selected only one subset with three attributes with the highest correlation with the target attribute. The detailed attribute selection approach results are available in the Appendix 2. As mentioned before, the Appendix 3 shows the specific adapted attribute selection criteria within the Chi-Square, CFS and Wrapper for the MLP. That process was performed to increase algorithm computational performance and results. Finally, the Random Forest algorithm did not need the attribute selection approach as it has this built-in functionality.

Table 4 below depicts the results achieved with 10-Fold Cross Validation for each attribute selection combined with SMOTE class balancing. It was kept the training option by cross-validation. The same exercise displayed in Table 5 was done but considering the Hold Out test option with 90% of the original dataset for training and 10% for test validation. Initially, the ratio of 75% (training) and 25% (test) was performed but has shown lower performance leading to the 90/10 ratio final decision. The Random Forest algorithm has a built-in feature selection solution in its algorithm and therefore was not subjected to Chi-Square, Wrapper and CFS methodologies as shown in Table 4. Conversely, the SVM, MLP and KNN utilized the former feature selection algorithms to improve performance as they do not possess attribute selection in their iterative logic.

Table 4: Results with Attribute Selection and Class Balancing SMOTE (10 fold Cross-Validation Test Option)

Method	SVM			MLP			KNN			RF		
	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa
χ^2	80,799	0,757	0,59	75,945	0,698	0,49	84,261	0,781	0,67			
Wrapper	78,966	0,612	0,37	75,445	0,507	0,30	80,299	0,733	0,59			
CFS	74,768	0,691	0,47	70,449	0,646	0,37	77,551	0,700	0,53			
Built-in										86,759	0,893	0,71

Table 5: Results with Attribute Selection and Class Balancing SMOTE (Hold Out Test Option)

Method	SVM			MLP			KNN			RF		
	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa
χ^2	78,222	0,564	0,41	73.333	0,474	0,36	77,333	0,556	0,35			
Wrapper	78,222	0,578	0,37	75,555	0,507	0,38	77,333	0,581	0,28			
CFS	73,333	0,468	0,31	74,222	0,482	0,31	75,555	0,509	0,33			
Built-in										84,000	0,763	0,52

Following the methodology, additional testing was conducted using the NCL Class Balancing approach combined with the abovementioned attribute selection approaches. Table 6 and 7 depict the results achieved:

Table 6: Results with Attribute Selection and Class Balancing NCL (10 Fold Cross Validation Test Option)

Method	SVM			MLP			KNN			RF		
	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa
χ^2	94,634	0,658	0,59	97,030	0,767	0,69	94,179	0,704	0,45			
Wrapper	94,117	0,681	0,56	95,756	0,675	0,57	92,955	0,571	0,14			
CFS	92,434	0,429	0,07	97,239	0,571	0,44	93,870	0,583	0,43			
Built-in										94,340	0,816	0,41

Table 7: Results with Attribute Selection and Class Balancing NCL (Hold Out Test Option)

Method	SVM			MLP			KNN			RF		
	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa
χ^2	79,555	0,813	0,28	76,444	0,556	0,23	76,888	0,833	0,11			
Wrapper	78,666	0,9	0,21	76,889	0,559	0,28	76,000	1,00	0,05			
CFS	75,555	1,0	0,02	76,889	0,75	0,13	76,444	0,714	0,10			
Built-in										76,888	0,833	0,11

Finally, additional testing was conducted using the NCL+SMOTE Class Balancing approach. Table 8 and 9 depict the results achieved:

Table 8: Results with Attribute Selection and Class Balancing NCL+SMOTE (10 Fold Cross Validation Test Option)

Method	SVM			MLP			KNN			RF		
	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa
χ^2	98,321	0,949	0,72	98,271	0,767	0,64	94,179	0,704	0,45			
Wrapper	96,621	0,707	0,45	98,101	0,842	0,58	96,231	0,455	0,06			
CFS	95,944	0,25	0,03	98,365	0,200	0,03	96,625	0,556	0,35			
Built-in										96,863	0,926	0,32

Table 9: Results with Attribute Selection and Class Balancing NCL+SMOTE (Hold out Test Option)

Method	SVM			MLP			KNN			RF		
	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa	Accuracy	Precision	Kappa
χ^2	79,555	0,813	0,28	74,666	0,476	0,14	76,000	0,75	0,06			
Wrapper	78,666	0,900	0,21	75,111	0,500	0,16	75,555	1,00	0,02			
CFS	75,555	1,00	0,02	76,888	0,75	0,13	75,555	0,667	0,04			
Built-in										76,444	1,00	0,07

Distinct results were achieved on the test validations scenarios. Overall, the best results achieved were those combining SMOTE attribute reduction technique with Cross-Validation and Hold Out test options. It was considered the most suitable methodological alternative to be replicated in real applications as the Hold Out test in comparison with the Cross-Validation technique did not lose as much as the other NCL and NCL + SMOTE techniques in terms of Accuracy, Kappa and Precision metrics. In addition to that, the NCL and NCL + SMOTE may have caused overfitting in the cross validation test scenario as the metrics achieved an excessive high level of accuracy and low performance in the hold out test option. None of these latter scenarios reached good Kappa metrics leading to the conclusion achieved of the SMOTE as the best balancing scenario.

In specific, the Random Forest algorithm performed better than the SMO, MLP and KNN (both in cross validation and hold out test options) within the SMOTE methodological scenario. It achieved Kappa metric results in the cross-validation test option considered as good (0,71) and regular for the Hold Out (0,52). Furthermore, the precision level of the positive class was considered as adequate reaching 89% and 76% in the cross-validation and hold out scenarios respectively. The KNN algorithm ranked in second place achieving better results in comparison to the ANN and SVM algorithms in the cross validation test scenario due to the Hyperparameter Adjustment done to optimize the level of accuracy (k=9 neighbors).

In order to further confirm the Random Forest Algorithm as the best suitable option of the SMOTE scenario, the ROC curve was plotted for both the Hold Out and Cross-Validation considering the best results of each algorithm. In order to build the ROC curve, the FPR, TPR and AUC values from the positive class were chosen (delayed shipments). Results are shown in the Table 10, 11 and Figures 6, 7.

Table 10: ROC Curve Metrics (SMOTE Cross-Validation)

Algorithm/ Attribute Selection Method	Data Matrix		
	FPR	TPR	AUC
MLP / Chi-Square	20	70	0,81
RF / Built-in	6	75	0,91
KNN / Chi-Square	15	84	0,90
SVM/Chi-Square	16	76	0,80

Table 11: ROC Curve Metrics (SMOTE Hold Out)

Algorithm/ Attribute Selection Method	Data Matrix		
	FPR	TPR	AUC
MLP/Wrapper	19	60	0,72
RF / Built-in	5	51	0,78
KNN / Chi-Square	12	44	0,76
SVM / Chi-Square	14	55	0,70

In addition to presenting the best combined result of Accuracy and Kappa, the RF algorithm also had the lowest false positive rate and the highest area value under the curve for both test options scenarios. The larger the area under the curve, the better the average classifier performance is. Having evaluated all the performance indicators, the RF algorithm was considered the best option, as detailed in Figure 6 and 7.

Figure 6: ROC Curve (SMOTE Cross-Validation)

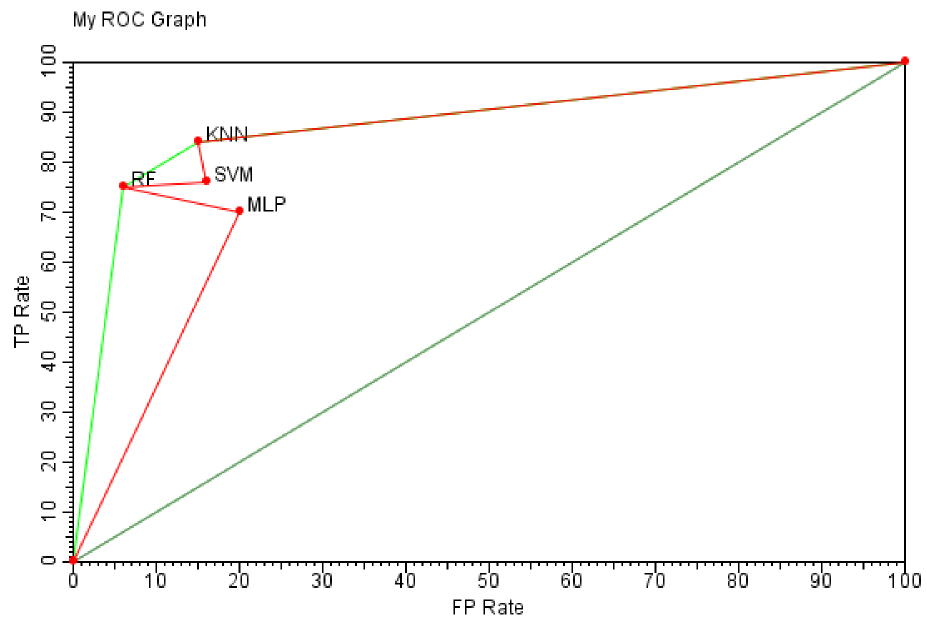
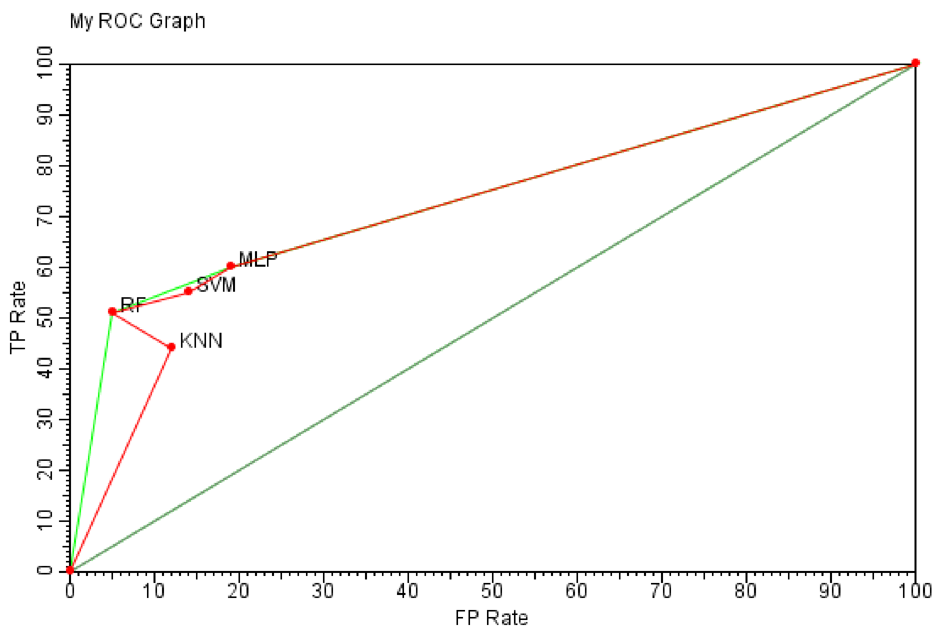


Figure 7: ROC Curve (SMOTE Hold Out)



The achievement of satisfactory results owns to the consistency of model and validation metrics application. Similar to the studies conducted by Liu et al. (2019), Etani (2019), Herrema et al. (2019) and Gui et al. (2020), the Random Forest algorithm outperformed others classifiers in the predictive transport delay binary classification problem. In addition, as shown by Gui et al. (2020), the use of a wider range of type of variables increased the model accuracy level in the binary classification task. The current study used not only data from origin or destination airports but rather expanded attribute selection to shipper, consignee and logistics service provider data increasing thereby the model predictability to 86% of accuracy.

The main variables that influence supply chain delay in this case study were also unveiled. The Random Forest Algorithm ranked the main attributes that increase supply chain capacity to predict delays, namely: Destination City, Shipment Priority Level, Shipment Date (Month/Week) and Consignee Location (State). That information could be used in mitigation actions and contribute to improve predictability of international shipment delays and the supply chain performance.

The Kappa Index associated with the Accuracy, Precision, True Positive, False Positive and AUC rates (ROC Curve) were key to better assess the predictive performance of the binary classification problem on a wider validation scope. The use of the Kappa Index added to the existing literature by providing an additional metric to assess the results achieved. Similar papers have concentrated more on interestingness and performance measures that did not include the Kappa as an analysis reference (GUI *et al*; 2020). This methodology has proven its applicability and is bound to provide supply chain practitioners a new tool of assessing and controlling risks in the international transportation processes. Conclusion remarks and future opportunities of research are presented in the next section as Step 6 of the KDD methodology (Knowledge Presentation).

6 CONCLUSION

Research on supply chain management combined with data analytics has significantly evolved in recent years. Many initiatives have been conducted to test and provide empirical evidence, in which the assessment of data mining models has taken the lead as the main source of comparative analysis. The challenge to expand this current investigation to wider frameworks of analysis and new methodologies has great value for the supply chain research community and practitioners. To expand this knowledge, this paper gathered data from an automotive international supply chain and applied data mining methodologies to predict air freight delays. The research hypothesis has been covered by the confirmation of the efficiency of machine learning techniques to predict delays.

The theoretical and managerial implications along with opportunities for future research were unveiled. First and foremost, the Random Forest algorithm has confirmed the literature as the most suitable classifier for prediction of transportation delay problems. Secondly, proactive actions could be performed when future shipments are classified by the algorithm as having the high probability of delay. A specific follow up could be done in each step of the process to guarantee that the risks are under control and early warning measures are taken avoiding supply chain impact. Thirdly, the key variables that influence the supply chain performance for each region were identified and could be used in mitigation actions that seek to optimize the possible outcomes of air freight transportation. For instance, origin and destination information associated with service prioritization level were amongst the key input data to better predict the outcome of the classification models.

This research also provided supply chain practitioners and researchers a new data mining approach regarding air freight delay management. The key element is to combine supplier, customer and transportation operational data to identify patterns which indicate higher probability of delay occurrence. Secondly, it was proposed a thorough data mining methodology that combines different sampling, attribute selection, dimensionality reduction and classification techniques. In short, this framework could be used as reference in future research or practical work not only in transportation but also on other supply chain risk management fields. Thirdly, the achievement of satisfactory results owns to the consistency of model and validation metrics application. Finally, this paper further expands current machine learning literature applied to air freight management which has been mostly focused on weather, airport structure, flight schedule, ground delay and congestion explanatory attributes.

As limitations of the study, the focus was mainly on air freight shipments. Further research on other logistics modals such as sea freight may yield different results. In addition, different algorithms can be applied to the dataset and further expand the classification techniques employed. Different combinations may lead to improved results. Thirdly, the larger the database the more accurate the algorithm might evolve over time. The research has used data from 2019 and this could be expanded in order to gain further insights. Last but not least, different operational situations were not cross validated to assess possible outcomes related to different conditions such as the Covid-19 impact in the international supply chain.

Taking everything into consideration, a consistent outlook of the main research as regards transportation management associated with data mining techniques has been provided. Companies are increasingly investing in freight initiatives that are bound to unleash unprecedented results based on the new concepts of innovation and optimization. However, this paper has brought evidence that there are potential niches of practical investigation in which there is widespread data available unexplored. As further research directions, it is advisable to expand the sample scope to other countries and continents in order to provide a wider overview of transportation management on a global level. The true impact of international shipping data mining applications will be reached when organizations commit their resources to fully capture the value of data inputs from different contexts of their entire supply chain and not only from their countries of operation.

REFERENCES

AGGARWAL, A. K.; DAVÈ, D. S. An Artificial Intelligence Approach to Curtailing the Bullwhip Effect in Supply Chains. **IUP Journal of Supply Chain Management**, v. 15, n. 4, 2018.

AHA, D.W.; KIBLER, D.; ALBERT, M.K. Instance-based learning algorithms. **Mach Learn,** v. 6, p. 37–66, 1991.

AKTER, S. *et al.* Transforming business using digital innovations: the application of AI, blockchain, cloud and data analytics, **Annals of Operations Research**, v. 1, n. 33, 2020.

ALBERGARIA, M.; JABBOUR, C. J. C. The role of big data analytics capabilities (BDAC) in understanding the challenges of service information and operations management in the sharing economy: Evidence of peer effects in libraries. **International Journal of Information Management**, v. 51, p. 102023, 2020.

AMIRKOLAII, K. N *et al.* Demand Forecasting for Irregular Demands in Business Aircraft Spare Parts Supply Chains by using Artificial Intelligence (AI). **IFAC-PapersOnLine**, v. 50, n. 1, p. 15221-15226, 2018.

AMO, S. D. Técnicas de Mineração de Dados. **Jornada de Atualização em Informática**, Universidade Federal de Uberlândia, 2010.

ANDERSSON, J.; JONSSON, P. Big data in spare parts supply chains: The potential of using product-in-use data in aftermarket demand planning. **International Journal of Physical Distribution & Logistics Management**, v. 48, n. 5, pp. 524-544, 2018.

ARYA, V. *et al.* An exploratory study on supply chain analytics applied to spare parts supply chain. **Benchmarking**, v. 24, n. 6, p. 1571-1580, 2017.

BARYANNIS, G.; DANI, S. Antoniou, G. Predicting supply chain risks using machine learning: The trade-off between performance and interpretability. **Future Generation Computer Systems,** v. 101, p. 993-1004, 2019.

BARYANNIS, G. *et al.* Supply chain risk management and artificial intelligence: state of the art and future research directions. **International Journal of Production Research,** v. 57, n. 7, pp. 2179-2202, 2019b.

BATISTA, G.; PRATI, R.; MONARD M. A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data. **SIGKDD Explorations**, v. 6, n. 1, p. 20-29, 2004.

BEN-DAYA, M.; HASSINI, E.; BAHROUN, Z. Internet of things and supply chain management: a literature review. **International Journal of Production Research**, p. 1-24, 2017.

BOLDT, L. C. *et al.* Forecasting Nike's Sales using Facebook Data. **IEEE International Conference on Big Data**, 2016.

BORADE, A. B.; SWEENEY, E. Decision support system for vendor managed inventory supply chain: A case study. **International Journal of Production Research**, v. 53, n. 16, p. 4789-4818, 2015.

BOWERS, M. R.; PETRIE, A. G.; HOLCOMB, M. C. Unleashing the Potential of Supply Chain Analytics. **MIT Sloan Management Review**, v. 59, n. 1, p. 14-16, 2017.

BREIMAN, L. Random Forests. **Machine Learning**, Boston, v.45, n.1, p.5-32, 2001.

BRINCH, M. Understanding the value of big data in supply chain management and its business processes: Towards a conceptual framework. **International Journal of Operations & Production Management**, v. 38, n. 7, p. 1589-1614, 2018.

BUMBLAUSKAS, D. *et al.* Smart Maintenance Decision Support Systems (SMDSS) based on corporate big data analytics. **Expert Systems with Applications**, v. 90, p. 303-317, 2017.

CANBEK, G. *et al*. Binary classification performance measures/metrics: A comprehensive visualized roadmap to gain new insights. **International Conference on Computer Science and Engineering (UBMK)**, p. 821-826, 2017.

CARBONNEAU, R.; VAHIDOV, R.; LAFRAMBOISE, K. Machine learning-based demand forecasting in supply chains. **International Journal of Intelligent Information Technologies**, v. 3, n. 4, p. 40-57, 2007.

CARBONNEAU, R.; LAFRAMBOISE, K.;VAHIDOV, R. Application of machine learning techniques for supply chain demand forecasting. **European Journal of Operational Research**, v. 184, n. 3, p. 1140-1154, 2008.

CARBONNEAU, R.; VAHIDOV, R.; LAFRAMBOISE, K. Forecasting supply chain demand using machine learning algorithms. In: **Machine Learning**: Concepts, Methodologies, Tools and Applications: IGI Global, p. 1652-1686, 2012.

CASTILLO-VILLAR, K. K.; HERBERT-ACERO, J. F. The effect of individual representation on the performance of a genetic algorithm applied to a supply chain network design problem. **International Journal of Supply Chain Management**, v. 2, n. 3, p. 17-24, 2013.

CAVALCANTE, I. M. *et al.* A supervised machine learning approach to data-driven simulation of resilient supplier selection in digital manufacturing. **International Journal of Information Management**, v. 49, p. 86-97, 2019.

CHAHARSOOGHI, S. K.; HEYDARI, J.; ZEGORDI, S. H. A reinforcement learning model for supply chain ordering management: An application to the beer game. **Decision Support Systems**, v. 45, n. 4, p. 949-959, 2008.

CHAVEZ, R. *et al.* Data-Driven Supply Chains, Manufacturing Capability and Customer Satisfaction. **Production Planning & Control**, v. 28, n. 11–12, p. 906–918, 2017.

CHAWLA *et al*. Synthetic Minority Over-sampling Technique. **Journal of Artificial Intelligence Research**. v. 16, p. 321-357, 2002.

CHEN, C.; XU, C. A Negotiation Optimization Strategy of Collaborative Procurement with Supply Chain Based on Multi-Agent System. **Mathematical Problems in Engineering**, 2018.

CHENG, O. K. M.; LAU, R. Y. K; DESTECH PUBLICAT, I. Exploring Big Data Analytics for Supply Chain Management. p. 1111-1117, 2016.

CHI, H. M. *et al.* Modeling and optimizing a vendor managed replenishment system using machine learning and genetic algorithms. **European Journal of Operational Research**, v. 180, n. 1, p. 174-193, 2007.

