

UNICAMP

UNIVERSIDADE ESTADUAL DE
CAMPINAS

Instituto de Matemática, Estatística e
Computação Científica

LUIZ ALVES DE SOUZA NETO

**Modelos Matemáticos e Algoritmos
Computacionais para a Predição de Risco em
Saúde Materna**

Campinas

2022

Luiz Alves de Souza Neto

Modelos Matemáticos e Algoritmos Computacionais para a Predição de Risco em Saúde Materna

Tese apresentada ao Instituto de Matemática,
Estatística e Computação Científica da Uni-
versidade Estadual de Campinas como parte
dos requisitos exigidos para a obtenção do
título de Doutor em Matemática Aplicada.

Orientador: Cristiano Torezzan

Este trabalho corresponde à versão fi-
nal da Tese defendida pelo aluno Luiz
Alves de Souza Neto e orientada pelo
Prof. Dr. Cristiano Torezzan.

Campinas

2022

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

So89m Souza Neto, Luiz Alves de, 1986-
Modelos matemáticos e algoritmos computacionais para a predição de
risco em saúde materna / Luiz Alves de Souza Neto. – Campinas, SP : [s.n.],
2022.

Orientador: Cristiano Torezzan.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de
Matemática, Estatística e Computação Científica.

1. Saúde materna. 2. Aprendizado de máquina. 3. Predição de risco. 4.
Algoritmos computacionais. 5. Modelos matemáticos. I. Torezzan, Cristiano,
1976-. II. Universidade Estadual de Campinas. Instituto de Matemática,
Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Mathematical models and computational algorithms for risk
prediction in maternal health

Palavras-chave em inglês:

Maternal health

Machine learning

Risk prediction

Computer algorithms

Mathematical models

Área de concentração: Matemática Aplicada

Titulação: Doutor em Matemática Aplicada

Banca examinadora:

Cristiano Torezzan [Orientador]

Rodolfo de Carvalho Pacagnella

Ana Estela Antunes da Silva

Petra Maria Bartmeyer

André Kazuo Takahata

Data de defesa: 26-07-2022

Programa de Pós-Graduação: Matemática Aplicada

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0002-9835-0118>

- Currículo Lattes do autor: <http://lattes.cnpq.br/0493392198514817>

**Tese de Doutorado defendida em 26 de julho de 2022 e aprovada
pela banca examinadora composta pelos Profs. Drs.**

Prof(a). Dr(a). CRISTIANO TOREZZAN

Prof(a). Dr(a). RODOLFO DE CARVALHO PACAGNELLA

Prof(a). Dr(a). ANA ESTELA ANTUNES DA SILVA

Prof(a). Dr(a). PETRA MARIA BARTMEYER

Prof(a). Dr(a). ANDRÉ KAZUO TAKAHATA

A Ata da Defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-Graduação do Instituto de Matemática, Estatística e Computação Científica.

Agradecimentos

Agradeço a Deus pela vida, saúde e portas de oportunidades abertas em meu caminho. Sem Ele nada disso seria possível.

Semelhantemente a gestação de uma criança que depende uma cadeia de pessoas até a sua concepção, esta tese tem a presença de diversas pessoas, de forma direta ou indireta contribuíram para esse desfecho. A essas pessoas, o meu sentimento de imensa gratidão.

À minha esposa, Sildenice Melo, pelo companheirismo, apoio e força nos momentos mais difíceis dessa trajetória.

À minha mãe, Antonieta, e meu pai, Raimundo, pelos incentivos, cuidados e, principalmente, a educação, mesmo com dificuldades, nunca me deixaram faltar. À minha sogra, Aldenir e meu sogro, Silvestre que também os considero como meus pais.

A meu orientador, Cristiano Torezzan, um grande irmão, exemplo de professor, que sem as suas contribuições, talento e direcionamento essa tese não seria possível de ser desenvolvida.

A Universidade Federal do Maranhão por permitirem o desenvolvimento de todo esse projeto acadêmico através do afastamento concedido.

À Universidade Estadual de Campinas, ao Instituto de Matemática e à Pós-Graduação em Matemática Aplicada e à Faculdade de Ciências Aplicadas. Servidores e professores de excelência nessa instituição marcaram a minha vida.

Aos “mosqueteiros”, Fabiano, Gesivaldo e Tacildo, nos momentos de conversas no cafezinho. Grandes amigos que o IMECC me presenteou, obrigado por tudo.

Aos componentes da banca, pelas sugestões para tese e para trabalhos futuros.

Ao professor Rodolfo da medicina, que disponibilizou tempo em longas conversas, mesmo à distância, ofereceu vários subsídios para esta tese.

*“Não vos amoldeis às estruturas deste mundo,
mas transformai-vos pela renovação da mente,
a fim de distinguir qual é a vontade de Deus:
o que é bom, o que Lhe é agradável, o que é perfeito.
(Bíblia Sagrada, Romanos 12,2)*

Resumo

A elevada demanda por diagnósticos cada vez mais complexos na área médica tem motivado pesquisas em diversos ramos científicos, em particular, na área de modelagem matemática e computacional. Trabalhos recentes têm mostrado que tais métodos podem atingir precisão compatível e até superar especialistas em tarefas de diagnóstico médico. No entanto, mesmo com resultados empíricos promissores, tais soluções ainda enfrentam barreiras para serem utilizadas na prática. Parte desse impasse pode ser explicado pela dificuldade de interpretar resultados obtidos com algoritmos que, em geral, são projetados para realizar tarefas de classificação supervisionada. Além disso, tais soluções não incorporam de maneira adequada informações advindas de calculadoras, ou scores de risco que são comumente utilizadas na prática. Visando contribuir cientificamente nessa problemática, a presente tese de doutorado apresenta duas principais contribuições originais no âmbito metodológico: a primeira, é um método denominado Clinical Probability Outcome – C3PO, que permite a combinação de calculadoras de risco e modelos baseados em apoio à decisão multicritério, com diferentes algoritmos de aprendizado de máquina, cuja resposta final é convertida na forma de probabilidades. Uma segunda contribuição da tese é um Modelo Multilinear por Coalizões Multiplicativas que permite a obtenção de indicadores de risco por meio de atribuição de pesos para cada variável e para todas as coalizões par-a-par. Tal modelo é inspirado na Integral de Choquet, mas possui características mais gerais, que permitem melhor adequação para aplicações em diagnóstico médico. Adicionalmente, a tese também contempla dois estudos empíricos, baseados em dados reais anonimizados, relativos à predição de necessidade de leitos de UTI para pacientes gestantes e predição de diagnóstico de COVID-19 com base em sintomas. Tais estudos servem como validação dos métodos propostos e, também, como investigações paralelas sobre a viabilidade de aplicação de métodos de aprendizado de máquina para predições em saúde da mulher. Diversos métodos clássicos, como Florestas Aleatórias, XGboost, GBM, Redes Neurais Profundas, SVM, Regressão Logística, Naive Bayes, KNN foram utilizados, tanto individualmente quanto em comitês de máquinas. Os resultados atestam a viabilidade da proposta, com acurácias acima de 93% nas predições. Além disso, como os resultados são apresentados na forma de probabilidades, os limiares de decisão podem ser ajustados pelo médico, em função do nível de sensibilidade ou especificidade desejado em cada aplicação. Finalmente, estudos adicionais podem ter amparo nos resultados desta tese, tanto no âmbito da aplicação do C3PO para diagnósticos variados, quando na possibilidade de aprimoramentos metodológicos, como o uso de regularização no Modelo Multilinear por Coalizões Multiplicativas visando reduzir o número de pesos necessários para o modelo.

Palavras-chave: saúde; aprendizado de máquina; predição de risco; materna.

Abstract

The increase of demand for complex diagnoses in the medical field has motivated research in several scientific fields, in particular, in mathematical and computational modeling. Recent work has shown that such methods can achieve comparable accuracy and even outperform specialists in medical diagnostic tasks. However, even with promising empirical results, such solutions still face barriers to be used in practice. Part of this impasse can be explained by the difficulty of interpreting results obtained with algorithms that, in general, are designed to perform supervised classification tasks. In addition, such solutions may not adequately incorporate information from risk calculators that are commonly used in practice. Aiming to contribute scientifically to this problem, the present doctoral thesis presents two main original contributions in the methodological scope: the first, is a method called Clinical Probability Outcome - C3PO, which allows the combination of risk calculators and models based on multi-criteria decision support, with different machine learning algorithms, such that the final answer is converted in terms of probabilities. A second contribution of the thesis is a Multilinear Model by Multiplicative Coalitions that allows obtaining risk indicators by assigning weights to each variable and to all pairwise coalitions. This model is inspired by the Choquet Integral, but it has more general characteristics, which allow better adaptation for applications in medical diagnosis. Additionally, the thesis also contemplates two empirical studies, based on anonymized real data, related to the prediction of the need for ICU beds for pregnant patients and the prediction of diagnosis of COVID-19 based on symptoms. Such studies serve as a validation of the proposed methods and as parallel investigations on the feasibility of applying machine learning methods for predictions in women's health. Several classic methods such as Random Forests, XGBoost, GBM, Deep Neural Networks, SVM, Logistic Regression, Naive Bayes, KNN were used, both individually and in machine committees. The results attest to the feasibility of the proposal, with accuracies above 93% in the predictions. Furthermore, as the results are presented in the form of probabilities, the decision thresholds can be adjusted by the physician, depending on the level of sensitivity or specificity desired in each application. Finally, additional studies may be supported by the results of this thesis, both in the scope of the application of C3PO for different diagnoses, and in the possibility of methodological improvements, such as the use of regularization in the Multilinear Model by Multiplicative Coalitions to reduce the number of weights necessary for the model.

Keywords: health; machine learning; risk prediction; maternal.

Lista de ilustrações

Figura 1 – Espaço amostral Ω	22
Figura 2 – Fluxograma de treinamento para uma regressão logística	27
Figura 3 – Gráfico da função sigmoide	28
Figura 4 – Região de separação das densidades de probabilidades pelo <i>threshold</i>	32
Figura 5 – localização dos pontos	33
Figura 6 – Telhas de Voronoi	33
Figura 7 – Exemplo de probabilidade com KNN	34
Figura 8 – Modelo matemático de um neurônio artificial	35
Figura 9 – Exemplo de uma Rede neural perceptron em várias camadas MLP	36
Figura 10 – Exemplo de uma separação dos dados pelo hiperplano por SVM	39
Figura 11 – Hiperplano de separação, margem máxima e vetores de suporte	39
Figura 12 – Distância entre H1 e H2	41
Figura 13 – Transformação do núcleo $\phi : L \rightarrow H$	42
Figura 14 – Exemplo de uma árvore de decisão	44
Figura 15 – Comparação entre a distribuição de probabilidade e entropia	44
Figura 16 – Grafo para o sintoma Cefaleia	50
Figura 17 – Grafo para o sintoma de Coriza	51
Figura 18 – Grafo para o sintoma de Febre	51
Figura 19 – Grafo para o sintoma Calafrio	52
Figura 20 – Primeiro ramo da árvore de decisão construída	53
Figura 21 – Etapas de um sistema de múltiplos classificadores. [Adaptada de Cruz et al [1].]	54
Figura 22 – Exemplo de um combinador	55
Figura 23 – Etapas do Método Aditivo Linear	57
Figura 24 – Solução ideal para o método TOPSIS	58
Figura 25 – Representação gráfica para três sintomas do paciente P_1	62
Figura 26 – Estrutura do C3PO	68
Figura 27 – Esquema do Módulo 1 do C3PO.	69
Figura 28 – Esquema do Módulo 2 do C3PO.	69
Figura 29 – A matriz P	72
Figura 30 – Curva ROC com AUC para combinação dos melhores classificadores	81
Figura 31 – Correlação das saídas em probabilidades do método aditivo	83
Figura 32 – Matriz de correlação entre as probabilidades do método TOPSIS	85
Figura 33 – Distribuição de frequência de probabilidades	86
Figura 34 – Curva ROC para um preditor com TOPSIS	87
Figura 35 – Representação em bloxplot para o primeiro cenário	88

Figura 36 – Boxplot para F1-Score e para a representação do controle pelo limiar de decisão	89
Figura 37 – Representação em boxplot das métricas para o segundo cenário	89
Figura 38 – Boxplot para F1-Score e para a representação do controle pelo limiar de decisão no segundo cenário	90
Figura 39 – Disposição dos pesos calculados pelo modelo multilinear	91
Figura 40 – Matriz de correlação das saídas de teste	94
Figura 41 – Curva ROC para o comitê de máquina para predição de leitos em UTIs com AUC de 0.90	98
Figura 42 – Matriz de correlação das saídas de teste	99
Figura 43 – Distribuição de probabilidades para o método aditivo linear	100
Figura 44 – Matriz de correlação entre as probabilidades de saída	100
Figura 45 – Distribuição de probabilidades da REDE	101
Figura 46 – Curva ROC para o método TOSPSIS aplicado na base da REDE . . .	101
Figura 47 – Representação gráfica para o primeiro cenário	103
Figura 48 – Representação gráfica para o primeiro cenário	103
Figura 49 – Representação em boxplot das métricas para o segundo cenário	104
Figura 50 – Representação em boxplot das métricas para o segundo cenário	104
Figura 51 – Representação em boxplot das métricas para o segundo cenário	105
Figura 52 – Disposição dos pesos calculados pelo modelo multilinear	106