CHOI, T.M.; CHIU, C.H.; CHAN, H.K. Risk management of logistics systems. **Transportation Research Part E**: Logistics and Transportation Review, v. 90, p. 1–6, 2016.

CHOI, T.-M. Incorporating social media observations and bounded rationality into fashion quick response supply chains in the big data era. **Transportation Research Part E**: Logistics and Transportation Review, v. 114, p. 386-397, 2018.

CHOPRA, S.; MEINDL, P. **Gestão da Cadeia de Suprimentos** – Estratégia, planejamento e operação. 6 ª Edição, São Paulo: Pearson-Prentice Hall, 2016.

CHRISTOPHER, M. **Logistics and Supply Chain Management**. Pitman Publishing, London, 1992.

CHUNG, S.H. *et al.* Data Science and analytics in aviation. **Transportation Research Part E**: Logistics and Transportation Review, v. 134, 2020.

COHEN, J. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, v. 20, pp. 37-46, 1960.

CÔRTE-REAL, N.; OLIVEIRA, T.; RUIVO, P. Assessing Business Value of Big Data Analytics in European Firms. **Journal of Business Research**, v. 70, p. 379–390, 2017.

COX, M.; ELLSWORTH, D. Application-Controlled Demand Paging for Out-of-Core Visualization, **Report NAS-97-010**, July 1997.

CURCIO, D. *et al.* Pharmaceutical routes optimization using artificial intelligence techniques, **Proceedings of the 4th Ieee Workshop on Intelligent Data Acquisition and Advanced Computing Systems**: Technology and Applications, p. 238-242, 2007.

DE SANTIS, R. B.; DE AGUIAR, E.P.; GOLIATT, L. Predicting Material Backorders in Inventory Management using Machine Learning. **IEEE Latin American Conference on Computational Intelligence (La-Cci)**, 2017.

DIETTERICH, T. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization, Machine Learning, v. 1, n. 22, 1998

DHL Logistics Trends. Disponível em: https://www.dhl.com/global-en/home/insights-and-innovation/insights/logistics-trend-radar.html, Acesso em: 02, Maio, 2021.

DIANA, T. Can machines learn how to forecast taxi-out time? A comparison of predictive models applied to the case of Seattle/Tacoma International Airport. **Transportation Research Part E**: Logistics and Transportation Review, v. 119, p. 149-164, 2018.

DIDEHKHANI, H.; JASSBI, J.; PILEVARI, N. Assessing flexibility in supply chain using adaptive neuro fuzzy inference system. **IEEE International Conference on Industrial Engineering and Engineering Management** (IEEM 2009), Hong-Kong, 2009.

DUBEY, R. *et al.* The Impact of Big Data on World-Class Sustainable Manufacturing. **The International Journal of Advanced Manufacturing Technology**. v. 84, n. 1–4, pp. 631–645, 2016.

DUBEY, R. *et al.* Big data and predictive analytics in humanitarian supply chains: Enabling visibility and coordination in the presence of swift trust. **The International Journal of Logistics Management**, v. 29, n. 2, p. 485-512, 2018.

DUBEY, R. *et al.* Big data and predictive analytics and manufacturing performance: integrating institutional theory, resource-based view and big data culture. **British Journal of Management**, v. 30, n. 2, p. 341-361, 2019a.

DUBEY, R. *et al.* Big data analytics and artificial intelligence pathway to operational performance under the effects of entrepreneurial orientation and environmental dynamism: A study of manufacturing organizations. **International Journal of Production Economics,** v. 107599, 2019b.

DUBEY, R. *et al.* Can big data and predictive analytics improve social and environmental sustainability? **Technological Forecasting and Social Change**, v. 144, p. 534.545, 2019c.

DUBEY, R.; GUNASEKARAN, A.; CHILDE, S. J. Big data analytics capability in supply chain agility. **Management Decision**, v. 57, n. 8, p. 2092-2112, 2019d.

DWEEKAT, A. J.; HWANG, G.; PARK, J. A supply chain performance measurement approach using the internet of things: Toward more practical SCPMS. **Industrial Management & Data Systems**, v. 117, n. 2, p. 267-286, 2017.

EFENDIGIL, T.; ÖNÜT, S.; KAHRAMAN, C., 2009. A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: A comparative analysis. **Expert Systems with Applications**, v. 36, n. 3 PART 2, p. 6697-6707, 2009.

EHRET, M.; WIRTZ, J. Unlocking value from machines: business models and the industrial internet of things. **Journal of Marketing Management**, v. 33, n. 1-2, p. 111-130, 2017.

ENGELSETH, P.; WANG, H. Big data and connectivity in long-linked supply chains. **Journal of Business & Industrial Marketing**, v. 33, n. 8, p. 1201-1208, 2018.

ER KARA, M.; OKTAY FIRAT, S. Ü.; GHADGE, A. A data mining-based framework for supply chain risk management. **Computers and Industrial Engineering,** v. 139, 2020.

ETANI, N., 2019. Development of a predictive model for on-time arrival flight of airliner by discovering correlation between flight and weather data. **Journal of Big Data,** v. 6, p. 85.

FARAHANI, R.; REZAPOUR, S.; KARDAR, L. Logistics Operations and Management: Concepts and Models, 2011.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, American Association for Artificial Intelligence, Califórnia, USA, v. 17, n. 3, p. 37-54, 1996.

FERNANDO, Y.; CHIDAMBARAM, R. R. M.; WAHYUNI-TD, I. S. The impact of Big Data analytics and data security practices on service supply chain performance. **Benchmarking**, v. 25, n. 9, p. 4009-4034, 2018.

FIX, E.; HODGES, J.L. Discriminatory analysis, nonparametric discrimination: Consistency properties. **Technical Report 4:** USAF School of Aviation Medicine, Randolph Field, Texas, 1951.

FRANK, E.; HALL, M.A.; WITTEN, I.H. The Weka Workbench, Fourth Edition, 2016.

FU, J.; FU, Y. An adaptive multi-agent system for cost collaborative management in supply chains. **Engineering Applications of Artificial Intelligence**, v. 44, p. 91-100, 2015.

GARG, V. K.; VISWANADHAM, N. EcoSupply: A Machine Learning Framework for Analyzing the Impact of Ecosystem on Global Supply Chain Dynamics. **Asia-Pacific Conference on Simulated Evolution and Learning**, p. 677-686, 2010.

GARDNER, M.W.; DORLING, S.R. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. **Atmospheric Environment,** v. 32, n. 14/15, p. 2627—2636, 1998

GERON, A. Hands-on machine learning with Scikit-Learn, Keras & TensorFlow. Concepts, Tools and Techniques to Build Intelligent Systems. Second Edition, 2019.

GIANNAKIS, M.; LOUIS, M. A multi-agent based framework for supply chain risk management. **Journal of Purchasing and Supply Management**, v. 17, n. 1, p. 23-31, 2011.

GIANNAKIS, M.; LOUIS, M. A multi-agent based system with big data processing for enhanced supply chain agility. **Journal of Enterprise Information Management**, v. 29, n. 5, p. 706-727, 2016.

GIANNOCCARO, I.; PONTRANDOLFO, P. Inventory management in supply chains: A reinforcement learning approach. **International Journal of Production Economics**, v. 78, n. 2, p. 153-161, 2002.

GSF (GLOBAL SHIPPERS FORUM Briefing). The value of air cargo to the global economy. Disponível em http://fma-agf.ca/CMFiles/Blog/GSFTakeaways/Thevalueofaircargototheglobaleconomy.pdf. Acesso em: 02, Maio, 2015.

GOVINDAN, K.; CHAUDHURI, A. Interrelationships of risks faced by third party logistics service providers: A DEMATEL based approach. **Transportation Research Part E**: Logistics and Transportation Review, v. 90, p. 177-195, 2016.

GRAVILI, G. *et al.* The influence of the Digital Divide on Big Data generation within supply chain management. **International Journal of Logistics Management**, v. 29, n. 2, p. 592-628, 2018.

GRAWE, S. J. Logistics innovation: a literature-based conceptual framework, **International Journal of Logistics Management**, v. 20, n. 3, p. 360-377, 2009.

GUI, G. *et al.* Flight Delay Prediction Based on Aviation Big Data and Machine Learning. **IEEE Transactions on Vehicular Technolog**y, v. 69, n. 1, p. 140-150, 2020.

GUNASEKARAN, A. *et al.* Big data and predictive analytics for supply chain and organizational performance. **Journal of Business Research**, v. 70, p. 308-317, 2017.

GUOSHENG, H.; GUOHONG, Z. Comparison on neural networks and support vector machines in suppliers' selection. **Journal of Systems Engineering and Electronics**, v. 19, n. 2, pp. 316-320, 2008.

GUPTA, S. Role of cloud ERP and big data on firm performance: a dynamic capability view theory perspective, **Management Decision**, v. 57, n. 8, p. 1857-1882, 2019.

GYULAI, D. *et al.* Lead time prediction in a flow-shop environment with analytical and machine learning approaches. **IFAC-PapersOnLine**, v. 51, n. 11, pp. 1029-1034, 2018.

HALL, M.A. **Correlation-based Feature Subset Selection for Machine Learning**. Hamilton, New Zealand, 1998.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining**: Concepts and Techniques, 3rd edition, Morgan Kaufmann, 2011.

HASSAN, A. P. Enhancing Supply Chain Risk Management by Applying Machine Learning to Identify Risks. **Springer**. p.191-205, 2019.

HAZEN, B. T. *et al.* 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.

HAZEN, B. T. *et al.* Big Data and predictive analytics for supply chain sustainability: A theory-driven research agenda. **Computers & Industrial Engineering**, v. 101, p. 592-598, 2016.

HAZEN, B. T. *et al.* Back in business: operations research in support of big data analytics for operations and supply chain management. **Annals of Operations Research**, v. 270, n. 1-2, p. 201-211, 2018.

HERREMA, F. *et al.* A machine learning model to predict runway exit at Vienna Airport. **Transportation Research Part E**: Logistics and Transportation Review, v. 131, pp. 329-342, 2019.

HIROMOTO, R. E.; HANEY, M.; VAKANSKI, A. A Secure Architecture for IoT with Supply Chain Risk Management. **Proceedings of the 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems**: Technology and Applications, 2017.

HO, W. *et al.* Supply chain risk management: a literature review. **International Journal of Production Research**, v. 53, n. 16, p. 5031–5069, 2015.

HOFMANN, E., 2017. Big Data and Supply Chain Decisions: The Impact of Volume, Variety and Velocity Properties on the Bullwhip Effect. **International Journal of Production Research**, v. 55, n. 17, pp. 5108–5126, 2017.

HOFMANN, E.; RUTSCHMANN, E. Big data analytics and demand forecasting in supply chains: a conceptual analysis. **The International Journal of Logistics Management**, v. 29, n. 2, p. 739-766, 2018.

HONG, G. H.; HA, S. H. Evaluating supply partner's capability for seasonal products using machine learning techniques. **Computers and Industrial Engineering**, v. 54, n. 4, p. 721-736, 2008.

HOPKINS, J.; HAWKING, P. Big Data Analytics and IoT in logistics: a case study. **International Journal of Logistics Management**, v. 29, n. 2, p. 575-591, 2018.

HU, H. *et al.* Toward scalable systems for big data analytics: **a technology tutorial IEEE Access**, v. 2, p. 652-687, 2014.

ILIE-ZUDOR, E. *et al.* Advanced predictive-analysis-based decision support for collaborative logistics networks. **Supply Chain Management:** An International Journal, v. 20, n. 4, pp. 369-388, 2015.

ITTMANN, H. W. The impact of big data and business analytics on supply chain management. **Journal of Transport and Supply Chain Management**, v. 9, n. 1, 2015.

IVANOV, D.; DOLGUI, A.; SOKOLOV, B. The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics. **International Journal of Production Research**, v. 57, n. 3, p. 829-846, 2019.

JAIN, N.; SRIVASTAVA, V. Data Mining Techniques: A Survey Paper. **International Journal of Research in Engineering and Technology**, v. 02, n.11, Nov-2013.

JAFARZADEH-GHOUSHCHI, S., RAHMAN, M.N.A. Performance study of artificial neural network modelling to predict carried weight in the transportation system. **International Journal of Logistics Systems and Management**, v. 24, n. 2, 2016.

JEBLE, S.; KUMARI, S.; PATIL, Y. Role of big data in decision making. **Operations and Supply Chain Management**, v. 11, n. 1, p. 36-44, 2018.

JIANG, C.; SHENG, Z. Case-based reinforcement learning for dynamic inventory control in a multi-agent supply-chain system. **Expert Systems with Applications**, v. 36, n. 3, p. 6520-6526, 2009.

KACHE, F.; SEURING, S. Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management. **International Journal of Operations & Production Management**, v. 37, n. 1, p. 10-36, 2017.

KAR, A. K. A hybrid group decision support system for supplier selection using analytic hierarchy process, fuzzy set theory and neural network. **Journal of Computational Science**, v. 6, p. 23-33, 2015.

KARTAL, H, *et al.* An integrated decision analytic framework of machine learning with multi-criteria decision making for multi-attribute inventory classification. **Computers & Industrial Engineering**, v. 101, p. 599-613, 2016.

KAUR, H.; SINGH, S. P. Heuristic modeling for sustainable procurement and logistics in a supply chain using big data. **Computers & Operations Research**, v. 98, p. 301-321, 2018.

KAZEMI, A.; FAZEL ZARANDI, M. H. An agent-based framework for building decision support system in supply chain management. **Journal of Applied Sciences**, v. 8, n. 7, pp. 1125-1137, 2008.

KHERBASHA, O.; MOCANA, M. A Review of Logistics and Transport Sector as a Factor of Globalization. 22nd International Economic Conference – **IECS**, 2015.

KIEKINTVELD, C. *et al.* Forecasting market prices in a supply chain game. **Electronic Commerce Research and Applications**, v. 8, n. 2, p. 63-77, 2009.

KOHAVI, R. Feature subset selection as search with probabilistic estimates, **AAAI Fall Symposium on Relevance**, p. 122-126, 1994

KOHAVI, R.; SOMMERFIELD, D. Feature subset selection using the wrapper model: Overfitting and dynamic search space topology. **Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95)**, 1995

KOHAVI, R.; JOHN, G.H. Wrappers for feature subset selection. **Artificial Intelligence**, v. 97, n. 1-2, p. 273-324, 1997.

KONG, F.; LI, J. Supply chain flexibility enhancement based on deep belief network. **Jisuanji Jicheng Zhizao Xitong/Computer Integrated Manufacturing Systems**, CIMS, v. 24, n. 5, p. 1292-1300, 2018.

KUMAR, V.; MINZ, S. Feature Selection: a literature review. **Smart Computing Review**, v.4, n. 3, June 2014.

KUMAR, D. *et al.* A fuzzy logic based decision support system for evaluation of suppliers in supply chain management practices. **Mathematical and Computer Modelling**, v. 58, n. 11-12, p. 1679-1695, 2013.

LAI, Y.; SUN, H.; REN, J. Understanding the determinants of big data analytics (BDA) adoption in logistics and supply chain management: An empirical investigation. **International Journal of Logistics Management**, v. 29, n. 2, p. 676-703, 2018.

LAU, R. Y. K.; ZHANG, W.; XU, W. Parallel Aspect-Oriented Sentiment Analysis for Sales Forecasting with Big Data. **Production and Operations Management**, v. 27, n. 10, p. 1775-1794, 2018.

LEE, C. K. H. A GA-based optimization model for big data analytics supporting anticipatory shipping in Retail 4.0. **International Journal of Production Research**, v. 55, n.2, p. 593-605, 2017.

LI, J. *et al.* Big Data in Product Lifecycle Management. **The International Journal of Advanced Manufacturing Technology**, v. 81, n.1–4, p. 667–684, 2015.

LIMA-JUNIOR, F.R.; CARPINETTI, L.C.R. Combining SCOR® model and fuzzy TOPSIS for supplier evaluation and management. **International Journal Production Economics**, v. 174, p. 128-141, 2016.

LIU, Y. *et al.* Using machine learning to analyze air traffic management actions: ground delay program case study. **Transportation Research Part E**: Logistics and Transportation Review, v. 131, pp. 80-95, 2019.

LIU, H.; SETIONO, R. Chi2: Feature selection and discretization of numeric attributes. **Proceedings of the IEEE 7th International Conference on Tools with Artificial Intelligence**, p. 388–391, 1995.

MA, H.; WANG, Y.; WANG, K. Automatic detection of false positive RFID readings using machine learning algorithms. **Expert Systems with Applications**, v. 91, p. 442-451, 2018.

MAHROOF, K. A human-centric perspective exploring the readiness towards smart warehousing: The case of a large retail distribution warehouse. **International Journal of Information Management**, v. 45, p. 176-190, 2019.

MAKRIDAKIS, S.; SPILIOTIS E.; ASSIMAKOPOULOS V. Statistical and Machine Learning forecasting methods: Concerns and ways forward. **PLoS ONE 13**, v. 3, 2018.

MANDAL, S. An examination of the importance of big data analytics in supply chain agility development: A dynamic capability perspective. **Management Research Review**, v. 41, n. 10, p. 1201-1219, 2018.

MANDAL, S. The influence of big data analytics management capabilities on supply chain preparedness, alertness and agility: An empirical investigation. **Information Technology & People**, v. 32, n. 2, p. 297-318, 2019.

MANI, V. *et al.* Mitigating supply chain risk via sustainability using big data analytics: Evidence from the manufacturing supply chain. **Sustainability** (Switzerland), v. 9, n. 4, 2017.

MATTHIAS, O. *et al.* Making sense of big data–can it transform operations management? **International Journal of Operations & Production Management**, v. 37, n. 1, p. 37-55, 2017.

MIKALEF, P. *et al.* Big data analytics capabilities and innovation: the mediating role of dynamic capabilities and moderating effect of the environment, **British Journal of Management**, v. 30 n. 2, p. 272-298.

MIN, H. Artificial intelligence in supply chain management: Theory and applications. **International Journal of Logistics Research and Applications**, v.13, n. 1, p. 13-39, 2010.

MOJAVERI, H. R. S. *et al.* Validation and selection between machine learning technique and traditional methods to reduce bullwhip effects: A data mining approach. **World Academy of Science, Engineering and Technology**, v. 37, p. 555-561, 2009.

MOKHTARINEJAD, M. *et al.* A novel learning based approach for a new integrated location-routing and scheduling problem within cross-docking considering direct shipment. **Applied Soft Computing Journal**, v. 34, p. 274-285, 2015.

MOKTADIR, M. A. *et al.* Barriers to big data analytics in manufacturing supply chains: A case study from Bangladesh. **Computers and Industrial Engineering**, v. 128, p. 1063-1075, 2019.

MORAGA, R. *et al.* Using neural networks to monitor supply chain behaviour. **International Journal of Computer Applications in Technology**, v. 40, n. 1-2, p. 53-63, 2011.

MORETTO, A. ; RONCHI, S. ; PATRUCCO, A. S. Increasing the effectiveness of procurement decisions: The value of big data in the procurement process. **International Journal of Rf Technologies-Research and Applications**, v. 8, n. 3, p. 79-103, 2017.

MUBAREK, A.M.; ADALI, E. Multilayer perceptron neural network technique for fraud detection. **International Conference on Computer Science and Engineering** (UBMK), 2017.

NITA, S. Application of big data technology in support of food manufacturers' commodity demand forecasting. **NEC Technical Journal**, v. 10, n. 1, p. 90-93, 2015.

NIU, B., DAI, Z., ZHUO, X. Co-opetition effect of promised-delivery-time sensitive demand on air cargo carriers' big data investment and demand signal sharing decisions. **Transportation Research Part E**: Logistics and Transportation Review, v. 123, p. 29-44, 2019.

ORJI, I. J.; WEI, S. An innovative integration of fuzzy-logic and systems dynamics in sustainable supplier selection: A case on manufacturing industry. **Computers and Industrial Engineering**, v. 88, p. 1-12, 2015.

PAPADOPOULOS, T. *et al*. The role of Big Data in explaining disaster resilience in supply chains for sustainability. **Journal of Cleaner Production**, v. 142, p. 1108-1118, 2017.

PARK, Y. B.; YOON, S. J.; YOO, J. S. Development of a knowledge-based intelligent decision support system for operational risk management of global supply chains. **European Journal of Industrial Engineering**, v. 12, n. 1, p. 93-115, 2018.

PEREIRA, M. M. *et al.* Predictive and Adaptive Management Approach for Omnichannel Retailing Supply Chains. **IFAC-PapersOnLine**, v. 51, n. 11, p. 1707-1713, 2018.

PIRAMUTHU, S. *et al.* Efficient Genetic Algorithm Based Data Mining Using Feature Selection with Hausdorff Distance. **Information and Technology Management**, v. 6, p. 315-331, 2005a.

PIRAMUTHU, S. Machine learning for dynamic multi-product supply chain formation. **Expert Systems with Applications**, v. 29, n. 4, p. 985-990, 2005b.

PIRAMUTHU, S. Knowledge-based framework for automated dynamic supply chain configuration - Production, manufacturing and logistics. **European Journal of Operational Research**, v. 165, n. 1, p. 219-230, 2005c.