Lista de tabelas

Tabela 1 – Tabela de sintomas de pacientes com suspeita de COVID-19	25
Tabela 2 – Tabela de probabilidades condicionais	26
Tabela 3 – Exemplo de aplicação	26
Tabela 4 – Tabela de sintomas de pacientes com suspeita de COVID-19	48
Tabela 5 – Tabela de Risco de Cefafeia	50
Tabela 6 – Método aditivo	57
Tabela 7 – Método TOPSIS	59
Tabela 8 – Sintomas de pacientes	61
Tabela 9 – Sintomas de pacientes	74
Tabela 10 – Tabela de probabilidades gerada para o treino na base da REBRACO .	77
Tabela 11 – Tabela de probabilidades gerada para o teste na base da REBRACO .	77
Tabela 12 – Resultados dos valores de AUC para cinco divisões da base	77
Tabela 13 – Resultados dos valores de taxas de verdadeiros positivos das saídas individuais	78
Tabela 14 – Resultados dos valores de taxas de verdadeiros negativos da saída individuais	78
Tabela 15 – Resultados para AUC	79
Tabela 16 – Taxa de verdadeiro positivo para combinação par-a-par	79
Tabela 17 – Taxa de verdadeiro negativo para a combinação par-a-par	79
Tabela 18 – Resultado para as saídas individuais, controlando pelo <i>threshold</i> igual a 0.5	80
Tabela 19 – Tabela de scores de preferências	82
Tabela 20 – Os <i>scores</i> extraídos do processo multiplicativo	83
Tabela 21 – Probabilidades geradas pelo método SAW	83
Tabela 22 – Resultado para o método aditivo linear	84
Tabela 23 – Scores do método TOPSIS	85
Tabela 24 – Probabilidades dos scores por regressão logística	85
Tabela 25 – Resultados para o método TOPSIS	86
Tabela 26 – Probabilidades de treinamento	94
Tabela 27 – Probabilidades de teste	94
Tabela 28 – Resultados para as saídas individuais da REDE com regressão logística- AUC	95
Tabela 29 – Resultados para as saídas individuais da REDE com regressão logística- TPR	95
Tabela 30 – Resultados para as saídas individuais da REDE com regressão logística- TNR	95

Tabela 31 – Resultados para a AUC com combinação par-a-par por agregação logística.	96
Tabela 32 – Resultados para a TVP com combinação par-a-par	96
Tabela 33 – Resultados para a TVN com combinação par-a-par com agregação por regressão logística	97
Tabela 34 – Resultados para as métricas controlando apenas pelo limiar de 0.5 . . .	97
Tabela 35 – Probabilidades para o método aditivo linear	99
Tabela 36 – Resultado para o método aditivo linear	99
Tabela 37 – Resultados para o método TOPSIS	101

Sumário

	Introdução	15
1	REVISÃO DA LITERATURA	18
1.1	Estudos Relacionados	18
1.2	Métodos de Aprendizado de Máquina	21
1.2.1	Pensamento Bayesiano	21
1.2.2	Classificador Naive Bayes	23
1.2.3	Regressão Logística	27
1.2.3.1	Saídas em perspectiva probabilística	30
1.2.4	KNN- K-Vizinhos Mais Próximos	32
1.2.5	Rede Neural Multicamadas ou Multilayer Perceptron	35
1.2.5.1	Neurônio artificial	35
1.2.6	Support Vector Machine- SVM	38
1.2.7	Métodos Baseados em <i>Ensemble Learning</i>	42
1.2.7.1	Árvore de Decisão e Floresta Aleatória	43
1.2.7.2	Gradient Boosting Machine-GBM	53
1.2.7.3	Comitê de Classificadores	54
1.2.7.3.1	Combinador Linear	55
1.2.7.3.2	Pesos das Máquinas	56
1.3	Métodos de Apoio à Decisão Multicritérios	56
1.3.1	Método Aditivo Linear	56
1.3.2	Método TOPSIS	57
1.3.3	Integral de Choquet	59
1.3.4	Métricas de Desempenho	63
2	O MÉTODO C3PO	64
2.1	Fundamentação do Método C3PO	65
2.1.1	Dois Paradigmas de Apoio à Decisão Clínica	66
2.1.2	Estrutura do C3PO	67
2.2	Um Novo Método Interpretável: Método Multilinear por Coalizões Multiplicativas	70
2.2.1	Programação Quadrática para a Estimação de Pesos com Base em Dados em Abordagem Supervisionada	71
3	APLICAÇÃO DO C3PO: PREDIÇÃO DE COVID-19 EM MULHERES GESTANTES	75

3.1	Apresentação dos Resultados	76
3.2	Resultado para o Módulo 1	76
3.2.1	Resultados Individuais para cada saída do C3PO	77
3.2.2	Agregando todas as saídas	78
3.2.3	Agregando as Saídas em Pares	78
3.2.4	Controle Somente pelo Limar de Corte	80
3.2.5	Comitê de Máquina	80
3.3	Resultado para o Módulo 2	81
3.3.1	Método Aditivo Linear	82
3.3.2	Método TOPSIS	84
3.4	C3PO: Aplicação do Modelo Multilinear por Coalizão Multiplicativa	87
3.4.1	Resultados	87
3.4.2	Conclusões para Modelo Multilinear por Coalizões Multiplicativas (M2CM)	90
3.4.3	Conclusão	91
4	APLICAÇÃO DO C3PO: PREDIÇÃO DE DESFECHOS PARA MULHERES GESTANTES	92
4.1	Apresentação dos Resultados	93
4.1.1	Resultado para o Módulo 1	93
4.1.2	Resultados Individuais para cada saída do C3PO	94
4.1.3	Agregando todas as saídas	95
4.1.4	Agregando as saídas em pares	95
4.1.5	Controle somente pelo limiar	97
4.1.6	Comitê de Máquina	97
4.2	Resultado para o Módulo 2	98
4.2.1	Definição dos pesos dos critérios da base da REDE	98
4.2.2	Método Aditivo Linear	99
4.2.3	Método TOPSIS	100
4.3	C3PO: Aplicação do Modelo Multilinear por Coalizão Multiplicativa	102
4.3.1	Resultados	102
4.3.2	Conclusão	106
5	CONSIDERAÇÕES FINAIS E PERSPECTIVAS	108
	REFERÊNCIAS	110

Introdução

Modelos matemáticos e algoritmos computacionais vêm moldando os rumos de várias atividades humanas e uma das mais promissoras aplicações têm sido na área médica [2]. Tais ferramentas podem tanto fornecer suporte para assistência aos pacientes, quanto auxiliar em processos administrativos. Exemplos dessas aplicações incluem o uso de ferramentas de Inteligência Artificial (IA) para apoio a diagnóstico [3, 4, 5], recomendação de tratamentos [6, 7], microbiologia clínica [8] e atividades administrativas em saúde [2, 9, 10].

Entre as várias e amplas temáticas relevantes em saúde humana, estudos com foco na gravidez representam uma área ativa de pesquisa, com importantes desafios que impactam positivamente a qualidade de vida da população [11]. Um exemplo desses desafios é a redução da morbimortalidade materna, um problema complexo e multifacetado, que pode ser beneficiado pelo desenvolvimento de sistemas inteligentes de apoio à decisão [12] capazes de ampliar, em abrangência e em qualidade, a assistência remota à pacientes.

Em vista disso, um dos fatores mais relevantes para diminuição da morbidade materna é a redução no tempo para o atendimento às pacientes [13, 14, 15]. Nesse contexto, o investimento em estratégias de diagnóstico oportuno e em tratamento ágil pode reduzir a complicação de algumas doenças ligadas à morbimortalidade materna. Modelos matemáticos e algoritmos computacionais são ferramentas que podem auxiliar nessa problemática em automatização de diagnósticos e calculadoras de risco.

Além de técnicas de aprendizado de máquina, o uso de métodos multicritério de apoio à decisão também vem ganhando espaço na área de saúde. Por exemplo, Nardo [16], utiliza métodos multicritério para priorização de pacientes infectados por COVID-19. Em [17], tais técnicas são empregadas para a seleção de instrumentos e dispositivos médicos em ambientes de incerteza. Métodos de classificação multicritério também são utilizados para o planejamento de terapia de câncer de mama, conforme descrito em [18].

No entanto, apesar da vasta literatura sobre Métodos de Aprendizado de Máquina e sobre Apoio à Decisão Multicritério, ainda há poucos trabalhos que integram esses dois paradigmas, sobretudo em aplicações em saúde. Essa tese de doutorado busca contribuir com esse tema por meio da proposição de um modelo original, que integra soluções baseadas em protocolos clínicos, métodos multicritério e algoritmos de aprendizado de máquina.

A principal contribuição deste trabalho é a proposição do método denominado *Clinical Probability Outcome* (C3PO), que permite a combinação de calculadoras de risco, modelos baseados em apoio à decisão multicritério e diferentes algoritmos de aprendizado de

máquina para a estimação de riscos ou predição de diagnósticos. O C3PO foi desenvolvido sob perspectiva Bayesiana e seus resultados de saída são sempre expressos em termos de probabilidades, visando se aproximar da forma de raciocinar na prática médica [19, 20] e assim reduzir as barreiras de uso do método.

Além do C3PO, esta tese também traz outras três contribuições originais complementares:

- Um Modelo Multilinear por Coalizões Multiplicativas que permite a obtenção de indicadores de risco por meio de atribuição de pesos para cada variável e para todas as coalizões par-a-par;
- Um estudo empírico, utilizando o C3PO, para a predição de diagnóstico de COVID-19 com base em sintomas, com base em dados reais anonimizados de pacientes gestantes, fornecidos pelo projeto de pesquisa REBRACO, da Faculdade de Ciências Médicas da UNICAMP [21, 22, 23].
- Um estudo empírico, utilizando o C3PO, para prever a necessidade de leitos de UTI para pacientes gestantes, com base em dados reais anonimizados fornecidos pela Rede Brasileira de Morbidade Materna Grave, um projeto do Departamento de Tocoginecologia da Faculdade de Ciências Médicas da UNICAMP [24].

Os estudos de caso realizados mostraram que é possível obter acurácias acima de 93% para os problemas estudados, mesmo utilizando o limiar clássico de decisão de 50%. Como a saída do C3PO é dada em termos de probabilidades, é possível obter resultados mais sensíveis ou mais específicos por meio da alteração dos limiares de decisão. Por exemplo, para aplicações de triagem, pode-se reduzir o limiar de decisão, tornando o método mais sensível (ou seja, aumentando a chance do resultado ser interpretado como positivo). Para aplicações de diagnóstico, ou com recursos escassos, pode-se tornar o método mais específico, aumentando o limiar de decisão.

Os resultados empíricos mostraram ainda que o Modelo Multilinear proposto tem desempenho superior a métodos multicritérios clássicos, como SAW, TOPSIS e Integral de Choquet e se aproxima de alguns métodos de aprendizado de máquina, tendo a vantagem de ser facilmente interpretável.

As demais partes dessa tese estão organizadas da seguinte forma: no Capítulo 1, apresentamos uma breve revisão da literatura sobre alguns métodos orientados a dados de aprendizado de máquina utilizados na tese, e sobre os métodos de apoio à decisão multicritério. No Capítulo 2, introduzimos a fundamentação e a estrutura do Modelo C3PO-*Clinical Probability Outcome* e, por fim, a proposta do método multilinear por coalizões multiplicativa. Nos capítulos posteriores, são investigadas as métricas de desempenho de classificação do C3PO e do método Multilinear por Coalizões Multiplicativas. Os resultados

estão divididos da seguinte forma: Capítulo 3, apresentamos os resultados empíricos da aplicação do C3PO com o uso do Módulo 1 e Módulo 2 para diagnóstico de COVID-19 em gestantes por meio da base de dados da REBRACO (Rede Brasileira em Estudo do COVID-19 em Obstetrícia), e, logo em seguida, um estudo comparativo para o método Multilinear por Coalizões Multiplicativas. No Capítulo 4, a aplicação do C3PO no estudo de caso na predição ou recomendação do uso de leitos de UTIs, por gestantes, por meio da base de dados da Rede Brasileira de Vigilância de Morbidade Materna Grave, e, logo após, um estudo comparativo para o método Multilinear por Coalizões Multiplicativas. No Capítulo 5, apresentamos as considerações finais e propostas para estudos futuros, como o aprimoramento do método Multilinear por Coalizões Multiplicativas em diagnósticos médicos como o uso de regularização por técnicas como o LASSO visando reduzir o número de pesos necessários para o modelo e expressar as coalizões por índices de interação. A tese é finalizada com a lista de referências bibliográficas.

1 Revisão da literatura

Nossa revisão descreve técnicas de aprendizado de máquina e métodos de apoio à decisão multicritério aplicada em abordagem à saúde da mulher em todo o processo de gravidez encontradas em revistas científicas relevantes publicadas nos repositórios EMBASE, PubMed, SCOPUS (Elsevier), Springer, IEEE Xplore e Google Scholar: questões relacionadas com termos relevantes em aprendizado de máquina, inteligência artificial, tomadas de decisão multicritérios e saúde da gestante como questões relacionadas a predição em leitos de UTIs e COVID-19.

1.1 Estudos Relacionados

A saúde da mulher gestante está diretamente relacionada com a do recém-nascido em que é um tema de extrema relevância para a população [25]. Estudos relacionados ao uso de modelos matemáticos e algoritmos computacionais podem auxiliar na predição de fatores de riscos e auxiliar médicos e gestores de saúde na tomada de decisões [26, 27, 28].

Fatores de riscos na gestação são componentes que revelam sobre a situação materna relacionada com o uso de leitos em UTI, prematuridade, peso ao nascer e atualmente a COVID-19 em gestantes e outros fatores [29].