PLATT, J. Fast Training of Support Vector Machines using Sequential Minimal Optimization. **Advances in Kernel Methods - Support Vector Learning**, 1998.

PONTRANDOLFO, P. *et al.* Global supply chain management: A reinforcement learning approach. **International Journal of Production Research**, v. 40, n. 6, p. 1299-1317, 2002.

PRASAD, S., ZAKARIA, R., ALTAY, N. Big data in humanitarian supply chain networks: A resource dependence perspective. **Annals of Operations Research**, v. 270, n. 1-2, p. 383-413, 2018.

PRATI, R.C. *et al.* Curvas ROC para avaliação de classificadores. **IEEE Latin America Transactions**, v. 6, n. 2, June 2008.

QUEIROZ, M. M.; TELLES, R. Big data analytics in supply chain and logistics: an empirical approach. **International Journal of Logistics Management,** v. 29, n. 2, p. 767-783, 2018.

QUINLAN, J. Induction of decision trees. **Machine Learning**, v. 1, n. 1, p. 81-106, 1986.

RAMPERSAD, G. Robot will take your job: Innovation for an era of artificial intelligence. **Journal of Business Research**, v. 116, p. 68-74, 2020.

RAUT, R.D. *et al.* A hybrid approach using data envelopment analysis and artificial neural network for optimizing 3PL supplier selection. **International Journal of Logistics Systems and Management**, v. 26, n. 2, 2017.

RAUT, R.D. *et al.*. Linking big data analytics and operational sustainability practices for sustainable business management. **Journal of Cleaner Production**, v. 224, pp. 10-24, 2019.

REHMAN, M.H.U. *et al.* Big data reduction framework for value creation in sustainable enterprises. **International Journal of Information Management**, v. 36, n. 6, p. 917-928, 2016.

REN, S. *et al.* A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: A framework, challenges and future research directions**. Journal of Cleaner Production**, v. 210, p. 1343-1365, 2019.

RICHEY, R. G *et al.* A global exploration of big data in the supply chain. **International Journal of Physical Distribution & Logistics Management**, v. 46, n. 8, pp. 710-739, 2016.

RODRIGUEZ, L.; DA CUNHA, C. Impacts of big data analytics and absorptive capacity on sustainable supply chain innovation: a conceptual framework. **Logforum**, v. 14, n. 2, p. 151-161, 2018.

ROßMANN, B. *et al.* The future and social impact of Big Data Analytics in Supply Chain Management: Results from a Delphi study. **Technological Forecasting and Social Change**, v. 130, pp. 135-149, 2018.

RUSSELL, S.; NORVIG, P. **Artificial intelligence**: a modern approach, Upper Saddle River, NJ: Prentice-Hall. 1995.

SANCHEZ-RODRIGUES, V., POTTER, A., NAIM, M. M. The impact of logistics uncertainty on sustainable transport operations**. International Journal of Physical Distribution & Logistics Management**, v.40, p.61–83, 2010.

SANCHEZ-RODRIGUES V. *et al.* Establishing a transport focused uncertainty model for the supply chain. **International Journal of Physical Distribution and Logistics Management** v. 38, n. 5, p. 388–411, 2008.

SANDERS, E. Air Cargo Market is Projected to Grow by 2025. CAAS (Cargo Airports & Airline Services), 2020. Disponível em: https://www.caasint.com/air-cargo-market-is-projected-to-grow-by-2025. Acesso em: 02, Maio, 2021.

SANDERS, N. R. How to use big data to drive your supply chain. **California Management Review**, v. 58, n. 3, p. 26-48, 2016.

SUPPLY CHAIN COUNCIL. Supply Chain Operations Reference Model. Supply Chain Council, 2012.

SCHAFER, C. Selecting a classification method by cross-validation. **Machine Learning**, v.13, n. 1, p. 135-143, October 1993.

SCHOENHERR, T.; SPEIER-PERO, C. Data science, predictive analytics, and big data in supply chain management: Current state and future potential. **Journal of Business Logistics**, v. 36, n. 1, p. 120-132, 2015.

SHAFIQUE, U.; QAISER, H. A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA, International Journal of Innovation and Scientific Research, v. 12, n. 1, p. 217-222, 2014.

SHAHRABI, J.; MOUSAVI, S. S.; HEYDAR, M. Supply chain demand forecasting: A comparison of machine learning techniques and traditional methods. **Journal of Applied Sciences**, v. 9, n. 3, p. 521-527, 2009.

SHOKOUHYAR, S. *et al.* Implementing a fuzzy expert system for ensuring information technology supply chain. **Expert Systems**, v. 36, n. 1, 2019.

SHUKLA, M.; TIWARI, M. K. Big-data analytics framework for incorporating smallholders in sustainable palm oil production. **Production Planning & Control**, v. 28, n. 16, p. 1365-1377, 2017.

SIMCHI-LEVI, D.; WU, M. X. Powering retailers' digitization through analytics and automation. **International Journal of Production Research**, v. 56, n. 1-2, p. 809-816, 2018.

SINGH, L. P.; CHALLA, R. T. Integrated Forecasting Using the Discrete Wavelet Theory and Artificial Intelligence Techniques to Reduce the Bullwhip Effect in a Supply Chain. **Global Journal of Flexible Systems Management**, v. 17, n. 2, p. 157-169, 2016.

SINGH, S. K.; EL-KASSAR, A. N. Role of big data analytics in developing sustainable capabilities. **Journal of Cleaner Production**, v. 213, p. 1264-1273, 2019.

SINGH, A.; SHUKLA, N.; MISHRA, N. Social media data analytics to improve supply chain management in food industries. **Transportation Research Part E**: Logistics and Transportation Review, v. 114, p. 398-415, 2018.

SIURDYBAN, A.; MØLLER, C. Towards intelligent supply chains: A unified framework for business process design. **International Journal of Information Systems and Supply Chain Management**, v.5, n.1, pp. 1-19, 2012.

SLIMANI, I. Configuration and implementation of a daily artificial neural network-based forecasting system using real supermarket data. **International Journal of Logistics Systems and Management**, v. 28, n. 2, p. 144-163, 2017.

SLIMANI, I., EL FARISSI, I., ACHCHAB, S. Application of game theory and neural network to study the behavioral probabilities in supply chain. **Journal of Theoretical and Applied Information Technology**, v. 82, n. 3, p. 411-416, 2015.

SMYTH, K. B. *et al.* Thirsty in an Ocean of Data? Pitfalls and Practical Strategies When Partnering With Industry on Big Data Supply Chain Research. **Journal of Business Logistics**, v. 39, n. 3, p. 203-219, 2018.

SODERO, A.; JIN, Y. H.; BARRATT, M. The social process of Big Data and predictive analytics use for logistics and supply chain management. **International Journal of Physical Distribution & Logistics Management**, v. 49, n. 7, 2019.

SUN, Z.-L. *et al.* Sales forecasting using extreme learning machine with applications in fashion retailing. **Decision Support Systems**, v. 46, n. 1, p. 411-419, 2008.

SUNIL, C.; MEINDL, P. **Supply Chain Management**: Strategy, Planning and Operation. Global Edition, 2015.

TIWARI, S.; WEE, H. M.;DARYANTO, Y. Big data analytics in supply chain management between 2010 and 2016: Insights to industries. **Computers & Industrial Engineering**, n. 115, p. 319-330, 2018.

TORGO, L. *et al.* Smote for Regression, **Portuguese Conference on Artificial Intelligence**, p. 378-389, 2013.

TRIPATHI, S.; GUPTA, M. Transforming towards a smarter supply chain. **International Journal of Logistics Systems and Management**, v. 36 n. 3, p.319–342, 2020.

TSANG, Y. P. *et al.* An Internet of Things (IoT)-based risk monitoring system for managing cold supply chain risks. **Industrial Management and Data Systems**, v. 118, n. 7, p. 1432-1462, 2018.

TSAO, Y. C. Managing default risk under trade credit: Who should implement Big-Data analytics in supply chains? **Transportation Research Part E**: Logistics and Transportation Review, v. 106, p. 276-293, 2017.

TSE, Y. K.; CHAN, T. M.; LIE, R. H. Solving complex logistics problems with multi-artificial intelligent system. **International Journal of Engineering Business Management**, v. 1, n. 1, p. 37-48, 2009.

URCIUOLI, L.; HINTSA, J. Improving supply chain risk management - can additional data help? **International Journal of Logistics Systems and Management**, v. 30, n. 2,

p. 195, 2018.

VAHDANI, B. *et al.* An artificial intelligence approach for fuzzy possibilistic-stochastic multi-objective logistics network design. **Neural Computing & Applications**, v. 25, p. 7-8, p. 1887-1902, 2014.

VALLURI, A.; NORTH, M. J.; MACAL, C. M. Reinforcement learning in supply chains. **International Journal of Neural Systems**, v. 19, n. 5, pp. 331-344, 2009.

VAPNIK, V.; A. CHERVONENKIS. A note on one class of perceptrons. **Automation and Remote Control**, v. 25, 1964.

VAPNIK, V., **Estimation of Dependences Based on Empirical Data**, Springer-Verlag, 1982

VIELLECHNER, A., SPINLER, S. Novel Data Analytics Meets Conventional Container Shipping: Predicting Delays by Comparing Various Machine Learning Algorithms. **Proceedings of the 53rd Hawaii International Conference on System Sciences**, 2020.

WALLACE, T.F.; DOUGHERTY, J.R. APICS Dictionary, 6ª Edição, American Production and Inventory Control Society, 1987.

WALLER, M. A.; FAWCETT, S. E. Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. **Journal of Business Logistics,** v. 34, n. 2, p. 77-84, 2013a.

WALLER, M. A., FAWCETT, S. E. Click here for a data scientist: Big data, predictive analytics, and theory development in the era of a maker movement supply chain. **Journal of Business Logistics**, 34, n. 4, pp. 249-252, 2013b.

WAMBA, S.F. *et al.* Big Data Analytics and Firm Performance: Effects of Dynamic Capabilities. **Journal of Business Research**, v. 70, p. 356–365, 2017.

WAMBA, S. F. *et al.* The performance effects of big data analytics and supply chain ambidexterity: The moderating effect of environmental dynamism, **International Journal of Production Economics**, v. 222, p.107498, 2020.

WANG, G. *et al.* Big data analytics in logistics and supply chain management: Certain investigations for research and applications. **International Journal of Production Economics**, v. 176, p. 98-110, 2016.

WANKE, P. *et al.* Fuzzy inference systems and inventory allocation decisions: Exploring the impact of priority rules on total costs and service levels. **Expert Systems with Applications**, v. 85, p. 182-193, 2017.

WIECZOREK, L.; IGNACIUK, P. Continuous Genetic Algorithms as Intelligent Assistance for Resource Distribution in Logistic Systems. **Data**, v. 3, n. 4, 2018.

WONG, J. T.; SU, C. T.; WANG, C. H. Stochastic dynamic lot-sizing problem using bi-level programming base on artificial intelligence techniques. **Applied Mathematical Modelling**, v. 36, n. 5, p. 2003-2016, 2012.

WU, P.-J.; CHEN, M.-C.; TSAU, C.-K. The data-driven analytics for investigating cargo loss in logistics systems. **International Journal of Physical Distribution & Logistics Management**, v. 47, n. 1, p. 68-83, 2017.

WU, K. J. *et al.* Toward Sustainability: Using Big Data to Explore the Decisive Attributes of Supply Chain Risks and Uncertainties. **Journal of Cleaner Production**, v.142, p. 663–676, 2017.

WU, P. J.; LIN, K. C. Unstructured big data analytics for retrieving e-commerce logistics knowledge. **Telematics and Informatics**, v. 35, n. 1, p. 237-244, 2018.

XU, Z. Y.; SUN, R.; SUN, Y. Z. An application of Artificial Neural Network on performance measurement of supply chain alliance. **Proceedings of the 13th International Conference on Industrial Engineering and Engineering Management**, v. 1-5: Industrial Engineering and Management Innovation in New-Era, 2006.

YADAV, S.; SHUKLA, S. Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. **IEEE 6th International Conference on Advanced Computing (IACC)**, p. 78-83, 2016.

YU, L. *et al.* Online big data-driven oil consumption forecasting with Google trends. **International Journal of Forecasting**, v.35, n.1, p. 213-223, 2019.

YU, B. *et al.* Flight delay prediction for commercial air transport: A deep learning approach. **Transportation Research Part E:** Logistics and Transportation Review, v. 125, pp. 203-221, 2019.

YUEN, J. S. M. *et al.* An intelligent-internet of things (IoT) outbound logistics knowledge management system for handling temperature sensitive products. **International Journal of Knowledge and Systems Science**, v. 9, n. 1, p. 23-40, 2018.

ZHAN, Y. *et al.* Unlocking the power of big data in new product development. **Annals of Operations Research**, v. 270, n. 1-2, p. 577-595, 2018.

ZHANG, Z.; FIGLIOZZI, M. A. A Survey of China's Logistics Industry and the Impacts of Transport Delays on Importers and Exporters. **Transport Reviews**, v.30, n. 2, p. 179-194, 2009.

ZHANG, R. *et al.* Learning to select supplier portfolios for service supply chain. **PLoS ONE**, v. 11, n. 5, 2016.

ZHANG, H.; XU, Z.; LU, J. F. Research of partner enterprise selection in supply chain management based on support vector machine. **Jisuanji Jicheng Zhizao Xitong/Computer Integrated Manufacturing Systems**, CIMS, v. 10, n. 7, p. 796-800, 2004.

ZHAO, R. *et al.* An Optimization Model for Green Supply Chain Management by Using a Big Data Analytic Approach. **Journal of Cleaner Production**, v. 142, p. 1085–1097, 2017.

ZHONG, R. Y. *et al.* Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. **Computers and Industrial Engineering**, v. 101, p. 572-591, 2016.

ZHONG, R. Y. *et al.* Big Data Analytics for Physical Internet-based intelligent manufacturing shop floors. **International Journal of Production Research**, v. 55, n. 9, p. 2610-2621, 2015.

ZHU, Y. *et al.* Comparison of individual, ensemble and integrated ensemble machine learning methods to predict China's SME credit risk in supply chain finance. **Neural Computing and Applications**, v.28, p. 41-50, 2017.

APPENDIX 1 – ATTRIBUTES CLEANING RESULT

Number	Attribute Name	Attribute Description
1	**City Lane**	Origin City - Destination City Route
2	**Country Lane**	Origin Country - Destination Country Route
3	**Origin**	Origin Airport
4	**Origin City**	Origin City Name
5	**Origin Country**	Origin Country
6	**Origin Country Reference**	Origin Consolidated Reference
7	**Origin Region**	Origin Macro Region
8	**Destination**	Destination Airport
9	**Destination City**	Destination City
10	**Ship Date**	Shipment Reference Date
11	**Priority Level**	Freight Forwarder Cargo Priority Level
12	**Service Type**	Type of Expedite Service Provided
13	**Shipper Account**	Supplier Identification Number
14	**Shipper Name**	Supplier Name
15	**Shipper City**	Supplier City
16	**Shipper State**	Supplier State
17	**Shipper Country**	Supplier Billing Country
18	**Consignee City**	Destination Customer City
19	**Consignee State**	Destination Customer State
20	**Consignee Country**	Destination Customer Country
21	**Customer Number**	Customer Number Identification
22	**Export Carrier**	Airline Company
23	**Carrier Code**	Airline Company Code Identification
24	**Freight Received**	Customer Shipment Authorization Date
25	**Docs From Shipper**	Customer Documents Availability Date
26	**Pickup**	Shipment Pick up Date
27	**ATD**	Actual Time of Departure Date
28	**ATA**	Actual Time of Arrival Date
29	**ETA**	Expected Time of Departure Date
30	**ETD**	Expected Time of Arrival Date
31	**Docs Received**	Destination Documents Hand Over Date
32	**Docs to Broker**	Broker Documents Availability Date
33	**POD Date**	Proof of Delivery Date
34	**Delivery Date**	Delivery Date
35	**Due Date**	Reference Contract Delivery Due Date
36	**Total Pieces**	Number of Shipment Handling Units
37	**Actual Weight (Kg)**	Cargo Weight
38	**Charge Weight (Kg)**	Cargo Chargeable Weight
39	**Service Level Airfreight**	Type of Service (expedite or standard service)
40	**US/CN Zone Identification**	United States and China Region Specification
41	**Week**	Shipment Calendar Week
42	**Year/Month**	Shipment Calendar Week (with Year)
43	**Business Unit Division**	Customer Plant/Business Unit Identification
44	**Weekday of Pick Up**	Monday to Sunday Pick up Day Information
45	**Reference Pickup Day**	Customer authorization based on best day of pick up verification
46	***Delay (Target Attribute)***	***Shipment delay occurrence (YES/NO)***

APPENDIX 2 – ATTRIBUTE SELECTION ALGORITHMS RESULT SMO

CFS (Correlation-based Feature Subset Selection)

Attribute Number	Attribute Name	Attribute Description
9	**Destination City**	Destination City
10	**Ship Date**	Shipment Reference Date
14	**Shipper Name**	Supplier Name

Chi-Square Statistical Test

Attribute Number	Attribute Name	Attribute Description
1	**City Lane**	Origin City - Destination City Route
2	**Country Lane**	Origin Country - Destination Country Route
3	**Origin**	Origin Airport
4	**Origin City**	Origin City Name
5	**Origin Country**	Origin Country
6	**Origin Country Reference**	Origin Consolidated Reference
7	**Origin Region**	Origin Macro Region
8	**Destination**	Destination Airport
9	**Destination City**	Destination City
10	**Ship Date**	Shipment Reference Date
11	**Priority Level**	Freight Forwarder Cargo Priority Level
12	**Service Type**	Type of Expedite Service Provided
13	**Shipper Account**	Supplier Identification Number
14	**Shipper Name**	Supplier Name
15	**Shipper City**	Supplier City
16	**Shipper State**	Supplier State
17	**Shipper Country**	Supplier Billing Country
18	**Consignee City**	Destination Customer City
19	**Consignee State**	Destination Customer State
20	**Consignee Country**	Destination Customer Country
21	**Customer Number**	Customer Number Identification
22	**Export Carrier**	Airline Company
23	**Carrier Code**	Airline Company Code Identification
24	**Freight Received**	Customer Shipment Authorization Date
25	**Docs From Shipper**	Customer Documents Availability Date
26	**Pickup**	Shipment Pick up Date
27	**ATD**	Actual Time of Departure Date
28	**ATA**	Actual Time of Arrival Date
29	**ETA**	Expected Time of Departure Date
30	**ETD**	Expected Time of Arrival Date
31	**Docs Received**	Destination Documents Hand Over Date
32	**Docs to Broker**	Broker Documents Availability Date
33	**POD Date**	Proof of Delivery Date
34	**Delivery Date**	Delivery Date
35	**Due Date**	Reference Contract Delivery Due Date
39	**Service Level Airfreight**	Type of Service (expedite or standard service)
40	**US/CN Zone Identification**	United States and China Region Specification
41	**Week**	Shipment Calendar Week
42	**Year/Month**	Shipment Calendar Week (with Year)
43	**Business Unit Division**	Customer Plant/Business Unit Identification
44	**Weekday of Pick Up**	Monday to Sunday Pick up Day Information
45	**Reference Pickup Day**	Customer authorization based on best day of pick up verification

Wrapper Support Vector Machine

Attribute Number	Attribute Name	Attribute Description
1	**City Lane**	Origin City - Destination City Route
2	**Country Lane**	Origin Country - Destination Country Route
3	**Origin**	Origin Airport
4	**Origin City**	Origin City Name
8	**Destination**	Destination Airport
9	**Destination City**	Destination City
10	**Ship Date**	Shipment Reference Date
11	**Priority Level**	Freight Forwarder Cargo Priority Level
12	**Service Type**	Type of Expedite Service Provided
13	**Shipper Account**	Supplier Identification Number
20	**Consignee Country**	Destination Customer Country
21	**Customer Number**	Customer Number Identification
22	**Export Carrier**	Airline Company
36	**Total Pieces**	Number of Shipment Handling Units
37	**Actual Weight (Kg)**	Cargo Weight
39	**Service Level Airfreight**	Type of Service (expedite or standard service)
40	**US/CN Zone Identification**	United States and China Region Specification
41	**Week**	Shipment Calendar Week
42	**Year/Month**	Shipment Calendar Week (with Year)
44	**Weekday of Pick Up**	Monday to Sunday Pick up Day Information
45	**Reference Pickup Day**	Customer authorization based on best day of pick up verification

APPENDIX 3 – ATTRIBUTE SELECTION ALGORITHMS RESULT MLP

CFS (Correlation-based Feature Subset Selection) (Adapted to MLP)

Attribute Number	Attribute Name	Attribute Description
9	**Destination City**	Destination City
14	**Shipper Name**	Supplier Name

Chi-Square Statistical Test (Adapted to MLP)

Attribute Number	Attribute Name	Attribute Description
2	**Country Lane**	Origin Country - Destination Country Route
7	**Origin Region**	Origin Macro Region
8	**Destination**	Destination Airport
9	**Destination City**	Destination City
12	**Service Type**	Type of Expedite Service Provided
39	**Service Level Airfreight**	Type of Service (expedite or standard service)
42	**Year/Month**	Shipment Calendar Week (with Year)
43	**Business Unit Division**	Customer Plant/Business Unit Identification
44	**Weekday of Pick Up**	Monday to Sunday Pick up Day Information
45	**Reference Pickup Day**	Customer authorization based on best day of pick up verification