Esses métodos já se mostraram bastante eficientes em outras áreas, como em bioquímica, na procura de novas combinações de materiais e farmacológicos na busca de novos medicamentos, na agropecuária e agricultura, na clínica de imagens médicas e vários outros exemplos [30, 31, 32, 33, 34, 35, 36].

Apesar de avanços significativos no uso de técnicas computacionais na saúde, há uma falta de pesquisas que apliquem os métodos de aprendizado de máquina para entender como os tratamentos afetam a gravidez na sua evolução e no desfecho materno e natalício [11, 12]. Dessa maneira, a maior lacuna na literatura que identificamos diz respeito ao uso de métodos de aprendizado de máquinas para otimizar estudos em análises de regressão e classificação de riscos para as gestantes e recém-nascidos, principalmente no contexto do SUS.

Vários estudos foram sistematizados e catalogados em relação a cenários vinculados à saúde da mulher, de forma a fazer inferências estatísticas médicas a descrever variáveis de morbidades de forma específica [37, 38]. Além disso, a OMS- Organização Mundial da Saúde e outras agências usam essa informação em revisões sistemáticas, o que pode levar a uma melhor visão geral das necessidades de cuidado à saúde materna e a

orientações relacionadas [39].

Dessa forma, de acordo com Ding [40], a previsão de mortalidade na UTI pode refletir a gravidade da doença ou o prognóstico dos pacientes, obter uma alocação mais razoável de recursos médicos e ajudar os médicos a tomarem decisões. Assim, prever e prevenir certos tipos de doenças são preocupações nas pesquisas que envolvem algoritmos de predições na área da saúde.

O tratamento de variáveis torna-se importante nas etapas que antecedem os ajustes da fase de treinamento de um preditor, como em estudo publicado por Sufriyana [41], em que o número de variáveis reduzidas alcançaram uma área sob a curva (AUC) de 88% na predição de pré-eclâmpsia e mulheres grávidas normotensivas. Dessa forma, o cuidado no tratamento dos dados é vital para a aplicação de algoritmos de aprendizado de máquina, como a seleção das melhores variáveis e a identificação de dados faltantes. À vista disso, diferentes técnicas de aprendizado de máquina podem mostrar diferentes sensibilidades para a seleção de seus valores de parâmetro [42].

Na mesma linha de saúde da mulher gestante, o estudo realizado por Venkatesh [43] teve como objetivo prever o risco de hemorragia no pós-parto, antes da admissão do trabalho de parto, usando aprendizado de máquina e modelos estatísticos com base nos dados de partos que ocorreram nos Estados Unidos entre (2002–2008). O estudo explorou técnicas como regressão logística, florestas aleatórias e XGBoost, e mediu o desempenho dos modelos pela estatística C (isto é, índice de concordância), calibração e curvas de decisão e validação cruzada. Os modelos foram construídos na primeira fase de dados (2002–2006) e validados externamente (ou seja, temporalmente) na segunda fase (2007–2008). Dentre os modelos analisados, aqueles que utilizam gradiente descendente extremo, como o XGBoost, apresentaram melhores resultados atingindo até 80% de acurácia.

Dentro de pesquisas que envolvem admissão de pacientes em UTIs, podemos atentar na pesquisa de Moreira [44], na qual foi realizada a aplicação de uma rede neural de multicamadas (MLP) com 100 parturientes diagnosticadas com hipertensão com objetivo de prever se o parto ocorrerá antes ou após 34 semanas de gestação. Também foi possível prever o tipo de parto, previsão do parto, o resultado para o recém-nascido se será pequeno para a idade gestacional e a admissão em UTI. Nessa pesquisa é proposta a utilização de uma técnica heurística de inspiração biológica, conhecida como otimização por enxame de partículas (OEP), para reduzir o custo computacional do método baseado em RNA denominado perceptron multicamadas (MLP), sem reduzir sua taxa de precisão. O estudo apresentou uma precisão de 0.930, *recall* de 0.920, medida *F-score* de 0.925 e *AUC* de 0.958.

Em [45] foi aplicado modelos preditivos para classificação binária dos natimortos, usando vários classificadores de aprendizado de máquina: regressão logística regularizada, árvores de decisão baseadas em árvores de classificação e regressão, floresta aleatória,

XGBoost e uma rede neural perceptron multicamadas e validação cruzada. Os resultados encontrados, para todos os modelos, são as AUCs variando de 0.59 (intervalo de confiança-IC 95%: 0.58, 0.60) a 0.84 (IC 95%: 0.83, 0.85) entre as taxas de falsos positivos e verdadeiros positivos.

No estudo de [45], uma das conclusões das abordagens foi que grande parte das complicações são potencialmente previsíveis e identificadas no pré-natal com base em uma combinação de variáveis como: anomalias congênitas, características maternas e histórico médico, mas que a maior sensibilidade é alcançada usando complicações atuais da paciente na gravidez.

Ferramentas ou aplicativos assistenciais de e-Saúde, que podem ser implementadas e utilizadas por gestantes e profissionais. Conforme estudo publicado no artigo por Akbulut [46], em que foram analisadas nove técnicas de aprendizado de máquinas e testadas em 96 mulheres para prever o *status* de anomalia fetal com base nos dados maternos e clínicos, obtiveram um resultado utilizando florestas aleatórias com precisão de 89.5%.

Novas variáveis que não são comuns, no âmbito da saúde da mulher, podem surgir, como as doenças que afetam em seu pré-natal e na saúde do feto, como é o caso do Zika-Vírus e COVID-19. Em cima dessa problemática, vários estudos vêm sendo elaborados pelo mundo atualmente, [47, 48, 49] são alguns exemplos que levantam a necessidade de abordagem em novas metodologias preditivas. Dessa maneira, no artigo de Asma Khalil [50] foi estudada a alteração na incidência de natimortos e parto prematuro durante a pandemia de COVID-19, com resultado de incidência de natimortos significativamente maior durante o período pandêmico ($n = 16$ [9.31 por 1000]) nascimentos comparando nascimentos em *St George's University Hospital*, Londres.

A abordagem envolvendo COVID-19 e aprendizado de máquina para a predição de casos futuros de problemas de saúde em mulheres ou na criança ainda não chegaram a resultados e discussões adequadas, requerendo ainda um pouco mais de tempo para monitorar as evidências que, por ocasião, irão surgir e fornecerão subsídios para estudo.

Para complementação de estudos concernentes à saúde da mulher desta tese, é importante buscar outros fatores que estão relacionados com o tema atualmente em grandes pesquisas no mundo, como o mostrado em artigo publicado por Gaoa [51] fez o uso de aprendizagem profunda para predição de nascimento prematuro extremo (NPE) antes da 28ª semanas de idade gestacional. A pesquisa utilizou registros eletrônicos de saúde por meio dos modelos RNN(Rede Neural Recorrente) *ensemble* treinados em dados balanceados artificialmente, alcançaram uma maior AUC (0.827) e sensibilidade (0.965).

Nesse estudo de Gaoa, [51] foram identificados diversos fatores preponderantes, como variáveis preditoras, gravidez gemelar, comprimento cervical curto, distúrbio hiper-

tensivo, lúpus eritematoso sistêmico e sulfato de hidroxicloroquina os quais auxiliam como fatores de risco em níveis significativos que influenciam diretamente no NPE, portanto, percebe-se que o reconhecimento de atributos ou características graves na saúde da mulher gestante através do acompanhamento médico contribui para que o algoritmo forneça uma boa performance de predição.

1.2 Métodos de Aprendizado de Máquina

Aprendizado de máquina é uma área da inteligência artificial que estuda algoritmos, métodos e técnicas como ferramentas computacionais relacionadas à otimização de um critério de desempenho e adquirindo novos conhecimentos através de experiências e observações extraídas dos dados disponibilizados, utilizando-se de teorias da estatística na construção de modelos matemáticos, fazendo-se inferências a partir de uma amostra [52]. Dessa forma, neste capítulo veremos um breve estudo de alguns modelos de algoritmos matemáticos.

1.2.1 Pensamento Bayesiano

O teorema de Bayes recebe este nome devido ao pastor e matemático inglês Thomas Bayes (1701 – 1761), que destacou na sua formulação que a probabilidade é uma opinião metódica e a inferência de dados nada mais é que a revisão de tal opinião com novas informações relevantes [53].

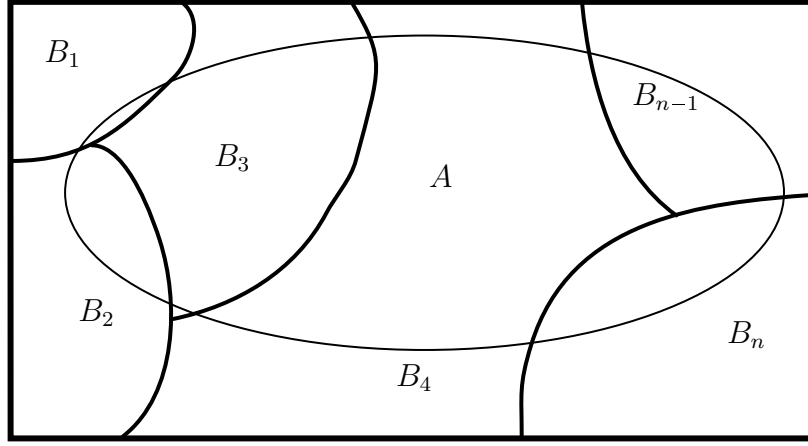
Uma premissa básica do pensamento bayesiano é que os atributos estão entrelaçados através de uma probabilidade. A abordagem do nosso trabalho levará em consideração a distribuição condicional de atributos alvo, dado os atributos de entradas.

A teoria de decisão bayesiana é uma abordagem estatística que faz parte dos fundamentos de problemas de classificação. As decisões, baseadas em probabilidades bayesianas para a classificação, assumem que o problema é colocado em termos de probabilidades conhecidas.

Eles podem prever as probabilidades de associação de classe, como a probabilidade de uma determinada combinação de eventos pertencer a uma classe específica.

Demonstração

Seja o espaço amostral Ω , podemos ilustrar como segue na Figura 1. O espaço amostral é o conjunto de todas as possíveis saídas de um experimento. O espaço amostral é particionado em subconjuntos de forma que esses não tem interseção entre si. E dentro desse espaço, vamos ter um evento de interesse A [54].

Figura 1 – Espaço amostral Ω

Definição 1. Sejam B_1, B_2, \dots, B_n subconjuntos formam uma partição do espaço amostral se:

- $B_i \cap B_j = \emptyset, i \neq j$;
- $\bigcup_{i=1}^n B_i = \Omega$;
- $P(B_i) \geq 0, i = 1, \dots, n$.

O \emptyset é o elemento nulo das operações de conjunto e sendo que $A \cap \emptyset = \emptyset$ e $A \cup \emptyset = A$, temos o seguinte:

- $A = (A \cap B_1) \cup (A \cap B_2) \cup \dots \cup (A \cap B_n)$;
- $A = \bigcup_{i=1}^n (A \cap B_i) (I)$.

A probabilidade condicional é dada por: $P(A|B) = \frac{P(A \cap B)}{P(B)}$, daí temos: $P(A \cap B) = P(A|B) P(B) (I)$ ou $P(B|A) = \frac{P(A \cap B)}{P(A)}$, daí temos: $P(A \cap B) = P(B|A) P(A) (II)$

Aplicando a probabilidade da união de conjuntos em (I), temos: $P(A) = P(\bigcup_{i=1}^n (A \cap B_i)) = \sum_{i=1}^n P(A \cap B_i) = \sum_{i=1}^n P(A|B_i) P(B_i)$. Então generalizando para todo B_i , temos:

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A|B_i) P(B_i)}{\sum_{j=1}^n P(A|B_j) P(B_j)} \quad (1.1)$$

O paradigma bayesiano não apenas oferece uma abordagem sólida para a quantificação da incerteza em modelos de aprendizado, mas também fornece uma estrutura matemática para entender muitas técnicas de regularização e estratégias probabilísticas de aprendizado que já são usadas no aprendizado clássico.

1.2.2 Classificador Naive Bayes

A partir do teorema de Bayes, podemos extrair resumidamente, para efeitos de classificação, a seguinte fórmula que estrutura a base de classificação em termos de probabilidade [55]:

$$P(H|data) = \frac{P(data|H) P(H)}{P(data)} \quad (1.2)$$

Onde os termos da fórmula de classificação podem ser dados como, $P(H|data)$, chamada de posteriori, que é a probabilidade que a hipótese é verdadeira dado o conjunto de dados coletados.

O vetor $data$ é uma coleção de medidas que contém atributos $[x_1, x_2, \dots, x_n]$, e H um conjunto que contém os elementos de uma classe alvo com $[y_1, \dots, y_m]$.

$P(data|H)$, chamada verossimilhança, sendo a probabilidade de coletar ou gerar os dados quando a hipótese é verdadeira.

$P(H)$, chamada priori, termo esse que designa a probabilidade da hipótese de ser verdadeira antes de coletar os dados.

$P(data)$, chamada de evidência ou probabilidade marginal, termo que se refere a probabilidade de coletar os dados sob todas as hipóteses ou classes.

A regra de decisão para a classificação pode ser dada conforme é descrito logo abaixo:

$$\begin{cases} Se & P(Y_i|X) > P(Y_j|X), & \text{classifique } x \in Y_i \\ Se & P(Y_i|X) < P(Y_j|X), & \text{classifique } x \in Y_j \end{cases}$$

Em termos práticos, o teorema de Bayes é uma formulação matemática do senso comum, ou seja, classifica segundo a classe mais provável. O classificador probabilístico Naive Bayes assume que os atributos das variáveis X são independentes.