Wrapper Artificial Neural Network (Multilayer Perceptron) - Original

Attribute Number	Attribute Name	Attribute Description
1	**City Lane**	Origin City - Destination City Route
3	**Origin**	Origin Airport
6	**Origin Country Reference**	Origin Consolidated Reference
7	**Origin Region**	Origin Macro Region
8	**Destination**	Destination Airport
9	**Destination City**	Destination City
10	**Ship Date**	Shipment Reference Date
11	**Priority Level**	Freight Forwarder Cargo Priority Level
12	**Service Type**	Type of Expedite Service Provided
13	**Shipper Account**	Supplier Identification Number
16	**Shipper State**	Supplier State
17	**Shipper Country**	Supplier Billing Country
19	**Consignee State**	Destination Customer State
20	**Consignee Country**	Destination Customer Country
36	**Total Pieces**	Number of Shipment Handling Units
38	**Charge Weight (Kg)**	Cargo Chargeable Weight
39	**Service Level Airfreight**	Type of Service (expedite or standard service)
40	**US/CN Zone Identification**	United States and China Region Specification
41	**Week**	Shipment Calendar Week
42	**Year/Month**	Shipment Calendar Week (with Year)
43	**Business Unit Division**	Customer Plant/Business Unit Identification
44	**Weekday of Pick Up**	Monday to Sunday Pick up Day Information
45	**Reference Pickup Day**	Customer authorization based on best day of pick up verification

Wrapper Artificial Neural Network (Multilayer Perceptron) - Adapted

Attribute Number	Attribute Name	Attribute Description
6	**Origin Country Reference**	Origin Consolidated Reference
7	**Origin Region**	Origin Macro Region
8	**Destination**	Destination Airport
9	**Destination City**	Destination City
11	**Priority Level**	Freight Forwarder Cargo Priority Level
12	**Service Type**	Type of Expedite Service Provided
19	**Consignee State**	Destination Customer State
38	**Charge Weight (Kg)**	Cargo Chargeable Weight
39	**Service Level Airfreight**	Type of Service (expedite or standard service)
42	**Year/Month**	Shipment Calendar Week (with Year)
43	**Business Unit Division**	Customer Plant/Business Unit Identification
44	**Weekday of Pick Up**	Monday to Sunday Pick up Day Information
45	**Reference Pickup Day**	Customer authorization based on best day of pick up verification

APPENDIX 4 – ATTRIBUTE SELECTION ALGORITHMS KNN

CFS (Correlation-based Feature Subset Selection)

Attribute Number	Attribute Name	Attribute Description
9	**Destination City**	Destination City
10	**Ship Date**	Shipment Reference Date
14	**Shipper Name**	Supplier Name

Chi-Square Statistical Test

Attribute Number	Attribute Name	Attribute Description
1	**City Lane**	Origin City - Destination City Route
2	**Country Lane**	Origin Country - Destination Country Route
3	**Origin**	Origin Airport
4	**Origin City**	Origin City Name
5	**Origin Country**	Origin Country
6	**Origin Country Reference**	Origin Consolidated Reference
7	**Origin Region**	Origin Macro Region
8	**Destination**	Destination Airport
9	**Destination City**	Destination City
10	**Ship Date**	Shipment Reference Date
11	**Priority Level**	Freight Forwarder Cargo Priority Level
12	**Service Type**	Type of Expedite Service Provided
13	**Shipper Account**	Supplier Identification Number
14	**Shipper Name**	Supplier Name
15	**Shipper City**	Supplier City
16	**Shipper State**	Supplier State
17	**Shipper Country**	Supplier Billing Country
18	**Consignee City**	Destination Customer City
19	**Consignee State**	Destination Customer State
20	**Consignee Country**	Destination Customer Country
21	**Customer Number**	Customer Number Identification
22	**Export Carrier**	Airline Company
23	**Carrier Code**	Airline Company Code Identification
24	**Freight Received**	Customer Shipment Authorization Date
25	**Docs From Shipper**	Customer Documents Availability Date
26	**Pickup**	Shipment Pick up Date
27	**ATD**	Actual Time of Departure Date
28	**ATA**	Actual Time of Arrival Date
29	**ETA**	Expected Time of Departure Date
30	**ETD**	Expected Time of Arrival Date
31	**Docs Received**	Destination Documents Hand Over Date
32	**Docs to Broker**	Broker Documents Availability Date
33	**POD Date**	Proof of Delivery Date
34	**Delivery Date**	Delivery Date
35	**Due Date**	Reference Contract Delivery Due Date
39	**Service Level Airfreight**	Type of Service (expedite or standard service)
40	**US/CN Zone Identification**	United States and China Region Specification
41	**Week**	Shipment Calendar Week
42	**Year/Month**	Shipment Calendar Week (with Year)
43	**Business Unit Division**	Customer Plant/Business Unit Identification
44	**Weekday of Pick Up**	Monday to Sunday Pick up Day Information
45	**Reference Pickup Day**	Customer authorization based on best day of pick up verification

Wrapper KNN

Number	Attribute Name	Attribute Description
1	**City Lane**	Origin City - Destination City Route
2	**Country Lane**	Origin Country - Destination Country Route
8	**Destination**	Destination Airport
9	**Destination City**	Destination City
11	**Priority Level**	Freight Forwarder Cargo Priority Level
12	**Service Type**	Type of Expedite Service Provided
20	**Consignee Country**	Destination Customer Country
25	**Docs From Shipper**	Customer Documents Availability Date
26	**Pickup**	Shipment Pick up Date
28	**ATA**	Actual Time of Arrival Date
32	**Docs to Broker**	Broker Documents Availability Date
33	**POD Date**	Proof of Delivery Date
35	**Due Date**	Reference Contract Delivery Due Date
39	**Service Level Airfreight**	Type of Service (expedite or standard service)
40	**US/CN Zone Identification**	United States and China Region Specification

APPENDIX 5 – PAPER INTERNATIONAL JOURNAL OF LOGISTICS SYSTEMS AND MANAGEMENT

DAYRELL MENDONÇA, G.; LIMA JUNIOR, O.F. (xxxx) 'Artificial intelligence applied to supply chain operations management: a systematic literature review', *Int. J. Logistics Systems and Management*, Vol. X, No. Y, pp.000-000 (Approved to publication)

Qualis Capes B2 – Engenharias I

Int. J. Logistics Systems and Management, Vol. X, No. Y, xxxx 1

Artificial intelligence applied to supply chain operations management: a systematic literature review

Guilherme Dayrell Mendonça* and
Orlando Fontes Lima Junior

Transportation Department,
Faculty of Civil Engineering, Architecture and Urbanism (FEC),
Campinas State University (UNICAMP),
P.O. Box 13083-852, Rua Albert Einstein,
951 – Cidade Universitária Zeferino Vaz – Campinas, São Paulo, Brazil
Email: g226791@dac.unicamp.br Email: oflimaj@fec.unicamp.br
*Corresponding author

Abstract: Artificial intelligence (AI) has been a key driver to reduce operational uncertainty and improve performance in supply chain management. Due to the advent of new data gathering technologies (IoT) and greater storage capacity, big data analytics (BDA) is rapidly growing as one of the main fields within AI research. We examined a representative sample of AI works applied to SCM from 2000 to 2020 and analysed them considering the main areas of the SCOR model framework of operations. The systematic literature review was based on a meta-synthesis methodology. The main research questions addressed were: 1) What are the main research methodologies used in AI SCM literature? 2) In what areas of SCM operations is AI (including BDA) mostly applied? 3) What are the most used AI models? The discussion addressing these three questions reveals a number of research gaps, which leads to future research directions.

Keywords: artificial intelligence; supply chain management; logistics; data mining; big data analytics; BDA; machine learning; supply chain operations reference; SCOR model.

Reference to this paper should be made as follows: Dayrell Mendonça, G. and Lima Junior, O.F. (xxxx) 'Artificial intelligence applied to supply chain operations management: a systematic literature review', *Int. J. Logistics Systems and Management*, Vol. X, No. Y, pp.000-000.

Biographical notes: Guilherme Dayrell Mendonça holds a BS in Industrial Engineering from Federal University of Minas Gerais and holds a graduate degree in Business Management from Dom Cabral Foundation. He is an MSc candidate in Transportation at the School of Civil Engineering, Architecture and Urban Planning of the State University of Campinas (UNICAMP) and graduate candidate at the MIT GCLOG (Graduate Certificate in Logistics) Program (Class 2021). His research interests are on transportation management, strategy and artificial intelligence applied to supply chain management. He has 17 years working experience on the logistics and supply chain management fields.

Orlando Fontes Lima Junior is a Professor of the Department of Geotechnics and Transportation at the Faculty of Civil Engineering, Architecture and Urbanism (FEC) at UNICAMP. He obtained his Doctorate and Master's in Transport Engineering from the University of São Paulo. He completed his Graduate degree in Naval Engineering at the University of São Paulo. He holds a Post-doctorate from the State University of Campinas and from Bournemouth University. His research interests are on logistics and supply chain management.

1 Introduction

Supply chain management frequently deals with operational instabilities caused by internal and external factors. Achieving good performance depends on the ability to foster information sharing and analysis between companies (Green, 2001). However, the full potential of data analytics is still to be achieved as companies progressively implement projects in this area. The analytics capability has a direct effect on supply chain agility and competitive advantage. Organisational flexibility also plays an important moderation role on the path unifying the agility and competitive dimensions (Dubey et al., 2019d). Srinivasan and Swink (2018) analysed data from 191 global firms which indicated that both demand and supply visibility are associated with the development of analytics capability. Similarly, analytics capability is shown to be more strongly associated with operational performance when supply chain organisations also possess organisational flexibility that is needed to quickly and efficiently act according to analytics-generated insights.

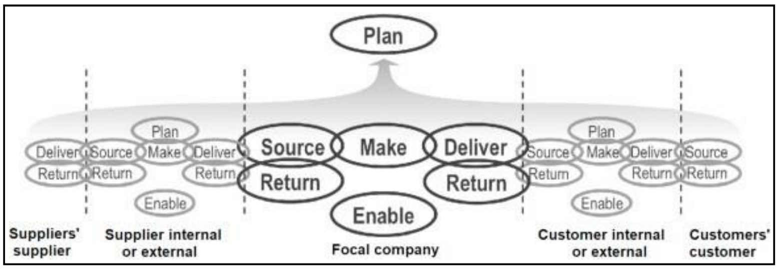
Research in artificial intelligence (AI) applied to the supply chain area is rapidly growing not only in supply and demand management but also in other application areas such as operations optimisation. The main objective is to increase performance and reduce operational uncertainty. AI, blockchain, cloud and data analytics technologies are driving the development of transformative business models with new platforms that automate processes, match demand and supply, dynamically define pricing and make real-time decisions (Akter et al., 2020). These methodologies have an intrinsic correlation with the enhanced use of information datasets to gain further insights into daily decision making. Hofmann and Rutschmann (2018) studied large retailers that implemented the technique of advanced shipment of products based on prescriptive models of demand forecasting. One of the main benefits achieved was to avoid the recurrence of inventory peaks of finished products through proactive mitigation actions.

The development of internet of things (IoT) devices combined with greater data storage capacity, expanded data analytics and AI applications led to the development of the big data analytics (BDA) research field in SCM. The term big data was first coined by Cox and Ellsworth (1997) in an article that indicated the eminent limitation of information storage on hardware resources caused by the exponential growth of information available in computer systems. In addition, BDA was defined as the application of advanced techniques of data mining, statistical analysis, predictive and prescriptive analysis of very large databases aimed at generating value to the organisational decision-making process (Tiwari et al., 2018). Mikalef et al. (2018) argued that the main source of competitive edge, especially in highly dynamic and turbulent environments, will result from the ability of companies to strengthen their organisational capabilities through the targeted use of big data and business analytics. In addition, to realise the value of BDA, it is necessary not only to put them into action by generating data-driven information for specific organisational capabilities, but also to take steps to harness the insights.

In recent years, several literature review papers have addressed the adoption of AI in supply chain management. In Guo et al. (2011), the systematic review identified the main trend of AI application in the apparel industry. Min (2015) reviewed the application of GA algorithms in supply chain modelling. Buyukozkan and Gocer (2018) investigated the state-of-the-art literature of digital supply chain not only in terms of big data but also new technologies such as cloud computing and robotics in SCM. Analyses have been carried out on BDA technology resources (Hu et al., 2014), development capabilities (Arunachalam et al., 2018), main research fields and data science techniques employed (Barbosa et al., 2018; Govidan et al., 2018; Nguyen et al., 2018; Wang et al., 2016), combined with IoT (Aryal et. al., 2018), application in manufacturing (O'Donovan et al., 2015) related to SCM framework of value creation (Brinch, 2018; Wamba et al., 2015), clusters of practical (Mishra et al., 2018), process-oriented (Chehbi-Gamoura et al., 2020) and operational (Choi et al., 2018; Lamba and Singh, 2017) applications.

While these studies have been able to provide insight into the field through structured reviews and classification into future research themes, there is further potential to investigate the main SC operational areas. This paper considered the AI literature review regarding the six supply chain operations reference (SCOR®) management fields, namely: plan, source, make, deliver, return and enable. As illustrated in Figure 1, the business processes proposed by the SCOR® model embrace various tiers along the supply chain and include a set of management practices recognised by companies in many industries (Didehkhani et al., 2009; Lima-Junior and Carpinetti, 2016).

Figure 1 Major management processes proposed by the SCOR® model



Source: SCC (2012)

The plan processes aggregate demand and supply to develop a course of action which best meets sourcing, production, and delivery requirements. Source processes procure goods and services to meet planned or actual demand. Make processes transform product to a finished state to meet planned or actual demand. Deliver processes provide finished goods and services to meet planned or actual demand, typically including order management, transportation management, and distribution management. Return processes are associated with returning or receiving returned products for any reason. These processes extend into post-delivery customer support. Finally, enable processes are associated with the management of the supply chain. These processes include management of: business rules, performance, data, resources, facilities, contracts, supply chain network management, managing regulatory compliance and risk management (SCC, 2012). Specifically, this paper examined a representative sample of works using a systematic literature review based on meta-synthesis methodology in order to address three research questions:

a What are the main research methodologies used in AI SCM literature?

b In what areas of SCM operations is AI (including BDA) mostly applied?

c What are the most used AI models?

The discussion addressing these three questions reveals a number of research gaps, which leads to future research directions. This study contributes to the SCM literature by identifying how supply chain operation processes through the lens of the SCOR model were developed using AI techniques (not only BDA), and which models were used to support value achievement. It also contributes to practitioners who may find potential benefits in such process investigation as it provides an integrated scope of analysis considering the entire supply chain.

The paper is organised into six sections. Following this introduction, Section 2 presents an introductory concept review on AI and BDA applied to SCM. Section 3 describes the methodological approach utilised. Section 4 details the descriptive statistics of recent publications. Finally, Section 5 presents content analysis and Section 6 shows the conclusions and future research recommendations.

2 Literature review

According to Russell and Norvig (1995), AI is known for its ability to think like humans, act like humans, think rationally, and acts rationally. Thus, with respect to these distinctive features, AI can also be classified into a number of sub-fields:

1 artificial neural networks (ANN) and rough set theory (‘thinking humanly’)

2 machine learning, expert systems, and genetic algorithms (‘acting humanly’)

3 fuzzy logic (‘thinking rationally’)

4 agent-based systems (‘acting rationally’).

Min (2010) defines the first group as the development of computational capacity based on interconnected memory systems that are able to learn from experience, recognise patterns, group objects and process ambiguous and abstract information. In the logistics field, the ANN can be used in autonomous vehicles and self-driving-cars, applied in problems of lot-sizing and machine set-up time. These solutions have proven to outperform traditional optimisation algorithms in the field of operational research. The second category consists of machine learning, expert systems and genetic algorithms. The machine learning functionality is primarily intended to enable computers to learn without necessarily being programmed for such activity. Its application has already been used to predict collaborative behaviour in supply chain management. Conversely, expert systems are programs which emulate problems through logical reasoning based on human knowledge and have already been used in traffic control, maintenance scheduling and other situations. Genetic algorithms simulate the natural evolutionary process generating new organisms for problem solving and have already been applied to problems in the design of transport networks. Thirdly, fuzzy logic system based on the input of experts, which defines the quality of a particular solution, can be used to support the process of choosing suppliers, inventory level control, among others. Finally, the main characteristic of agent-based systems is the division of problems into subgroups to be treated by a specific computational agent, and has been used in several processes of the supply chain planning. Tripathi and Gupta (2020) emphasised the important role AI plays in network planning, procurement, and consumer interaction. According to these authors, AI is a key enabler in the framework proposed for transforming supply chains to smarter systems.

AI has evolved considerably in several areas of theoretical and practical development of the supply chain. A study conducted by DHL in conjunction with IBM identified the latest applications of AI in industry best practices (DHL, 2018). Applications include end user support solutions, voice interaction solutions with end consumers, machine learning applied to social networks, creation of expert content, identification of information standards, robots in the retail operation, autonomous vehicles, robots in manufacturing and predictive management of demand. Specifically, industrial and logistics AI applications in conjunction with technologies such as cloud enterprise resource planning have a positive correlation to resolve high uncertainties and gain more operational competitive advantages than other competitors in the dynamically changing market (Gupta et al., 2019).

The adoption of such tools presupposes a more advanced stage of technological development, an entrepreneurial culture of investment and an adaptive thinking environment to cope with these transformative innovations (Rampersad, 2020; Dubey et al., 2019a). The research by Dubey et al. (2019b) confirmed the importance entrepreneurial orientation has in allowing companies to sense dynamic market changes and enhance their performance by improving their decision-making ability utilising BDA-AI. In general, a company’s ability to promote innovation in logistics is positively correlated with the generation of competitive advantage in the market in which it operates (Grawe, 2009).

Albergaria and Jabbour (2020) suggest the vital importance of adopting BDA capabilities to deal with large amounts of real-world data in order to understand the complexities of the sharing economy. According to Tiwari et al. (2018), the biggest challenge for supply chain professionals today is finding the best way to deal with the growing availability of large information bases. Among the possible benefits of using this new BDA tool is the construction of agile operations with greater capacity to monitor events, thus increasing the possibility of performing adaptive actions (Dubey et al., 2019a). Wamba et al. (2020) suggest that BDA has positive effects on improving supply chain agility, supply chain adaptability and performance measures (cost and operational performance).

Bowers et al. (2017) argued that for a company to benefit from the practice of supply chain BDA, it is necessary to reduce the reaction time after receiving the information. As an example, Hanesbrands Inc., a US capital goods company, decided to adjust its machine learning algorithms to better react when a supplier rescheduling occurs. Additionally, other possible reactions range from adjusting the freight rate to re-sequencing production schedules to avoid a line stop. However, these cases are exceptions, since according to the authors; few companies are able to transform the high availability of information into competitive advantage and value for the end customer. The study by Kamble and Gunasekaran (2020) on big data-driven supply chain (BDDSC) proposes a new framework to better measure operational performance in real-time, with proactive decision-making regarding shortcomings and better overall target and achieving added value.

Within this context of value creation, according to Brinch (2018), the use of big data tools offers three possible dimensions of analysis:

1 Discovery value that describes the company's ability to structure a reliable database.

2 Creation value which represents the capacity to transform information into a source of decision-making.

3 Capture value in which the company achieves improvement in the operational or financial results through the use of big data.

Based on these concepts an evaluation model can be created to see how the value is being managed in a company's BDA process. Complementary to the above mentioned, important elements for gaining business value from big data investments include recruiting people with good technical and managerial understanding of big data and analytics, fostering a culture of organisational learning, and embedding big‐data decision‐making into the organisation's structure. Hence, it is the combined effect of these resources that will enable a firm to develop a BDA and achieve value gains (Mikalef et al., 2019).

In addition to companies, BDA can also be used for operational decisions related to the humanitarian supply chain, increasing coordination and integration by providing greater visibility of each agent's capabilities in the temporary flow of supplies (Dubey et al., 2018; Papadopoulos et al., 2017). Another possible positive impact advocated by Hazen et al. (2018), is the possibility to transform the supply chain into a more sustainable organisation by broadening the field of analysis for environmental and social issues of the process. Likewise, in Dubey et al. (2019c), the empirical results indicate that BDA offers significant benefits to both social and environmental related initiatives and performance in supply chains.

Different sources can be used to build the database, such as IoT products or machines. Utilising these sources has proven to increase the companies' ability to measure operational performance since it allows real-time analysis along the supply chain (Dweekat et al., 2017). For example, IoT applied to cargo vehicles is becoming a key source of information on drivers' conduct and their relationship with fuel consumption and vehicle depreciation (Hopkins and Hawking, 2018). Yerpude and Singhal (2020) studied how IoT supported the smart supply chain management. According to the authors, the data provided by these devices will play a crucial role in the supply chain management for online retail growth. Businesses will increase future growth based on automatically generated data from the IoT, which will help them become a much more agile and competitive supply chain.