Ao adotar a regra de decisão bayesiana o erro obtido na classificação é o mínimo possível, ou seja, essa teoria mostra que o classificador bayesiano é ótimo com relação à minimização da chance de erro [56].

Embora, na teoria, a decisão bayesiana ofereça o menor erro possível de classificação, na prática, possui uma grande limitação, que é o fato de não conhecermos essa distribuição $P(Y_i|X)$, a distribuição de probabilidade conjunta. Essa probabilidade define o modelo gerador dos dados, então, inferimos que essa probabilidade e essa inferência depende do método que usamos e da qualidade e da quantidade de dados que temos.

Para definir $P(Y_i|X)$, calculamos a verossimilhança $P(X|Y_i)$ e $P(Y_i)$, sendo a distribuição a priori e $P(X)$ é a evidência ou termo normalizador, então, assim, estimamos a função de verossimilhança.

Definição 2. Seja X um vetor de características de variáveis aleatórias e Y um conjunto de classes Y_1, \dots, Y_n , podemos dizer que cada transformação da função geradora pode realizar n ações ao invés de apenas classificar os X_1, X_2, \dots, X_n . Para cada ação X_i há um custo associado para cada Y_j . Para tomar uma decisão existe um custo que podemos representar por $\lambda(\alpha_i|Y_j)$, é um valor dado.

A perda esperada ou risco de se tomar uma ação decisória α_i ao observar X é dado por: $R(\alpha_i|X) = \sum_{j=1}^n \lambda(\alpha_i|Y_j) P(Y_j|X)$ é chamada de risco condicional ou risco médio de uma ação α_i . Temos que $\alpha(X)$ é uma função que seleciona α_i com base em X [57].

O risco total é considerado como a integral do risco médio, em termos gerais, podemos expressar através da seguinte formulação, $R = \int R(\alpha(X)|X) P(X) dx$, dessa forma, podemos pensar em uma maneira de maximizar o risco total.

O $\alpha(X)$ deve sempre escolher a ação α_i que tem menor risco condicional $R(\alpha_i|X)$, o que recai na regra de decisão bayesiana. Esta regra para o cálculo de R mínimo é o risco de Bayes que fornece o melhor resultado possível. Dessa forma, para uma classificação binária, temos o seguinte, $\lambda_{ij} = \lambda(\alpha_i|Y_j)$, α_i = classificar como Y_i .

$$R(\alpha_1|X) = \lambda_{11}P(Y_1|X) + \lambda_{12}P(Y_2|X)$$

$$R(\alpha_2|X) = \lambda_{21}P(Y_1|X) + \lambda_{22}P(Y_2|X)$$

Em que λ_{ij} é o custo em classificar um objeto da classe i como sendo verdadeiro da classe j , em que $R(\alpha_1|X)$ o risco de classificar sendo da classe ou o risco condicional esperado.

Quando tomamos a ação α_1 (isto é, classificamos como pertencente à classe Y_1) se $R(\alpha_2|X) > R(\alpha_1|X)$.

Podemos reescrever essa desigualdade como:

$$\lambda_{21}P(Y_1|X) + \lambda_{22}P(Y_2|X) > \lambda_{11}P(Y_1|X) + \lambda_{12}P(Y_2|X)$$

$$(\lambda_{21} - \lambda_{11})P(Y_1|X) > (\lambda_{12} - \lambda_{22})P(Y_2|X)$$

$$(\lambda_{21} - \lambda_{11})P(X|Y_1)P(Y_1) > (\lambda_{12} - \lambda_{22})P(X|Y_2)P(Y_2)$$

$$\frac{P(X|Y_1)}{P(X|Y_2)} > \frac{(\lambda_{12} - \lambda_{22})P(Y_2)}{(\lambda_{21} - \lambda_{11})P(Y_1)} = \theta,$$

se $\lambda_{12} - \lambda_{22}$ e $\lambda_{21} - \lambda_{11}$ são positivos. Uma alternativa a decisão seria Y_1 , que segue imediatamente sob a suposição razoável de que $\lambda_{21} > \lambda_{11}$.

O valor de theta θ chamamos de *threshold* ou *Likelihood ratio* (razão de verossimilhança). Esta forma de regra de decisão concentra-se na dependência de X das densidades de probabilidade. Assim, a regra de decisão de Bayes pode ser interpretada como uma decisão sobre Y_1 se a razão de verossimilhança exceder um valor limite que é independente da observação X .

Se a ação α_i for realizada e o verdadeiro estado de natureza for Y_j , então a decisão é correta se $i = j$, é errônea se $i \neq j$.

Em casos de modelos com várias classes de saída, podemos inferir os mesmos argumentos, dessa forma, temos que o risco condicional médio da seguinte maneira:

$$R(\alpha_i|X) = \sum_{j=1}^n \lambda(\alpha_i|Y_j) P(Y_j|X) = \sum_{j \neq i} P(Y_j|X) = 1 - P(Y_i|X). \quad (1.3)$$

Dessa forma, a tomada de decisão em relação ao que está expresso acima é relacionado a quando se decide por Y_i se $P(Y_i|X) > P(Y_j|X)$ para todo $j \neq i$. Se a função de perda é a zero-um ou escada, isso se resume a escolher a classe com maior posteriori.

Vimos, então, como a teoria da probabilidade nos fornece uma estrutura matemática consistente para quantificar e manipular a incerteza. A teoria da probabilidade permite tomar decisões ideais em situações que envolvem incerteza, como aquelas encontradas no reconhecimento de padrões.

O exemplo abaixo, dado pela Tabela 1, exemplifica a tomada de decisão através do classificador Naive Bayes. Temos um conjunto de atributos de amostras retiradas de uma tabela amostrada de um hospital, em que são representados pacientes, em níveis diferentes de COVID-19.

Tabela 1 – Tabela de sintomas de pacientes com suspeita de COVID-19

Calafrio	Coriza	Febre	Cefaleia	Risco COVID-19
Baixo	Sim	Não	Baixa	Alto
Médio	Sim	Não	Média	Alto
Médio	Não	Não	Média	Moderado
Médio	Não	Não	Forte	Alto
Médio	Não	Não	Forte	Baixo
Médio	Não	Sim	Forte	Baixo
Baixo	Não	Não	Baixa	Alto
Baixo	Não	Sim	Forte	Moderado
Alto	Não	Não	Forte	Baixo
Alto	Sim	Sim	Forte	Baixo
Alto	Sim	Não	Baixa	Alto
Alto	Sim	Não	Média	Moderado
Alto	Sim	Não	Forte	Baixo
Baixo	Sim	Não	Média	Alto

Dessa maneira, podemos calcular as probabilidades de verossimilhanças, conforme a Tabela 1, levando em consideração que as probabilidades, a priori, podem ser dadas diretamente com vista no número de ocorrência em cada atributo.