In summary, AI can create value in areas such as consumer behaviour, supply chain visibility and transparency, operational and maintenance efficiency, information management, responsiveness, and the generation of new business opportunities based on market trends. In Zhan et al. (2018) the findings reveal that big data can offer customer involvement so as to provide valuable input for developing new products, hence smaller market risks. Conversely, its development limitations are usually related to IT infrastructure, human resources and knowledge, and openness to information exchange in the supply chain (Kache and Seuring, 2017). Urciuoli and Hintsa (2018) argue the importance of data to support risk management initiatives and any information sharing barrier may jeopardise all benefits of information sharing between companies.

3 Methodology

In the supply chain literature, several applications of AI (including big data) are described and validated in different operational, tactic and strategic levels of management. However, there is still a shortage of studies that summarise all these initiatives into a single literature review framework of analysis combining the main AI research methodologies, supply chain areas of application and main analytical models used. In the present paper the main research objective is to fill this gap and provide researchers and practitioners a broader overview of AI in supply chain management in line with the SCOR model. The model's main management areas (plan, source, make, deliver, return, enable) will be used as a conceptual reference for assessing and classifying the papers according to their content and area of impact. As a result, a broader overview of the evolution of the field in the last two decades will be provided considering a technical cross-referenced literature review process with descriptive findings.

The search methodology was a systematic literature review and meta-synthesis. For Kamal and Irani (2014) the objective of a systematic and structured literature review is to observe and understand the past trends and existing patterns/themes in the research area, evaluate contributions and summarise knowledge, thereby identifying limitations, implications and potential directions of further research. The systematic literature review approach ensures it is auditable and repeatable, so that this method overcomes the perceived weaknesses of a narrative review (Wong et al., 2012a).

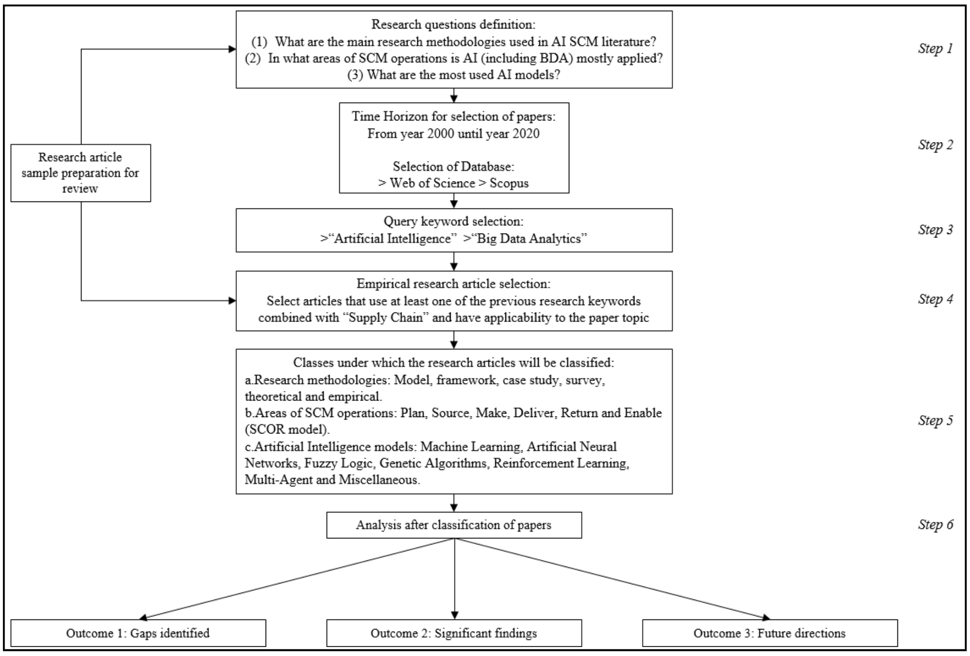
This approach covers the following research steps: step 1 – define the research question, providing the drivers for the literature review; step 2 – set the search strategy – define the databases and search period; step 3 – define the inclusion or exclusion criteria – choose the appropriate keywords for selecting the papers; step 4 – search the articles – select the first group of papers according to the strategy (step 2) and based on inclusion/exclusion criteria (step 3); step 5 – analyse the papers – in-depth review of the papers selected in step 4 and categorisation of research fields; step 6 – conclusion and observation of findings (Soni and Kodali, 2011).

A schematic representation of literature review methodology adopted in the paper is shown in Figure 2 and described below:

- Step 1: based on the research objectives, the main research questions to be answered were:
  - a What are the main research methodologies used in AI SCM literature?
  - b In what areas of SCM operations is AI (including BDA) mostly applied?
  - c What are the most used AI models?
- Step 2: time horizon ranging from year 2000 until year 2020; Scopus and Web of Science were the databases chosen.
- Step 3: the following keywords were used as query inputs: 'AI', 'BDA'.

- Step 4: paper selection based on articles titles/abstracts, which use one or more keywords defined in step 3 combined with addition keyword 'supply chain', considering the time horizon in step 2 and research applicability to the article topic.
- Step 5: papers were classified under the following classes:
  a Research methodologies: model, framework, case study, survey, theoretical and empirical.
  b Areas of SCM operations: plan, source, make, deliver, return and enable (SCOR model).
  c AI models: machine learning, ANN, fuzzy logic, genetic algorithms, reinforcement learning, multi-agent and miscellaneous.
- Step 6: the objective of this paper is dominantly descriptive in nature. Thus, it is not suitable for applying statistical methodologies in deductions or for any inferential purpose using hypothesis testing. In this step, all efforts are directed towards critically analysing the classified articles so as to identify research gaps in AI content in SCM as well as to present significant findings from the existing literature.

Figure 2 Schematic representation of literature review methodology



A total of 542 papers were identified in Scopus and 412 in Web of Science databases (step 4). All abstracts were reviewed to ensure suitability to the research objectives in the present paper. The final dataset was composed of 144 articles. As the main literature review papers were discussed in the Introduction section, they were not considered for the purpose of descriptive and analytical statistics. The detailed paper classification and analysis are shown in the next section of the paper (step 5). Research gaps will be provided in the conclusion section (step 6).

4 Analysis

The findings are presented in two distinct areas of analysis:

1 time distribution

2 main journals.

By means of statistical analysis, we considered AI papers all works that discussed and applied AI in SCM not using big data. BDA papers are those that actually investigated AI in SCM within the context of big data application. The purpose of this classification is to include in the literature review process papers that either focuses on AI or BDA. We believe that an integrated analytical review of both AI and BDA papers could shed more precise and broader light on the evolution of AI within SCM as a whole.

4.1 Time distribution

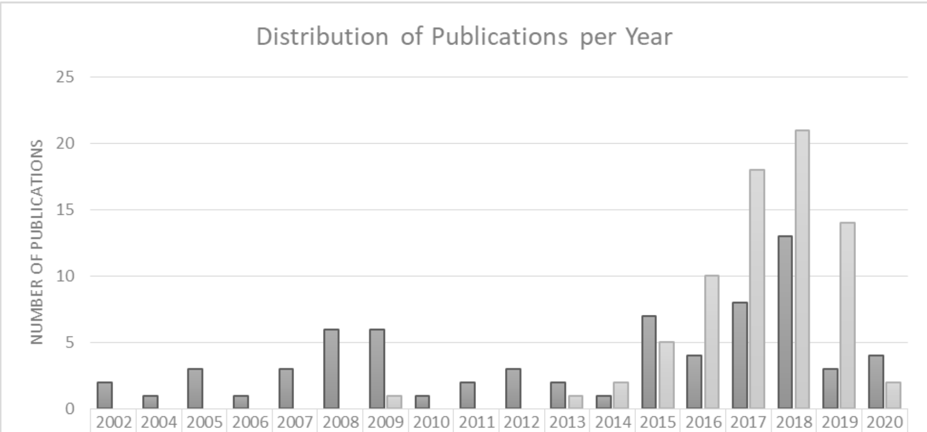
The selected articles range from the years 2002 until 2020. This time frame is divided into two different periods of distinct research trends:

1 predominance of AI papers, which accounts for approximately 20 % of published material (2002–2013)

2 rampant growth of BDA publications which outnumbered AI publications from 2014 until the present time.

Figure 3 represents the evolution of the number of publications over the years. It suggests that the application of BDA in the SCM area is a fast-growing and fruitful research field.

Figure 3 Amount of publications per year (AI and BDA)



	2002	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
AI	2	1	3	1	3	6	6	1	2	3	2	1	7	4	8	13	3	4
BDA							1				1	2	5	10	18	21	14	2

Table 1 Major journals based on number of publications and year (AI and BDA)

Main journals	*2*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*Total*
Expert Systems with Applications	0	0	2	0	0	0	2	0	0	0	0	0	0	0	2	1	0	*0*	7
International Journal of Production Research	1	0	0	0	0	0	0	0	0	0	0	0	1	0	3	1	1	*0*	7
International Journal of Logistics Systems and Management	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*1*	*2*	*1*	*0*	*2*	*6*
Computers and Industrial Engineering	0	0	0	0	0	1	0	0	0	0	0	0	1	3	0	0	1	*0*	6
Journal of Cleaner Production	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	3	*0*	6
International Journal of Information Management	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	1	5
Transportation Research Part E: Logistics and Transportation Review	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	1	*0*	4
International Journal of Physical Distribution and Logistics Management	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	*0*	4
Journal of Business Research	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	1	4
International Journal of Production Economics	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	4
Journal of Business Logistics	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	3
IFAC-PapersOnLine	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	3
European Journal of Operational Research	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	3
International Journal of Logistics Management	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	3
Production and Operations Management	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	*0*	3
Annals of Operations Research	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	*0*	3
Decision Support Systems	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2
Production Planning and Control	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	*0*	2
Neural Computing and Applications	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	*0*	2
The International Journal of Advanced Manufacturing Technology	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	*0*	2
Journal of Applied Sciences	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	*0*	2
Benchmarking	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	*0*	2
Jisuanji Jicheng Zhizao Xitong/Computer Integrated Manufacturing Systems, CIMS	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	*0*	2
Total	2	1	3	0	1	5	3	0	0	0	1	2	4	8	20	20	9	5	84

4.2 Journals

The selected 144 papers are from 77 different journals, of which only 23 published more than one paper. Table 1 illustrates the distribution of the reference papers in these 23 journals. AI research has attracted real interest from highly regarded academics as most of these papers have been published by journals with high impact factors in recent years.

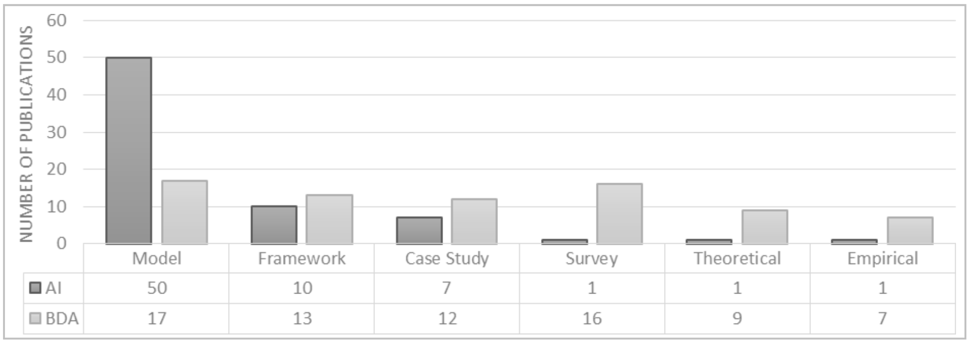
5 Results and discussion

5.1 What is the main research methodologies used in AI SCM literature?

In order to answer the first research question, a methodological classification was applied to all selected articles based on six different categories: model application, framework development, case study, survey, theoretical discussion and empirical investigation.

Figure 4 shows that model development tops the ranking with approximately half of total publications (46 %). That is, most of the literature discussion focuses on developing mathematical approaches and applications in supply chain research problems. The main areas of concentration of the model papers are supply chain planning and operational optimisation. The former group consists of papers concentrated on demand forecasting (Carbonneau et al., 2012), inventory management (De Santis et al., 2017), bullwhip effect mitigation (Aggarwal and Dave, 2018) and predictive and adaptive management approach for Omnichannel retailing supply chains (Pereira et al., 2018). However, operational optimisation is represented mostly by studies that focus on improving transportation routing decisions (Mokhtarinejad et al., 2015), lead time shop floor prediction (Gyulai et al., 2018) and dynamic lot-sizing (Wong et al., 2012b).

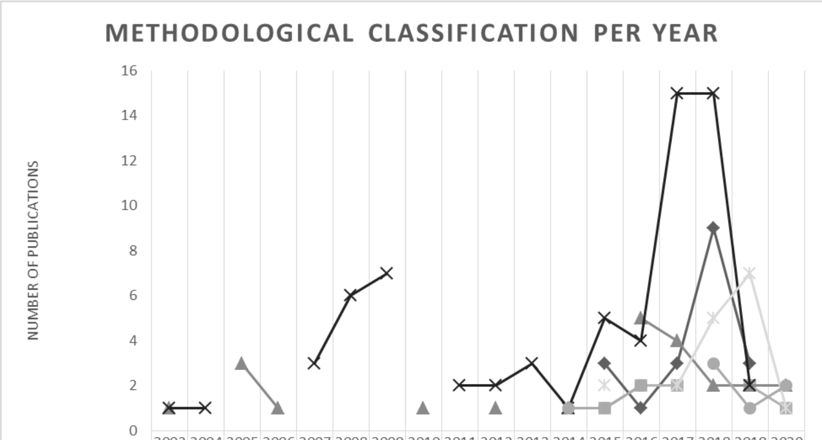
Figure 4 Methodological classification



The second group of publication concentrates on case studies and framework development, which could be an indication of increasing practical application of developed models and concepts. Similarly to the model papers, case studies are largely applied to supply chain planning (Andersson and Jonsson, 2018) and operational optimisation areas (Borade and Sweeney, 2015). Conversely, framework papers are mainly focused on conceptual development regarding data driven supply chains (Chavez et al., 2017), sustainable manufacturing (Dubey et al., 2016), supply chain risk analytics (Ivanov et al., 2019) and disaster resilience (Papadopoulos et al., 2017).

Figure 5 shows the evolution of methodological approaches over the years. Since 2016, there has been a clear increase of papers dedicated to practical investigation such as surveys, case studies and model development. Finally, theoretical development and empirical investigation are the lowest ranked approaches.

Figure 5 Methodological classification over the years.



	2002	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Case Study													3	1	3	9	3	
Empirical													1	2	2		2	1
Framework	1		3	1				1		1		1		5	4	2	2	2
Model	1	1			3	6	7		2	2	3	1	5	4	15	15	2	
Survey													2		2	5	7	1
Theoretical												1	1	2		3	1	2

5.2 In what areas of SCM operations is AI (including BDA) mostly applied?

Tables 2 and 3 illustrate, respectively, all AI and BDA papers divided by research methodology and area of classification regarding the SCOR model, namely: enable, source, make, deliver, plan and return. While AI papers are mostly concentrated on model application, BDA research spans over other areas such as survey, empirical and theoretical investigation and is more equally distributed. This may well show that BDA is still developing its conceptual and practical background as a recent area of study in supply chain.

Regarding SCOR classification, plan and enable areas are the most representative. Enable ranks first in BDA papers (Table 3) and second for AI publications (Table 2). Conversely, plan ranks first in AI papers (Table 2) and second in BDA only studies (Table 3). Overall, the SCOR plan area is the most representative if we consider both BDA and AI papers.

The plan category consists of papers essentially focused on further understanding and applying AI and BDA on demand management. The main plan papers in Table 2 highlight: predictive approaches (Pereira et al., 2018), Bullwhip effect mitigation (Mojaveri et al., 2009; Singh and Challa, 2016), demand forecasting improvement (Amirkolaii et al., 2017; Carbonneau et al., 2008; Efendigil et al., 2009), development of vendor management inventory technique (Chi et al., 2007) and inventory management optimisation (Giannoccaro and Pontrandolfo, 2002). Furthermore, the majority of BDA papers in Table 3 are related to demand forecasting (Hofmann and Rutschmann, 2018; Lau et al., 2018; Lee, 2017; Nita, 2015; Yu et al., 2019).

Table 2 Most relevant research fields (AI) regarding SCOR model categories

Research method/authors	*Enable*	*Source*	*Make*	*Deliver*	*Plan*	*Return*
Case study	*5*	*2*	*1*	*3*	*4*	*0*
Borade and Sweeney (2015)	0	1	0	1	1	0
Ma et al. (2018)	1	0	1	0	0	0
Mahroof (2019)	1	0	0	1	1	0
Orji and Wei (2015)	1	1	0	0	0	0
Slimani (2017)	0	0	0	1	1	0
Tsang et al. (2018)	1	0	0	0	1	0
Urciuoli and Hintsa (2018)	1	0	0	0	0	0
Framework	*7*	*2*	*1*	*0*	*6*	*1*
Garg and Viswanadham (2010)	1	0	0	0	0	1
Kartal et al. (2016)	0	1	0	0	1	0
Piramuthu and Sikora (2005)	0	0	0	0	1	0
Piramuthu (2005a)	1	1	0	0	1	0
Piramuthu (2005b)	1	0	0	0	0	0
Pontrandolfo et al. (2002)	0	0	1	0	1	0
Rampersad (2020)	1	0	0	0	0	0
Siurdyban and Moller (2012)	1	0	0	0	1	0
Tripathi and Gupta (2020)	1	0	0	0	0	0
Xu et al. (2006)	1	0	0	0	1	0
Model	*26*	*16*	*6*	*22*	*42*	*0*
Aggarwal and Dave (2018)	0	0	0	0	1	0
Amirkolaii et al. (2017)	0	0	0	1	1	0
Carbonneau et al. (2007)	0	0	0	1	1	0
Carbonneau et al. (2008)	0	0	0	1	1	0
Carbonneau et al. (2012)	0	0	0	1	1	0
Castillo-Villar and Herbert-Acero (2013)	1	0	1	1	1	0
Cavalcante et al. (2019)	1	1	0	0	1	0
Chaharsooghi et al. (2008)	0	0	0	0	1	0
Chen and Xu (2018)	1	1	0	0	0	0
Chi et al. (2007)	0	1	0	1	1	0
Cui et al. (2018)	1	0	0	1	1	0
Curcio et al. (2007)	0	0	0	1	1	0
De Santis et al. (2017)	0	1	0	0	1	0
Efendigil et al. (2009)	0	0	0	1	1	0
Fu and Fu (2015)	0	1	0	0	1	0
Giannakis and Louis (2011)	1	1	1	1	1	0

Table 2 Most relevant research fields (AI) regarding SCOR model categories (continued)

Research method/authors	*Enable*	*Source*	*Make*	*Deliver*	*Plan*	*Return*
Model	*26*	*16*	*6*	*22*	*42*	*0*
Giannoccaro and Pontrandolfo (2002)	0	1	1	1	1	0
Guosheng and Guohong (2008)	1	1	0	0	0	0
Gyulai et al. (2018)	0	0	1	0	1	0
Hiromoto et al. (2017)	1	0	0	0	0	0
Hong and Ha (2008)	1	1	0	0	1	0
Ilie-Zudor et al. (2015)	1	0	0	1	1	0
Jafarzadeh-Ghoushchi and Rahman (2016)	1	0	0	1	1	0
Kar (2015)	1	1	0	0	0	0
Kazemi and Fazel Zarandi (2008)	1	0	1	0	1	0
Kiekintveld et al. (2009)	0	0	0	0	1	0
Kong and Li (2018)	0	0	0	0	1	0
Kumar et al. (2013)	0	1	0	0	1	0
Mojaveri et al. (2009)	0	0	0	0	1	0
Mokhtarinejad et al. (2015)	0	0	0	1	1	0
Moraga et al. (2011)	1	0	0	0	1	0
Park et al. (2018)	1	0	0	0	1	0
Pereira et al. (2018)	1	1	0	1	1	0
Raut et al. (2017)	1	1	0	0	1	0
Shahrabi et al. (2009)	0	0	0	1	1	0
Shokouhyar et al. (2019)	1	0	0	0	0	0
Singh and Challa (2016)	1	0	0	1	1	0
Slimani et al. (2015)	0	0	0	0	1	0
Sun et al. (2008)	0	0	0	1	1	0
Tse et al. (2009)	1	0	0	0	1	0
Vahdani et al. (2014)	0	0	0	0	1	0
Valluri et al. (2009)	1	0	0	0	1	0
Wanke et al. (2017)	0	0	0	1	1	0
Wieczorek and Ignaciuk (2018)	1	1	0	1	1	0
Wong et al. (2012b)	0	0	1	0	1	0
Wu et al. (2017b)	1	0	0	1	1	0
Yuen et al. (2018)	1	0	0	1	1	0
Zhang et al. (2004)	1	1	0	0	0	0
Zhang et al. (2016)	1	1	0	0	0	0
Zhu et al. (2017)	1	0	0	0	0	0
Total of papers	*38*	*20*	*8*	*25*	*52*	*1*

In contrast to the previous SCOR category, BDA is more prominent in the enable research field. Enable papers in Table 3 are mostly surveys on service supply chains and on the development of capabilities such as agility and preparedness (Fernando et al., 2018; Mandal, 2018; Roßmann et al., 2018), risk management (Engelseth and Wang, 2018; Ivanov et al., 2019; Mani et al., 2017; Zhao et al., 2017; Wu et al., 2017a) and theoretical construction aimed at future applications (Hazen et al., 2014, 2016; Singh and El-Kassar, 2019; Smyth et al., 2018; Zhong et al., 2016). AI enable papers in Table 2 have a distinct framework focused on development of business process design (Piramuthu, 2005a, 2005b; Siurdyban and Møller, 2012; Xu et al., 2006). As main technological enabler, IoT applications in supply chain management are also highlighted in Table 2 (Ehret and Wirtz, 2017; Hiromoto et al., 2017; Ma et al., 2018; Tsang et al., 2018; Yuen et al., 2018).

Table 3 Most relevant research fields (BDA) regarding SCOR model categories

Research method/author	*Enable*	*Source*	*Make*	*Deliver*	*Plan*	*Return*
Case study	*6*	*3*	*0*	*8*	*9*	*1*
Andersson and Jonsson (2018)	0	0	0	1	1	0
Boldt et al. (2016)	0	0	0	1	1	0
Engelseth and Wang (2018)	1	1	0	0	1	0
Gravili et al. (2018)	0	0	0	0	1	0
Hopkins and Hawking (2018)	1	0	0	1	0	1
Matthias et al. (2017)	1	1	0	1	1	0
Moktadir et al. (2019)	1	0	0	0	0	0
Moretto et al. (2017)	1	1	0	0	0	0
Nita (2015)	0	0	0	1	1	0
Singh et al. (2018)	0	0	0	1	1	0
Yu et al. (2019)	0	0	0	1	1	0
Zhan et al. (2018)	1	0	0	1	1	0
Empirical	*7*	*2*	*2*	*4*	*4*	*0*
Ittmann (2015)	1	1	0	1	1	0
Niu et al. (2019)	1	0	0	1	1	0
Richey et al. (2016)	1	0	0	0	0	0
Sanders (2016)	1	1	1	1	1	0
Sodero et al. (2019)	1	0	0	0	0	0
Tsao (2017)	1	0	0	1	1	0
Zhong et al. (2015)	1	0	1	0	0	0
Framework	*9*	*4*	*2*	*4*	*9*	*5*
Arya et al. (2017)	0	1	0	1	1	0
Chavez et al. (2017)	0	1	0	0	1	0
Cheng and Lau (2016)	1	0	0	0	1	0
Dubey et al. (2016)	0	1	0	0	0	1
Hu et al. (2014)	1	0	0	0	0	0
Ivanov et al. (2019)	1	0	0	0	0	0
Jeble et al. (2018)	1	0	0	1	1	0

Table 3 Most relevant research fields (BDA) regarding SCOR model categories (continued)

Research method/author	*Enable*	*Source*	*Make*	*Deliver*	*Plan*	*Return*
Framework	*9*	*4*	*2*	*4*	*9*	*5*
Papadopoulos et al. (2017)	1	0	0	1	1	1
Rehman et al. (2016)	1	0	0	0	1	0
Ren et al. (2019)	1	0	1	0	1	1
Rodriguez and Da Cunha (2018)	1	0	0	0	0	1
Shukla and Tiwari (2017)	0	1	1	1	1	1
Wang et al. (2016)	1	0	0	0	1	0
Model	*12*	*5*	*3*	*6*	*12*	*5*
Choi (2018)	1	0	0	1	1	0
Côrte-Real et al. (2017)	0	1	0	0	1	1
Ehret and Wirtz (2017)	1	0	1	0	1	0
Giannakis and Louis (2016)	1	1	1	1	1	0
Hofmann (2017)	1	1	0	0	1	0
Jiang and Sheng (2009)	1	0	0	1	1	0
Kaur and Singh (2018)	0	1	0	0	1	1
Lau et al. (2018)	0	0	0	1	1	0
Lee (2017)	1	0	0	1	1	0
Prasad et al. (2018)	1	0	0	0	0	1
Simchi-Levi and Wu (2018)	1	0	0	1	1	0
Bumblauskas et al. (2017)	0	0	1	0	0	0
Waller and Fawcett (2013)	1	0	0	0	0	0
Wamba et al. (2017)	0	0	0	0	1	0
Wu et al. (2017a)	1	0	0	0	1	1
Wu and Lin (2018)	1	0	0	0	0	0
Zhao et al. (2017)	1	1	0	0	0	1
Survey	*15*	*0*	*1*	*0*	*6*	*2*
Chen et al. (2015)	0	0	0	0	1	0
Dubey et al. (2019a)	1	0	0	0	0	0
Dubey et al. (2019b)	1	0	0	0	0	0
Dubey et al. (2019c)	1	0	0	0	0	1
Dubey et al. (2019d)	1	0	0	0	0	0
Fernando et al. (2018)	1	0	0	0	1	0
Gunasekaran et al. (2017)	1	0	0	0	1	0
Gupta et al. (2019)	1	0	0	0	0	0
Lai et al. (2018)	1	0	0	0	0	0
Mandal (2018)	1	0	0	0	1	0
Mandal (2019)	1	0	0	0	1	0
Mani et al. (2017)	1	0	1	0	1	0

Table 3 Most relevant research fields (BDA) regarding SCOR model categories (continued)

Research method/author	*Enable*	*Source*	*Make*	*Deliver*	*Plan*	*Return*
Survey	*15*	*0*	*1*	*0*	*6*	*2*
Mikalef et al. (2019)	1	0	0	0	0	0
Raut et al. (2019)	1	0	0	0	0	1
Roßmann et al. (2018)	1	0	0	0	0	0
Schoenherr and Speier-Pero (2015)	1	0	0	0	0	0
Wamba et al. (2020)	1	0	0	0	0	0
Theoretical	*6*	*0*	*1*	*1*	*2*	*2*
Albergaria and Jabbour (2020)	1	0	0	0	0	0
Hazen et al. (2014)	1	0	1	0	0	0
Hazen et al. (2016)	1	0	0	0	0	0
Hazen et al. (2018)	0	0	0	0	1	0
Hofmann and Rutschmann (2018)	0	0	0	1	1	0
Li et al. (2015)	0	0	0	0	0	1
Singh and El-Kassar (2019)	1	0	0	0	0	1
Smyth et al. (2018)	1	0	0	0	0	0
Zhong et al. (2016)	1	0	0	0	0	0
Total	*55*	*14*	*9*	*22*	*41*	*15*

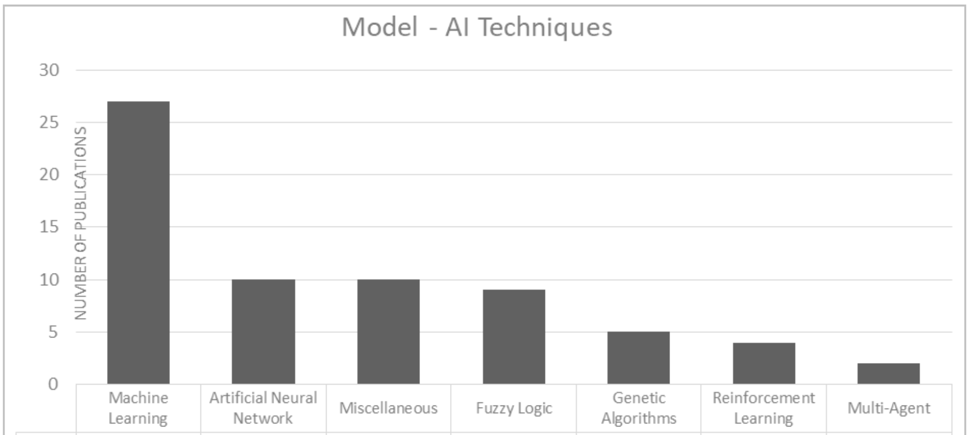
In third place, deliver SCOR category in Table 3 consist of papers that cover topics such as analytics to improve distribution practices (Singh et al., 2018; Simchi-Levi and Wu, 2018) and operations management (Giannakis and Louis, 2016) using BDA methodologies. Conversely, Deliver papers in Table 2 are more focused on route and resource optimisation (Curcio et al., 2007; Mokhtarinejad et al., 2015; Wieczorek and Ignaciuk, 2018), finished goods warehouse management (Mahroof, 2019), inventory allocation (Wanke et al., 2017) and network design (Ilie-Zudor et al., 2015).

Finally, source and return SCOR areas were the least explored, indicating further potential for future research. Source papers concentrated mainly on optimising supplier selection (Cavalcante et al., 2019; Chen and Xu, 2018; Guosheng and Guohong, 2008; Kar, 2015; Moretto et al., 2017; Orji and Wei, 2015; Raut et al., 2017; Zhang et al., 2004, 2016). Selected return papers consisted of works focused on supply chain sustainability (Garg and Viswanadham, 2010; Hopkins and Hawking, 2018; Kaur and Singh, 2018; Raut et al., 2019; Ren et al., 2019; Rodriguez and Da Cunha, 2018; Shukla and Tiwari, 2017; Singh and El-Kassar, 2019).

5.3 What are the most used AI models?

Model methodology accounted for 67 papers out of a total 144 selected for literature review in this paper (Figure 4). Within this category, Figure 6 shows that considerable research has been developed to apply machine learning (27 papers) and ANN (ten papers) in supply chain management. Miscellaneous category (ten papers), which comprise multiple algorithm application, fuzzy logic (nine papers), genetic algorithms (five papers), reinforcement learning (four papers) and multi-agent (two papers) were also used as AI investigation techniques.

Figure 6 AI models – techniques

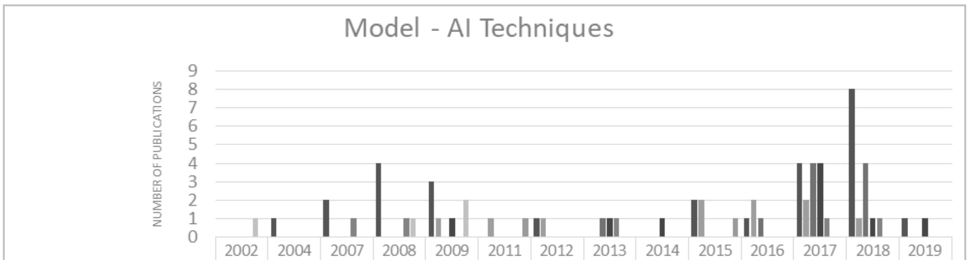


	Machine Learning	Artificial Neural Network	Miscellaneous	Fuzzy Logic	Genetic Algorithms	Reinforcement Learning	Multi-Agent
Total	27	10	10	9	5	4	2

Machine learning papers are more concentrated in supply chain planning, operational optimisation and supplier selection areas. Supervised learning methods such as support vector machines (Guosheng and Guoghong, 2008) are the most employed. ANN are growing also as main analytical tool with applications in transportation (Jafarzadeh-Ghoushchi and Rahman, 2016), supplier selection (Kar, 2015) and monitoring of supply chain behaviour (Moraga et al., 2011).

Figure 7 shows a growth trend of machine learning and miscellaneous algorithm application mainly from 2015 to 2018. The potential for applying other techniques such as reinforcement learning is yet to be explored. Papers on reinforcement learning are mostly concentrated on operational optimisation topics such as dynamic inventory control (Jiang and Sheng, 2009) and general supply chain management (Valluri et al., 2009).

Figure 7 AI models – techniques over the years



	2002	2004	2007	2008	2009	2011	2012	2013	2014	2015	2016	2017	2018	2019
Machine Learning		1	2	4	3		1			2	1	4	8	1
Artificial Neural Network					1	1	1			2	2	2	1	
Miscellaneous								1			1	4	4	
Fuzzy Logic					1			1	1			4	1	1
Genetic Algorithms			1	1				1				1	1	
Reinforcement Learning	1			1	2									
Multi-Agent						1				1				

6 Conclusions, scope for future work and limitations

Research on AI in supply chain management has significantly evolved in recent years. Many initiatives have been conducted to test and provide empirical evidence, in which the predictive approach has taken the lead as the main source of competitive advantage, especially in AI applications. The challenge to expand this investigation to wider frameworks of analysis and new methodologies has great potential for the supply chain research community and practitioners. To consolidate this knowledge, this paper examined a representative sample of works, using a systematic review with meta-syntheses methodology.

The main contributions of the paper to the literature on AI applied to supply chain operations are to:

a Provide an unprecedented integrated literature review which combined both BDA and AI papers applied to SCM in a comparative methodological framework of research analysis since the year 2000.

b Investigate AI in SCM operations considering the six areas of the SCOR model, including enable, which has not been done before.

c Review AI model implementation based on the main categories of application and techniques employed in SCM.

d Propose an integrated conceptual analysis that provided a cross-referenced outlook of main practices not only in supply chain management areas but also regarding the use of modelling and algorithm.

The theoretical and managerial implications along with opportunities for future research were also unveiled. First and foremost, there is a need to further increase research on source and return SCOR areas, especially when considering only AI papers. Secondly, AI application in sourcing and procurement is mostly related to improving supplier selection processes and can be expanded to other areas of inbound operations. Thirdly, there is a concentration of works related to improving demand forecasting in the plan SCOR area. New practices such as anticipatory shipping, which could revolutionise the industry, have not yet been fully tested and validated. Fourth, enable analysis showed how critical is to consider this dimension when assessing the application of AI in SCM. Topics such as resources (including human resources), contracts, managing regulatory compliance and risk management have played a pivotal role in facilitating the implementation of models in real business settings. This literature review has proven this statement, as the enable area ranked second in the overall analysis and has the potential to be further addressed. Fifth, AI techniques with proven efficiency such as reinforcement learning could be further applied along with machine learning and neural networks. Sixth, IoT combined with AI were addressed by only six papers, indicating that this integration is in its early days. Last but not least, quantitative research is concentrated in model application, mainly in planning and operational optimisation areas.

As limitations of the study, our review was based on the literature of AI and BDA using 144 articles published from 2000 to 2019. The results may vary depending on the keywords chosen. Secondly, the conceptual framework was based on meta synthesis methodology from Soni and Kodali (2011). Different methodologies could be used in future works to avoid recurrent results. Finally, the classification proposed could vary according to the researcher's interpretation of results. The main limitations of this approach are related to the selected scientific databases, (i.e., Scopus and Web of Science), document type, (i.e., articles), language, (i.e., English) and phrases researched, which can exclude some papers. The papers not included in the dataset could be pertinent to the field, but it is not likely they would change the results of this review.

Taking everything into consideration, a consistent outlook of the main research as regards AI and big data has been provided. Companies are increasingly investing in analytics solutions that are bound to unleash unprecedented results based on the new concepts of smart and data driven supply chains. However, this paper has brought evidence that there are niches of excellence in which these new techniques have already been broadly applied, (e.g., demand planning) while others still lack further research and practical implementation (e.g., sustainability and reverse logistics). The true impact of digital firms will be reached when organisations commit their resources to fully capture the value of data management in the entire supply chain. Processes related to source, make, deliver, return, plan and enable management should be considered as a unified data analysis approach from which patterns and predictive actions should be performed. This literature review paper, organised in the light of the SCOR model, illustrates how unbalanced these AI and big data supply chain initiatives are currently being applied. Corporate leadership should invest in training, culture change management, centralised information technologies and supply chain and business strategy alignment to amplify the outreach of AI, thus bring the digital transformation to the forefront of supply chain daily activities.

References

Aggarwal, A.K. and Davè, D.S. (2018) 'An artificial intelligence approach to curtailing the bullwhip effect in supply chains', *IUP Journal of Supply Chain Management*, Vol. 15, No. 4, pp.51–58.

Akter, S., Michael, K., Uddin, M.R., McCarthy, G. and Rahman, M. (2020) 'Transforming business using digital innovations: the application of AI, blockchain, cloud and data analytics', *Annals of Operations Research*, Vol. 1, No. 33, pp.1–33.

Albergaria, M and Jabbour, C.J.C. (2020) 'The role of big data analytics capabilities (BDAC) in understanding the challenges of service information and operations management in the sharing economy: evidence of peer effects in libraries', *International Journal of Information Management*, Vol. 51, No. 102023, pp.1–13.

Amirkolaii, K.N., Baboli, A., Shahzad, M.K. and Tonadre, R. (2017) 'Demand forecasting for irregular demands in business aircraft spare parts supply chains by using artificial intelligence (AI)', *IFAC-PapersOnLine*, Vol. 50, No. 1, pp.15221–15226.

Andersson, J. and Jonsson, P. (2018) 'Big data in spare parts supply chains: the potential of using product-in-use data in aftermarket demand planning', *International Journal of Physical Distribution and Logistics Management*, Vol. 48, No. 5, pp.524–544.

Arunachalam, D., Kumar, N. and Kawalek, J.P. (2018) 'Understanding big data analytics capabilities in supply chain management: unravelling the issues, challenges and implications for practice', *Transportation Research Part E: Logistics and Transportation Review*, Vol. 114, pp.416–436.

Arya, V., Sharma, P., Singh, A. and De Silva, P.T.M. (2017) 'An exploratory study on supply chain analytics applied to spare parts supply chain', *Benchmarking*, Vol. 24, No. 6, pp.1571–1580.

Aryal, A., Liao, Y., Nattuthurai, P. and Li, B. (2018) 'The emerging big data analytics and IoT in supply chain management: a systematic review', *Supply Chain Management: An International Journal*, Vol. 25, No. 2, pp.141–156.

Barbosa, M.W., Vicente, A.D., Ladeira, M.B. and de Oliveira, M.P.V. (2018) 'Managing supply chain resources with big data analytics: a systematic review', *International Journal of Logistics-Research and Applications*, Vol. 21, No. 3, pp.177–200.

Boldt, L.C., Vinayagamoorthy, V., Winder, F., Schnittger, M., Ekran, M., Mukkamala, R.R., Lassen, N.B., Flesch, B., Hussain, A. and Vatrapu, R. (2016) 'Forecasting Nike's sales using Facebook data' in *IEEE International Conference on Big Data*.

Borade, A.B. and Sweeney, E. (2015) 'Decision support system for vendor managed inventory supply chain: a case study', *International Journal of Production Research*, Vol. 53, No. 16, pp.4789–4818.

Bowers, M.R., Petrie, A.G. and Holcomb, M.C. (2017) 'Unleashing the potential of supply chain analytics', *MIT Sloan Management Review*, Vol. 59, No. 1, pp.14–16.

Brinch, M. (2018) 'Understanding the value of big data in supply chain management and its business processes: towards a conceptual framework', *International Journal of Operations and Production Management*, Vol. 38, No. 7, pp.1589–1614.

Bumblauskas, D., Gemmill, D., Igou, A. and Anzengruber, J. (2017) 'Smart maintenance decision support systems (SMDSS) based on corporate big data analytics', *Expert Systems with Applications*, Vol. 90, pp.303–317.

Buyukozkan, G. and Gocer, F. (2018) 'Digital supply chain: literature review and a proposed framework for future research', *Computers in Industry*, Vol. 97, pp.157–177.

Carbonneau, R., Laframboise, K. and Vahidov, R. (2008) 'Application of machine learning techniques for supply chain demand forecasting', *European Journal of Operational Research*, Vol. 184, No. 3, pp.1140–1154.

Carbonneau, R., Vahidov, R. and Laframboise, K. (2007) 'Machine learning-based demand forecasting in supply chains', *International Journal of Intelligent Information Technologies*, Vol. 3, No. 4, pp.40–57.

Carbonneau, R., Vahidov, R. and Laframboise, K. (2012) 'Forecasting supply chain demand using machine learning algorithms', in *Machine Learning: Concepts, Methodologies, Tools and Applications*, pp.1652–1686, IGI Global, USA.

Castillo-Villar, K.K. and Herbert-Acero, J.F. (2013) 'The effect of individual representation on the performance of a genetic algorithm applied to a supply chain network design problem', *International Journal of Supply Chain Management*, Vol. 2, No. 3, pp.17–24.

Cavalcante, I.M., Frazzon, E.M., Forcellini, F.A. and Ivanov, D. (2019) 'A supervised machine learning approach to data-driven simulation of resilient supplier selection in digital manufacturing', *International Journal of Information Management*, Vol. 49, pp.86–97.

Chaharsooghi, S.K., Heydari, J. and Zegordi, S.H. (2008) 'A reinforcement learning model for supply chain ordering management: an application to the beer game', *Decision Support Systems*, Vol. 45, No. 4, pp.949–959.

Chavez, R., Yu, W., Jacobs, M.A. and Feng, M. (2017) 'Data-driven supply chains, manufacturing capability and customer satisfaction', *Production Planning and Control*, Vol. 28, Nos. 11–12, pp.906–918.

Chehbi-Gamoura, S., Derrouiche, R., Damand, D. and Barth, M. (2020) 'Insights from big data analytics in supply chain management: an all-inclusive literature review using the SCOR model', *Production Planning and Control*, Vol. 31, No. 5, pp.355–382.

Chen, C. and Xu, C. (2018) 'A negotiation optimization strategy of collaborative procurement with supply chain based on multi-agent system', *Mathematical Problems in Engineering*, Vol. 2018, No. 4653648, pp.1–8.

Chen, D.Q., Preston, D.S. and Swink, M. (2015) 'How the use of big data analytics affects value creation in supply chain management', *Journal of Management Information Systems*, Vol. 32, No. 4, pp.4–39.

Cheng, O.K.M. and Lau, R.Y.K. (2016) 'Exploring big data analytics for supply chain management', *International Conference on Management, Economics and Social Development (ICMESD 2016)*, pp.1111–1117.

Chi, H.M., Ersoy, O.K., Moskowitz, H. and Ward, J. (2007) 'Modeling and optimizing a vendor managed replenishment system using machine learning and genetic algorithms', *European Journal of Operational Research*, Vol. 180, No. 1, pp.174–193.

Choi, T.M., Wallace, S.W. and Wang, Y. (2018) 'Big data analytics in operations management', *Production and Operations Management*, Vol. 27, No. 10, pp.1868–1883.

Choi, T-M. (2018) 'Incorporating social media observations and bounded rationality into fashion quick response supply chains in the big data era', *Transportation Research Part E: Logistics and Transportation Review*, Vol. 114, pp.386–397.

Côrte-Real, N., Oliveira, T. and Ruivo, P. (2017) 'Assessing business value of big data analytics in European firms', *Journal of Business Research*, Vol. 70, pp.379–390.

Cox, M. and Ellsworth, D. (1997) *Application-Controlled Demand Paging for Out-of-Core Visualization*, in Report NAS-97-010.

Cui, R., Gallino, S., Moreno, A. and Zhang, D.J. (2018) 'The Operational Value of Social Media information', *Production and Operations Management*, Vol. 27, No. 10, pp.1749–1769.

Curcio, D., Longo, F., Mirabelli, G. and Papoff, E. (2007) 'Pharmaceutical routes optimization using artificial intelligence techniques', in *Proceedings of the 4th IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, pp.238–242.

De Santis, R.B., De Aguiar, E.P. and Goliatt, L. (2017) 'Predicting Material backorders in inventory management using machine learning', in *Proceedings of the IEEE Latin American Conference on Computational Intelligence (La-Cci).*

DHL (2018) Logistics Trends [online] https://www.dhl.com/en/about_us/logistics_insights/dhl_trend_research/trendradar.html#.W6djlntKi00 (accessed 2 October 2019).

Didehkhani, H., Jassbi, J. and Pilevari, N. (2009) 'Assessing flexibility in supply chain using adaptive neuro fuzzy inference system', in *Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management*, Hong-Kong.

Dubey, R., Gunasekaran, A., Childe, S.J., Wamba, S.F. and Papadopoulos, T. (2016) 'The impact of big data on world-class sustainable manufacturing', *The International Journal of Advanced Manufacturing Technology*, Vol. 84, Nos. 1–4, pp.631–645.

Dubey, R., Luo, Z., Gunasekaran, A., Akter, S., Hazen, B.T. and Douglas, M.A. (2018) 'Big data and predictive analytics in humanitarian supply chains: enabling visibility and coordination in the presence of swift trust', *The International Journal of Logistics Management*, Vol. 29, No. 2, pp.485–512.

Dubey, R., Gunasekaran, A., Childe, S.J., Blome, C. and Papadopoulos, T. (2019a) 'Big data and predictive analytics and manufacturing performance: integrating institutional theory, resource‐based view and big data culture', *British Journal of Management*, Vol. 30, No. 2, pp.341–361.

Dubey, R., Gunasekaran, A., Childe, S.J., Bryde, D.J., Giannakis, M., Foropon, C., Roubaud, D. and Hazen, B.T. (2019b) 'Big data analytics and artificial intelligence pathway to operational performance under the effects of entrepreneurial orientation and environmental dynamism: a study of manufacturing organisations', *International Journal of Production Economics*, Vol. 226, No. 107599, pp.1–12.

Dubey, R., Gunasekaran, A., Childe, S.J., Papadopoulos, T., Luo, Z., Wamba, S.F. and Roubaud, D. (2019c) 'Can big data and predictive analytics improve social and environmental sustainability?', *Technological Forecasting and Social Change*, Vol. 144, pp.534–545.

Dubey, R., Gunasekaran, A. and Childe, S.J. (2019d) 'Big data analytics capability in supply chain agility', *Management Decision*, Vol. 57, No. 8, pp.2092–2112.

Dweekat, A.J., Hwang, G. and Park, J. (2017) 'A supply chain performance measurement approach using the internet of things: toward more practical SCPMS', *Industrial Management and Data Systems*, Vol. 117, No. 2, pp.267–286.

Efendigil, T., Önüt, S. and Kahraman, C. (2009) 'A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: a comparative analysis', *Expert Systems with Applications*, Vol. 36, No. 3, pp.6697–6707.

Ehret, M. and Wirtz, J. (2017) 'Unlocking value from machines: business models and the industrial internet of things', *Journal of Marketing Management*, Vol. 33, Nos. 1–2, pp.111–130.

Engelseth, P. and Wang, H. (2018) 'Big data and connectivity in long-linked supply chains', *Journal of Business and Industrial Marketing*, Vol. 33, No. 8, pp.1201–1208.

Fernando, Y., Chidambaram, R.R.M. and Wahyuni-Td, I.S. (2018) 'The impact of big data analytics and data security practices on service supply chain performance', *Benchmarking*, Vol. 25, No. 9, pp.4009–4034.

Fu, J. and Fu, Y. (2015) 'An adaptive multi-agent system for cost collaborative management in supply chains', *Engineering Applications of Artificial Intelligence*, Vol. 44, pp.91–100.

Garg, V.K. and Viswanadham, N. (2010) 'EcoSupply: a machine learning framework for analyzing the impact of ecosystem on global supply chain dynamics', in *Proceedings of the Asia-Pacific Conference on Simulated Evolution and Learning*, pp.677–686.

Giannakis, M. and Louis, M. (2011) 'A multi-agent based framework for supply chain risk management', *Journal of Purchasing and Supply Management*, Vol. 17, No. 1, pp.23–31.

Giannakis, M. and Louis, M. (2016) 'A multi-agent based system with big data processing for enhanced supply chain agility', *Journal of Enterprise Information Management*, Vol. 29, No. 5, pp.706–727.

Giannoccaro, I. and Pontrandolfo, P. (2002) 'Inventory management in supply chains: a reinforcement learning approach', *International Journal of Production Economics*, Vol. 78, No. 2, pp.153–161.

Govidan, K., Cheng, T.C.E., Mishra, N. and Shukla, N. (2018) 'Big data analytics and application for logistics and supply chain management', *Transportation Research Part E: Logistics and Transportation Review*, Vol. 114, pp.343–349.

Gravili, G., Benvenuto, M., Avram, A. and Viola, C. (2018) 'The influence of the digital divide on big data generation within supply chain management', *International Journal of Logistics Management*, Vol. 29, No. 2, pp.592–628.

Grawe, S.J. (2009) 'Logistics innovation: a literature-based conceptual framework', *International Journal of Logistics Management*, Vol. 20, No. 3, pp.360–377.

Green, F.B. (2001) 'Managing the unmanageable: integrating the supply chain with new developments in software', *Supply Chain Management*, Vol. 6, No. 5, pp.208–211.

Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S.F., Childe, S.J., Hazen, B. and Akter, S. (2017) 'Big data and predictive analytics for supply chain and organizational performance', *Journal of Business Research*, Vol. 70, pp.308–317.

Guo, Z.X., Wong, W.K., Leung, S.Y.S. and Min, L. (2011) 'Applications of artificial intelligence in the apparel industry: a review', *Textile Research Journal*, Vol. 81, No. 18, pp.1871–1892.

Guosheng, H. and Guohong, Z. (2008) 'Comparison on neural networks and support vector machines in suppliers' selection', *Journal of Systems Engineering and Electronics*, Vol. 19, No. 2, pp.316–320.

Gupta, S., Qian, X., Bhushan, B. and Luo, Z. (2019) 'Role of cloud ERP and big data on firm performance: a dynamic capability view theory perspective', *Management Decision*, Vol. 57, No. 8, pp.1857–1882.

Gyulai, D., Pfeiffer, A., Nick, G., Gallina, V., Sihn, W. and Monostori, L. (2018) 'Lead time prediction in a flow-shop environment with analytical and machine learning approaches', *IFAC-PapersOnLine*, Vol. 51, No. 11, pp.1029–1034.

Hazen, B.T., Boone, C.A., Ezell, J.D. and Jones-Farmer L.A. (2014) '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*, Vol. 154, pp.72–80.

Hazen, B.T., Skipper, J.B., Boone, C.A. and Hill, R.R. (2018) 'Back in business: operations research in support of big data analytics for operations and supply chain management', *Annals of Operations Research*, Vol. 270, Nos. 1–2, pp.201–211.

Hazen, B.T., Skipper, J.B., Ezell, J.D. and Boone, C.A. (2016) 'Big data and predictive analytics for supply chain sustainability: a theory-driven research agenda', *Computers and Industrial Engineering*, Vol. 101, pp.592–598.

Hiromoto, R.E., Haney, M. and Vakanski, A. (2017) 'A secure architecture for IoT with supply chain risk management', *in Proceedings of the 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*.

Hofmann, E. (2017) 'Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect', *International Journal of Production Research*, Vol. 55, No. 17, pp.5108–5126.

Hofmann, E. and Rutschmann, E. (2018) 'Big data analytics and demand forecasting in supply chains: a conceptual analysis', *The International Journal of Logistics Management*, Vol. 29, No. 2, pp.739–766.

Hong, G.H. and Ha, S.H. (2008) 'Evaluating supply partner's capability for seasonal products using machine learning techniques', *Computers and Industrial Engineering*, Vol. 54, No. 4, pp.721–736.

Hopkins, J. and Hawking, P. (2018) 'Big data analytics and IoT in logistics: a case study', *International Journal of Logistics Management*, Vol. 29, No. 2, pp.575–591.

Hu, H., Wen, Y., Chua, T.S. and Li, X. (2014) 'Toward scalable systems for big data analytics: a technology tutorial', *IEEE Access*, Vol. 2, pp.652–687.

Ilie-Zudor, E., Ekárt, A., Kemeny, Z., Buckingham, C., Welch, P. and Monostori, L. (2015) 'Advanced predictive-analysis-based decision support for collaborative logistics networks', *Supply Chain Management: An International Journal*, Vol. 20, No. 4, pp.369–388.

Ittmann, H.W. (2015) 'The impact of big data and business analytics on supply chain management', *Journal of Transport and Supply Chain Management*, Vol. 9, No. 1, pp.1–9.

Ivanov, D., Dolgui, A. and Sokolov, B. (2019) 'The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics', *International Journal of Production Research*, Vol. 57, No. 3, pp.829–846.

Jafarzadeh-Ghoushchi, S. and Rahman, M.N.A. (2016) 'Performance study of artificial neural network modelling to predict carried weight in the transportation system', *International Journal of Logistics Systems and Management*, Vol. 24, No. 2, pp.200–212.

Jeble, S., Kumari, S. and Patil, Y. (2018) 'Role of big data in decision making', *Operations and Supply Chain Management*, Vol. 11, No. 1, pp.36–44.

Jiang, C. and Sheng, Z. (2009) 'Case-based reinforcement learning for dynamic inventory control in a multi-agent supply-chain system', *Expert Systems with Applications*, Vol. 36, No. 3, pp.6520–6526.

Kache, F. and Seuring, S. (2017) 'Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management', *International Journal of Operations and Production Management*, Vol. 37, No. 1, pp.10–36.

Kamal, M.M. and Irani, Z. (2014) 'Analysing supply chain integration through a systematic literature review: a normative perspective', *Supply Chain Management: An International Journal*, Vol. 19, Nos. 5–6, pp.523–557.

Kamble, S.S. and Gunasekaran, A. (2020) 'Big data driven supply chain performance measurement system: a review and framework for implementation', *International Journal of Production Research*, Vol. 58, No. 1, pp.65–86.

Kar, A.K. (2015) 'A hybrid group decision support system for supplier selection using analytic hierarchy process, fuzzy set theory and neural network', *Journal of Computational Science*, Vol. 6, pp.23–33.

Kartal, H., Oztekin, A., Gunasekaran, A. and Cebi, F. (2016) 'An integrated decision analytic framework of machine learning with multi-criteria decision making for multi-attribute inventory classification', *Computers and Industrial Engineering*, Vol. 101, pp.599–613.

Kaur, H. and Singh, S.P. (2018) 'Heuristic modeling for sustainable procurement and logistics in a supply chain using big data', *Computers and Operations Research*, Vol. 98, pp.301–321.

Kazemi, A. and Fazel Zarandi, M.H (2008) 'An agent-based framework for building decision support system in supply chain management', *Journal of Applied Sciences*, Vol. 8, No. 7, pp.1125–1137.

Kiekintveld, C., Miller, J., Jordan, P.R., Callender, L.F. and Wellman, M.P. (2009) 'Forecasting market prices in a supply chain game', *Electronic Commerce Research and Applications*, Vol. 8, No. 2, pp.63–77.

Kong, F. and Li, J. (2018) 'Supply chain flexibility enhancement based on deep belief network', *Jisuanji Jicheng Zhizao Xitong/Computer Integrated Manufacturing Systems CIMS*, Vol. 24, No. 5, pp.1292–1300.

Kumar, D., Singh, J., Singh and O.M. and Seema (2013) 'A fuzzy logic based decision support system for evaluation of suppliers in supply chain management practices', *Mathematical and Computer Model*ling, Vol. 58, Nos. 11–12, pp.1679–1695.

Lai, Y., Sun, H. and Ren, J. (2018) 'Understanding the determinants of big data analytics (BDA) adoption in logistics and supply chain management: an empirical investigation', *International Journal of Logistics Management*, Vol. 29, No. 2, pp.676–703.

Lamba, K. and Singh, S.P. (2017) 'Big data in operations and supply chain management: current trends and future perspectives', *Production Planning and Control*, Vol. 28, Nos. 11–12, pp.877–890.

Lau, R.Y.K., Zhang, W. and Xu, W. (2018) 'Parallel aspect-oriented sentiment analysis for sales forecasting with big data', *Production and Operations Management*, Vol. 27, No. 10, pp.1775–1794.

Lee, C.K.H. (2017) 'A GA-based optimization model for big data analytics supporting anticipatory shipping in Retail 4.0.', *International Journal of Production Research*, Vol. 55, No. 2, pp.593–605.

Li, J., Tao, F., Cheng, Y. and Zhao, L. (2015) 'Big data in product lifecycle management', *The International Journal of Advanced Manufacturing Technology*, Vol. 81, Nos. 1–4, pp.667–684.

Lima-Junior, F.R. and Carpinetti, L.C.R. (2016) 'Combining SCOR® model and fuzzy TOPSIS for supplier evaluation and management', *International Journal Production Economics*, Vol. 174, pp.128–141.

Ma, H., Wang, Y. and Wang, K. (2018) 'Automatic detection of false positive RFID readings using machine learning algorithms', *Expert Systems with Applications*, Vol. 91, pp.442–451.

Mahroof, K. (2019) 'A human-centric perspective exploring the readiness towards smart warehousing: the case of a large retail distribution warehouse', *International Journal of Information Management*, Vol. 45, pp.176–190.

Mandal, S. (2018) 'An examination of the importance of big data analytics in supply chain agility development: a dynamic capability perspective', *Management Research Review*, Vol. 41, No. 10, pp.1201–1219.

Mandal, S. (2019) 'The influence of big data analytics management capabilities on supply chain preparedness, alertness and agility: an empirical investigation', *Information Technology and People*, Vol. 32, No. 2, pp.297–318.

Mani, V., Delgado, C., Hazen, B.T. and Patel, P. (2017) 'Mitigating supply chain risk via sustainability using big data analytics: Evidence from the manufacturing supply chain', *Sustainability (Switzerland)*, Vol. 9, No. 4, pp.1–21.

Matthias, O., Fouweather, I., Gregory, I. and Vernon, A. (2017) 'Making sense of big data – can it transform operations management?', *International Journal of Operations and Production Management*, Vol. 37, No. 1, pp.37–55.

Mikalef, P., Boura, M., Lekakos, G. and Krogstie, J. (2019) 'Big data analytics capabilities and innovation: the mediating role of dynamic capabilities and moderating effect of the environment', *British Journal of Management*, Vol. 30, No. 2, pp.272–298.

Mikalef, P., Pappas, I.O., Krogstie, J. and Giannakos, M. (2018) 'Big data analytics capabilities: a systematic literature review and research agenda', *Information Systems and e.Business Management*, Vol. 16, No. 3, pp.547–578.

Min, H. (2010) 'Artificial intelligence in supply chain management: theory and applications', *International Journal of Logistics Research and Applications*, Vol. 13, No. 1, pp.13–39.

Min, H. (2015) 'Genetic algorithm for supply chain modelling: basic concepts and applications', *International Journal of Services and Operations Management*, Vol. 22, No. 2, pp.143–164.

Mishra, D., Gunasekaran, A., Papadopoulos, T. and Childe, S.J. (2018) 'Big data and supply chain management: a review and bibliometric analysis', *Annals of Operations Research*, Vol. 270, Nos. 1–2, pp.313–336.

Mojaveri, H.R.S., Mousavi, S.S., Heydar, M. and Aminian, A. (2009) 'Validation and selection between machine learning technique and traditional methods to reduce bullwhip effects: a data mining approach', *World Academy of Science, Engineering and Technology*, Vol. 37, pp.555–561.

Mokhtarinejad, M., Ahmadi, A., Karimi, B. and Rahmati, S.H.A. (2015) 'A novel learning based approach for a new integrated location-routing and scheduling problem within cross-docking considering direct shipment', *Applied Soft Computing Journal*, Vol. 34, pp.274–285.

Moktadir, M.A., Ali, S.M., Paul, S.K. and Shukla, N. (2019) 'Barriers to big data analytics in manufacturing supply chains: a case study from Bangladesh', *Computers and Industrial Engineering*, Vol. 128, pp.1063–1075.

Moraga, R., Rabelo, L. Jones, A. and Vila, J. (2011) 'Using neural networks to monitor supply chain behaviour', *International Journal of Computer Applications in Technology*, Vol. 40, Nos. 1–2, pp.53–63.

Moretto, A., Ronchi, S. and Patrucco, A.S. (2017) 'Increasing the effectiveness of procurement decisions: the value of big data in the procurement process', *International Journal of RF Technologies-Research and Applications*, Vol. 8, No. 3, pp.79–103.

Nguyen, T., Li, Z., Spiegler, V., Ieromonachou, P. and Lin, Y. (2018) 'Big data analytics in supply chain management: a state-of-the-art literature review', *Computers and Operations Research*, Vol. 98, pp.254–264.

Nita, S. (2015) 'Application of big data technology in support of food manufacturers' commodity demand forecasting', *NEC Technical Journal*, Vol. 10, No. 1, pp.90–93.

Niu, B., Dai, Z. and Zhuo, X. (2019) 'Co-opetition effect of promised-delivery-time sensitive demand on air cargo carriers' big data investment and demand signal sharing decisions', *Transportation Research Part E: Logistics and Transportation Review*, Vol. 123, pp.29–44.

O'Donovan, P., Leahy, K., Bruton, K. and O'Sullivan, D.T.J. (2015) 'Big data in manufacturing: a systematic mapping study', *Journal of Big Data*, Vol. 2, No. 20, pp.1–22.

Orji, I.J. and Wei, S. (2015) 'An innovative integration of fuzzy-logic and systems dynamics in sustainable supplier selection: a case on manufacturing industry', *Computers and Industrial Engineering*, Vol. 88, pp.1–12.

Papadopoulos, T., Gunasekaran, A., Dubey, R., Altay, N., Childe, S.J. and Fosso-Wamba, S. (2017) 'The role of big data in explaining disaster resilience in supply chains for sustainability', *Journal of Cleaner Production*, Vol. 142, pp.1108–1118.

Park, Y.B., Yoon, S.J. and Yoo, J.S. (2018) 'Development of a knowledge-based intelligent decision support system for operational risk management of global supply chains', *European Journal of Industrial Engineering*, Vol. 12, No. 1, pp.93–115.

Pereira, M.M., de Oliveira, D.L., Portela Santos, P.P. and Frazzon, E.M. (2018) 'Predictive and adaptive management approach for omnichannel retailing supply chains', *IFAC-PapersOnLine*, Vol. 51, No. 11, pp.1707–1713.

Piramuthu, S. (2005a) 'Machine learning for dynamic multi-product supply chain formation', *Expert Systems with Applications*, Vol. 29, No. 4, pp.985–990.

Piramuthu, S. (2005b) 'Knowledge-based framework for automated dynamic supply chain configuration – production, manufacturing and logistics', *European Journal of Operational Research*, Vol. 165, No. 1, pp.219–230.

Piramuthu, S. and Sikora, R. (2005) 'Efficient genetic algorithm based data mining using feature selection with Hausdorff distance', *Information and Technology Management*, Vol. 6, pp.315–331.

Pontrandolfo, P., Gosavi, A., Okogbaa, O.G. and Das, T.K. (2002) 'Global supply chain management: a reinforcement learning approach', *International Journal of Production Research*, Vol. 40, No. 6, pp.1299–1317.

Prasad, S., Zakaria, R. and Altay, N. (2018) 'Big data in humanitarian supply chain networks: a resource dependence perspective', *Annals of Operations Research*, Vol. 270, Nos. 1–2, pp.383–413.

Rampersad, G. (2020) 'Robot will take your job: innovation for an era of artificial intelligence', *Journal of Business Research*, Vol. 116, pp.68–74.

Raut, R.D. Kamble, S.S., Kharat, M.G., Joshi, H., Singhal, C. and Kamble, S.J. (2017) 'A hybrid approach using data envelopment analysis and artificial neural network for optimizing 3PL supplier selection', *International Journal of Logistics Systems and Management*, Vol. 26, No. 2, pp.203–223.

Raut, R.D., Mangla, S.K., Narwane, V.S., Gardas, B.B., Priyadarshinee, P. and Narkhede, B.E. (2019) 'Linking big data analytics and operational sustainability practices for sustainable business management', *Journal of Cleaner Production*, Vol. 224, pp.10–24.

Rehman, M.H.H., Chang, V., Batool, A. and Wahh, T.Y. (2016) 'Big data reduction framework for value creation in sustainable enterprises', *International Journal of Information Management*, Vol. 36, No. 6, pp.917–928.

Ren, S., Zhang, Y.F., Liu, Y., Sakao, T., Huisingh, D. and Almeida, C. (2019) 'A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: a framework, challenges and future research directions', *Journal of Cleaner Production*, Vol. 210, pp.1343–1365.

Richey, R.G., Morgan, T.R., Lindsey-Hall, K. and Adams, F.G. (2016) 'A global exploration of big data in the supply chain', *International Journal of Physical Distribution and Logistics Management*, Vol. 46, No. 8, pp.710–739.

Rodriguez, L. and Da Cunha, C. (2018) 'Impacts of big data analytics and absorptive capacity on sustainable supply chain innovation: a conceptual framework', *Logforum*, Vol. 14, No. 2, pp.151–161.

Roßmann, B., Canzaniello, A., Von der Gracht, H. and Hartmann, E. (2018) 'The future and social impact of big data analytics in supply chain management: results from a Delphi study', *Technological Forecasting and Social Change*, Vol. 130, pp.135–149.

Russell, S. and Norvig, P. (1995) *Artificial Intelligence: A Modern Approach*, Prentice-Hall, Upper Saddle River, NJ.

Sanders, N.R. (2016) 'How to use big data to drive your supply chain', *California Management Review*, Vol. 58, No. 3, pp.26–48.

Supply Chain Council (SCC) (2012) *Supply Chain Operations Reference Model*, Supply Chain Council.

Schoenherr, T. and Speier‐Pero, C. (2015) 'Data science, predictive analytics, and big data in supply chain management: current state and future potential', *Journal of Business Logistics*, Vol. 36, No. 1, pp.120–132.

Shahrabi, J., Mousavi, S.S. and Heydar, M. (2009) 'Supply chain demand forecasting: a comparison of machine learning techniques and traditional methods', *Journal of Applied Sciences*, Vol. 9, No. 3, pp.521–527.

Shokouhyar, S., Seifhashemi, S., Siadat, H. and Ahmadi, M.M. (2019) 'Implementing a fuzzy expert system for ensuring information technology supply chain', *Expert Systems*, Vol. 36, No. 1, pp.e12339.

Shukla, M. and Tiwari, M.K. (2017) 'Big-data analytics framework for incorporating smallholders in sustainable palm oil production', *Production Planning and Control*, Vol. 28, No. 16, pp.1365–1377.

Simchi-Levi, D. and Wu, M.X. (2018) 'Powering retailers' digitization through analytics and automation', *International Journal of Production Research*, Vol. 56, Nos. 1–2, pp.809–816.

Singh, A., Shukla, N. and Mishra, N. (2018) 'Social media data analytics to improve supply chain management in food industries', *Transportation Research Part E: Logistics and Transportation Review*, Vol. 114, pp.398–415.

Singh, L.P. and Challa, R.T. (2016) 'Integrated forecasting using the discrete wavelet theory and artificial intelligence techniques to reduce the bullwhip effect in a supply Chain', *Global Journal of Flexible Systems Management*, Vol. 17, No. 2, pp.157–169.

Singh, S.K. and El-Kassar, A.N. (2019) 'Role of big data analytics in developing sustainable capabilities', *Journal of Cleaner Production*, Vol. 213, pp.1264–1273.

Siurdyban, A. and Møller, C. (2012) 'Towards intelligent supply chains: a unified framework for business process design', *International Journal of Information Systems and Supply Chain Management*, Vol. 5, No. 1, pp.1–19.

Slimani, I. (2017) 'Configuration and implementation of a daily artificial neural network-based forecasting system using real supermarket data', *International Journal of Logistics Systems and Management*, Vol. 28, No. 2, pp.144–163.

Slimani, I., El Farissi, I. and Achchab, S. (2015) 'Application of game theory and neural network to study the behavioral probabilities in supply chain', *Journal of Theoretical and Applied Information Technology*, Vol. 82, No. 3, pp.411–416.

Smyth, K.B., Croxton, K.L., Franklin, R. and Knemeyer, A.M. (2018) 'Thirsty in an ocean of data? Pitfalls and practical strategies when partnering with industry on big data supply chain research', *Journal of Business Logistics*, Vol. 39, No. 3, pp.203–219.

Sodero, A., Jin, Y.H. and Barratt, M. (2019) 'The social process of big data and predictive analytics use for logistics and supply chain management', *International Journal of Physical Distribution and Logistics Management*, Vol. 49, No. 7, pp.706–726.

Soni, G. and Kodali, R. (2011) 'A critical analysis of supply chain management content in empirical research', *Business Process Management Journal*, Vol. 17, No. 2, pp.238–266.

Srinivasan, R. and Swink, M. (2018) 'An investigation of visibility and flexibility as complements to supply chain analytics: an organizational information processing theory perspective', *Production and Operations Management*, Vol. 27, No. 10, pp.1849–1867.

Sun, Z-L., Choi, T-M., Au, K-F. and Yu, Y. (2008) 'Sales forecasting using extreme learning machine with applications in fashion retailing', *Decision Support Systems*, Vol. 46, No. 1, pp.411–419.

Tiwari, S., Wee, H.M. and Daryanto, Y. (2018) 'Big data analytics in supply chain management between 2010 and 2016: Insights to industries', *Computers and Industrial Engineering*, Vol. 115, pp.319–330.

Tripathi, S. and Gupta, M. (2020) 'Transforming towards a smarter supply chain', *International Journal of Logistics Systems and Management*, Vol. 36, No. 3, pp.319–342.

Tsang, Y.P., Choy, K.L., Wu, C.H., Ho, G.T.S., Lam, C.H.Y. and Koo, P.S. (2018) 'An internet of things (IoT)-based risk monitoring system for managing cold supply chain risks', *Industrial Management and Data Systems*, Vol. 118, No. 7, pp.1432–1462.

Tsao, Y.C. (2017) 'Managing default risk under trade credit: who should implement big-data analytics in supply chains?', *Transportation Research Part E: Logistics and Transportation Review*, Vol. 106, pp.276–293.

Tse, Y.K., Chan, T.M. and Lie, R.H. (2009) 'Solving complex logistics problems with multi-artificial intelligent system', *International Journal of Engineering Business Management*, Vol. 1, No. 1, pp.37–48.

Urciuoli, L. and Hintsa, J. (2018) 'Improving supply chain risk management – can additional data help?', *International Journal of Logistics Systems and Management*, Vol. 30, No. 2, p.195.

Vahdani, B., Dehbari, S., Naderi-Beni, M. and Kh, E.Z. (2014) 'An artificial intelligence approach for fuzzy possibilistic-stochastic multi-objective logistics network design', *Neural Computing and Applications*, Vol. 25, Nos. 7–8, pp.1887–1902.

Valluri, A., North, M.J. and Macal, C.M. (2009) 'Reinforcement learning in supply chains', *International Journal of Neural Systems*, Vol. 19, No. 5, pp.331–344.

Waller, M.A. and Fawcett, S.E. (2013) 'Click here for a data scientist: Big data, predictive analytics, and theory development in the era of a maker movement supply chain', *Journal of Business Logistics*, Vol. 34, No. 4, pp.249–252.

Wamba, S.F., Akter, S., Edwards, A., Chopin, G. and Gnanzou, D. (2015) 'How 'big data' can make big impact: findings from a systematic review and a longitudinal case study', *International Journal Production and Economics*, Vol. 165, pp.234–246.

Wamba, S.F., Dubey, R., Gunasekaran, A. and Akter, S. (2020) 'The performance effects of big data analytics and supply chain ambidexterity: the moderating effect of environmental dynamism', *International Journal of Production Economics*, Vol. 222, p.107498.

Wamba, S.F., Gunasekaran, A., Akter, S., Ren, S.J.F., Dubey, R. and Childe, S.J. (2017) 'Big data analytics and firm performance: effects of dynamic capabilities', *Journal of Business Research*, Vol. 70, pp.356–365.

Wang, G., Gunasekaran, A., Ngai, E.W.T. and Papadopoulos, T. (2016) 'Big data analytics in logistics and supply chain management: certain investigations for research and applications', *International Journal of Production Economics*, Vol. 176, pp.98–110.

Wanke, P., Alvarenga, H., Correa, H., Hadi-Vencheh, A. and Azad, M.A.K. (2017) 'Fuzzy inference systems and inventory allocation decisions: exploring the impact of priority rules on total costs and service levels', *Expert Systems with Applications*, Vol. 85, pp.182–193.

Wieczorek, L. and Ignaciuk, P. (2018) 'Continuous genetic algorithms as intelligent assistance for resource distribution in logistic systems', *Data*, Vol. 3, No. 4, pp.1–14.

Wong, C., Skipworth, H., Godsell, J. and Achimugu, N. (2012a) 'Towards a theory of supply chain alignment enablers: a systematic literature review', *Supply Chain Management: An International Journal*, Vol. 17, No. 4, pp.419–437.

Wong, J.T., Su, C.T. and Wang, C.H. (2012b) 'Stochastic dynamic lot-sizing problem using bi-level programming base on artificial intelligence techniques', *Applied Mathematical Modelling*, Vol. 36, No. 5, pp.2003–2016.

Wu, K.J., Liao, C.J., Tseng, M.L., Lim, M.K., Hu, J. and Tan, K. (2017a) 'Toward sustainability: using big data to explore the decisive attributes of supply chain risks and uncertainties', *Journal of Cleaner Production*, Vol. 142, pp.663–676.

Wu, P-J., Chen, M-C. and Tsau, C-K. (2017b) 'The data-driven analytics for investigating cargo loss in logistics systems', *International Journal of Physical Distribution and Logistics Management*, Vol. 47, No. 1, pp.68–83.

Wu, P.J. and Lin, K.C. (2018) 'Unstructured big data analytics for retrieving e-commerce logistics knowledge', *Telematics and Informatics*, Vol. 35, No. 1, pp.237–244.

Xu, Z.Y., Sun, R. and Sun, Y.Z. (2006) 'An application of artificial neural network on performance measurement of supply chain alliance', in *Proceedings of the 13th International Conference on Industrial Engineering and Engineering Management*, Industrial Engineering and Management Innovation in New-Era, Vols. 1–5.

Yerpude, S. and Singhal, T.K. (2020) 'IoT supported SMART supply chain management for effective online retail management (e-retail) – an empirical research', *International Journal of Logistics Systems and Management*, Vol. 36, No. 3, p.441.

Yu, L., Zhao, Y., Tang, L. and Yang, Z. (2019) 'Online big data-driven oil consumption forecasting with Google trends', *International Journal of Forecasting*, Vol. 35, No. 1, pp.213–223.

Yuen, J.S.M., Choy, K.L., Lam, H.Y. and Tsang, Y.P. (2018) 'An intelligent-internet of things (IoT) outbound logistics knowledge management system for handling temperature sensitive products', *International Journal of Knowledge and Systems Science*, Vol. 9, No. 1, pp.23–40.

Zhan, Y., Tan, K.H., Li, Y. and Tse, Y.K. (2018) 'Unlocking the power of big data in new product development', *Annals of Operations Research*, Vol. 270, Nos. 1–2, pp.577–595.

Zhang, H., Xu, Z. and Lu, J.F. (2004) 'Research of partner enterprise selection in supply chain management based on support vector machine', *Jisuanji Jicheng Zhizao Xitong/Computer Integrated Manufacturing Systems CIMS*, Vol. 10, No. 7, pp.796–800.

Zhang, R., Li, J., Wu, S. and Meng, D. (2016) 'Learning to select supplier portfolios for service supply chain', *PLoS ONE*, Vol. 11, No. 5, pp.1–19.

Zhao, R., Liu, Y., Zhang, N. and Huang, T. (2017) 'An optimization model for green supply chain management by using a big data analytic approach', *Journal of Cleaner Production*, Vol. 142, pp.1085–1097.

Zhong, R. Y., Xu, C., Chen, C and Huang, G.Q. (2015) 'Big data analytics for physical internet-based intelligent manufacturing shop floors', *International Journal of Production Research*, Vol. 55, No. 9, pp.2610–2621.

Zhong, R.Y., Newman, S.T., Huang, G.Q. and Lan, S. (2016) 'Big data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives', *Computers and Industrial Engineering*, Vol. 101, pp.572–591.

Zhu, Y., Xie, C., Wang, G.J. and Yan, X.G. (2017) 'Comparison of individual, ensemble and integrated ensemble machine learning methods to predict China's SME credit risk in supply chain finance', *Neural Computing and Applications*, Vol. 28, pp.41–50.

APPENDIX 6 – ABSTRACT 28th EUROMA (INTERNATIONAL EUROPEAN OPERATIONS MANAGEMENT ASSOCIATION CONFERENCE)

DAYRELL MENDONÇA, G.; LIMA JUNIOR, O.F. 'Machine Learning Algorithms Applied to Air Freight Delay Prediction', *Euroma*, 2021 (Approved but participation declined by authors)

Abstract based on Qualification text. Methodology and results have been revised for the final Master Thesis according to Qualification examining committee requests.

Qualis Capes A – Administração, Contabilidade e Turismo

Machine Learning Algorithms Applied to Air Freight Delay Prediction

Keywords: Artificial Intelligence, Data Mining, Supply Chain Risk Management

Topic(s): AI transformation and responsive SCM, Artificial Intelligence and Big Data Analytics in Operations and Supply Chain Management, Supply Chain Risk Management

Word count: 997

Purpose

The air freight cargo industry is increasingly facing operational challenges due to tougher global competition and higher service level requirements from customers. Machine Learning techniques are consequently being applied as an aviation supply chain risk management (SCRM) approach in order to predict delays, reduce operational uncertainty and reduce costs (Chung et. al., 2020).

Etani (2019) studied the application of Random Forest (RF) algorithm to predict delays based on weather conditions in Japanese airports. Herrema et al. (2019) also used RF for assessing runway capacity and utilization to avoid delays. Yu et. al. (2019) researched commercial air transport micro influential factors (e.g. air route situation and crowdedness degree of airport) that influence flight delays using Deep Learning algorithms. Congestion analysis was also a major field of study regarding delay causality in transport systems (Diana, 2018). Gui et. al. (2020) proposed a combined RF model application based on Big Data associated with flight delay factors such as airport, flight, air route and other operational information.

This study proposes the application of a supervised machine learning model that uses transportation data to predict intercontinental air freight import delivery performance in a Latin American automotive industry case study. The specific objective is to predict supply chain delay prior to cargo pick up at the supplier based on previous similar shipments. Overall, the research also adds to the current literature as both transport and supplier data is used in a specific machine learning application.

Design/Methodology/Approach

The KDD process (Knowledge Discovery Database) was applied as research methodology (Shafique and Qaiser, 2014). After data cleaning process, the air shipment database remained with 45 explanatory variables such as country, shipper and consignee information. The target attribute contained the information of whether there was a delay for each shipment (binary classification problem). Because the positive class (delayed shipments) represented only 25% of the total number of instances, the NCL (Neighbour Cleaning Cleaning) and Smote (Synthetic

Minority Oversampling Rule) algorithms were also applied to guarantee the positive class balance.

As the problem was characterized by being predictive, the task chosen in the data mining phase was of Classification using RF, SVM and KNN (K-Nearest Neighbours) algorithms with 10-fold Cross-Validation and Hold Out test options. In order to increase the performance of the KNN and SVM algorithms, the attribute selection approach with Chi-Square, Infogain, Gain Ratio and CFS (Correlation Feature Selection) methods was performed. Finally, the main criteria chosen to determine the best algorithm were Accuracy, Kappa Index, TPR (True Positive Rate), FPR (False Positive Rate), AUC (Area Under Curve) and ROC (Receiver Operator Characteristic Curve).

Findings

The SVM algorithm achieved the best result after a combined class balancing procedure (Smote) reaching an accuracy rate of 79% and a Kappa index of 0.44. Table 1 depicts the overall results:

Table 1-Application of Machine Learning Algorithms (with positive class balancing and feature selection)

Method	**Random Forest**		**SVM**		**KNN**	
	Accuracy	**Kappa**	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
NCL	76.888	0.11	80.888	0.35	77.777	0.25
SMOTE	80.444	0.30	**79.555**	**0.44**	78.222	0.39
NCL + SMOTE	77.777	0.15	80.888	0.33	79.111	0.28

In addition to presenting the best combined result of Accuracy and Kappa, the SVM algorithm also had a higher TPR and the second highest AUC value (Table 2):

Table 2- Algorithm Performance Criteria

Algorithms	**Performance Criteria**		
	FPR	**TPR**	**AUC**
KNN	13	52	0,69
Random Forest	6	23	0,77
SVM	13	57	0,72

Having evaluated all the performance indicators, the SVM algorithm was considered the best option as detailed in Figure 1 representing the ROC curve:

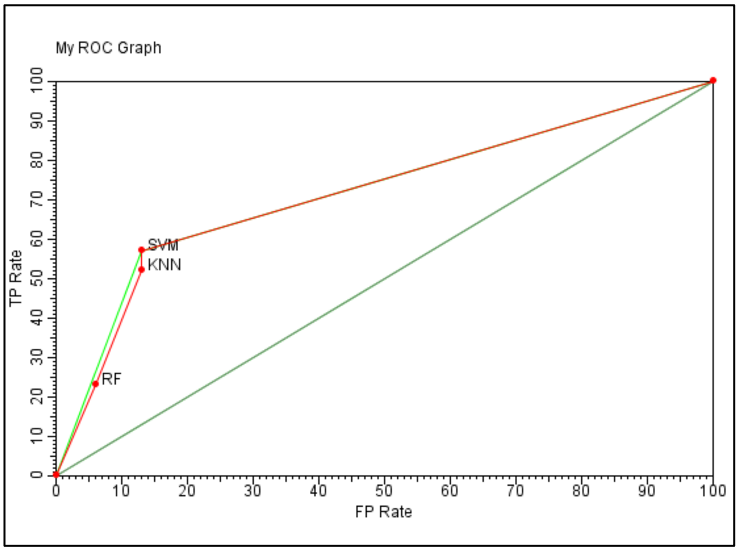


Figure 1-ROC curve (KNN, Random Forest and SVM)

Relevance/Contribution

As first main contribution, this paper aims to provide supply chain practitioners and researchers a new data mining approach regarding air freight delay management. The key element is to combine supplier, customer and transportation operational data to identify patterns which indicate higher probability of delay occurrence. Secondly, we propose a thorough data mining methodology that combines different sampling, attribute selection, dimensionality reduction and classification techniques. In short, this framework could be used as reference in future research or practical work not only in transportation but also on other supply chain risk management fields. Thirdly, the achievement of satisfactory results owns to the consistency of model and validation metrics application. Finally, this paper further expands current machine learning literature applied to aviation which has been mostly focused on weather, airport structure, flight schedule, ground delay and congestion explanatory attributes.

References

Chung, S.H., Ma, H.L., Hansen, M., Choi, T.M., (2020), "Data Science and analytics in aviation", *Transportation Research Part E: Logistics and Transportation Review*, Vol. 134.

Diana, T., (2018), "Can machines learn how to forecast taxi-out time? A comparison of predictive models applied to the case of Seattle/Tacoma International Airport", *Transportation Research Part E: Logistics and Transportation Review*, Vol. 119, pp. 149-164.

Etani, N., (2019), "Development of a predictive model for on-time arrival flight of airliner by discovering correlation between flight and weather data", *Journal of Big Data*, Vol. 6, No. 85.

Gui, G., Liu, F., Sun, J., Yang, J., Zhou, Z., Zhao, D., (2020). "Flight Delay Prediction Based on Aviation Big Data and Machine Learning", *IEEE Transactions on Vehicular Technology*, Vol. 69, No. 1, pp. 140-150.

Herrema, F., Curran, R., Hartjes, S., Ellejmi, M., Bancroft, S., Schultz, M., (2019),"A machine learning model to predict runway exit at Vienna Airport", *Transportation Research Part E: Logistics and Transportation Review*, Vol. 131, pp. 329-342.

Shafique, U, Qaiser, H., (2014), "A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA", *International Journal of Innovation and Scientific Research*, Vol. 12, No. 1, pp. 217-222.

Yu, B., Guo Z., Asian, S., Wang, H., Chen, G., (2019), "Flight delay prediction for commercial air transport: A deep learning approach". *Transportation Research Part E: Logistics and Transportation Review*, Vol. 125, pp. 203-221.