Com a tabela de probabilidades já construída como mostrado na Tabela 2, podemos realizar uma estimativa em relação a um caso novo e avaliar o paciente através

Tabela 2 – Tabela de probabilidades condicionais

Ri. COVID-19	Calafrio			Coriza		Febre		Cefaleia		
	Alto	Médio	Baixo	Sim	Não	Não	Sim	Baixa	Média	Alta
Alto	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{2}{6}$	$\frac{6}{6}$	0	$\frac{3}{6}$	$\frac{2}{6}$	$\frac{1}{6}$
Moderado	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	0	$\frac{2}{6}$	$\frac{1}{3}$
Baixo	$\frac{3}{5}$	$\frac{2}{5}$	0	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{2}{5}$	0	0	$\frac{5}{5}$

de um classificador *Naive Bayes* e estimar a probabilidade de pertencer a alguma classe de Risco de COVID-19.

Tabela 3 – Exemplo de aplicação

Paciente	
Sintomas	Grau dos sintomas
Calafrio	Alto
Coriza	Sim
Febre	Não
Cefaleia	Forte

Primeiro, precisamos calcular as probabilidades de verossimilhanças:

$$P(Alto|X) = \frac{6}{14} \frac{1}{6} \frac{4}{6} \frac{6}{6} \frac{1}{6} = 0.0079$$

$$P(Moderado|X) = \frac{3}{14} \frac{1}{3} \frac{1}{3} \frac{2}{3} \frac{1}{3} = 0.0052$$

$$P(Baixo|X) = \frac{5}{14} \frac{3}{5} \frac{2}{5} \frac{3}{5} \frac{5}{5} = 0.0514$$

Para o cálculo da evidência, como segue:

$$P(X) = 0.0079 + 0.0052 + 0.0514 = 0.0645$$

$$P(Alto|X) = \frac{0.0079}{0.0645} 100 = 12.24\%$$

$$P(Moderado|X) = \frac{0.0052}{0.0645} 100 = 8.06\%$$

$$P(Baixo|X) = \frac{0.0514}{0.0645} 100 = \mathbf{79.68\%}$$

Portanto, podemos dizer que a probabilidade do paciente avaliado esteja com COVID-19 é baixa através do método Naive Bayes, pois possui uma probabilidade de 79.69% de pertencer à classe baixa.

1.2.3 Regressão Logística

Iremos apresentar um método de aprendizado de máquina que possui grande aceitabilidade em relação aos resultados em vias de interpretabilidade. Dessa forma, é possível estabelecer um classificador linear a partir de uma abordagem baseada em regressão linear, segundo, Schein et al [58], os valores de saída dos classificadores são tomados, por meio de uma regra de decisão estabelecida entre um parâmetro de comparação. Uma maneira alternativa de obter a saída ao qual se pretende ter na tarefa de classificação seria ajustar os pesos considerando as saídas da regra de decisão \hat{d}_i [59].

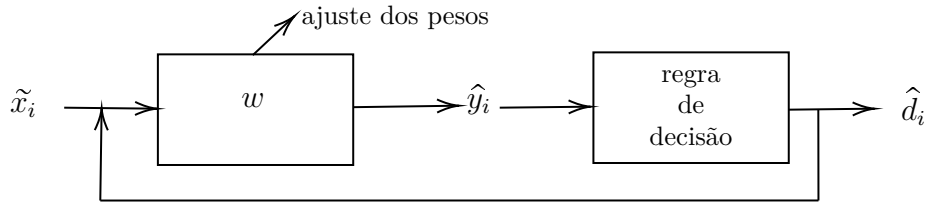


Figura 2 – Fluxograma de treinamento para uma regressão logística

Pela figura 2, temos que: \tilde{x}_i é o vetor de entrada de uma amostra e:

$$\hat{d}_i = \begin{cases} 1, & \text{classe 1} \\ 0, & \text{classe 2} \end{cases}$$

Portanto, o critério de ajuste do vetor de peso w é uma regra de decisão do tipo função degrau, onde h é uma função de decisão.

$$\hat{d}_i = h(\hat{y}_i) = \begin{cases} 1, & \hat{y} > 0.5 \\ 0, & \hat{y} < 0.5 \end{cases} \quad (1.4)$$

A abordagem baseada em quadrados mínimos pode ser dada da seguinte maneira:

Seja, $J(w) = \|d - \hat{d}\|_2^2 = \sum_{i=1}^N (d_i - \hat{d}_i)^2$ uma função objetivo de custo, onde \hat{d}_i , o vetor a ser estimado neste critério de aprendizagem, sendo N o número de amostras, temos o seguinte:

Dado na Equação 1.4, o vetor estimado \hat{y} pode ser calculado como $\hat{y}_i = w^t \tilde{x}_i$, assim substituímos na função de custo J , temos:

$$J(w) = \sum_{i=1}^N (d_i - h(\hat{y}_i))^2 = \sum_{i=1}^N (d_i - h(w^t \tilde{x}_i))^2 \quad (1.5)$$

Dessa forma, temos um problema de otimização resultante $\min_w J(w) = \sum_{i=1}^N (d_i - h(w^t \tilde{x}_i))$. Essa função apresenta algumas consequências indesejáveis, pois a função degrau

torna o problema de otimização difícil de ser resolvida, ou seja, não possui solução analítica. Uma forma de contornar esse obstáculo é buscar uma outra função que possa substituir em termos de existência de uma solução, assim, pode-se utilizar a função logística sigmoide padrão, como a que segue:

$$S(y) = \frac{1}{1 + e^{-y}} \quad (1.6)$$

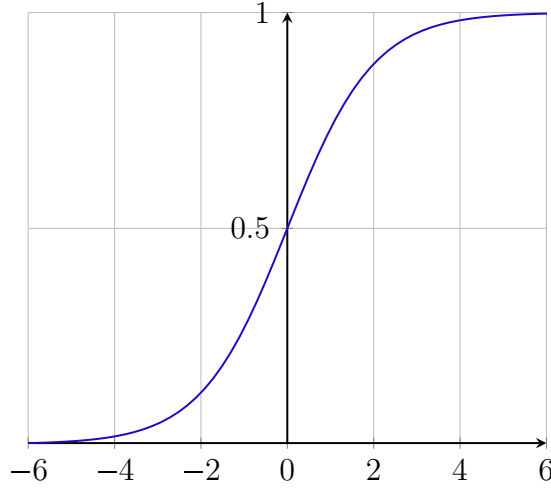


Figura 3 – Gráfico da função sigmoide

Como S é uma função diferenciável, podemos ter a seguinte derivada $\frac{dS(y)}{dy} = S(y)(1 - S(y)) = \frac{e^y}{(1 + e^y)^2}$.

Uma regressão logística é um mapeamento de espaço de atributos que é um rótulo não-linearmente separável, então, uma classificação zero ou um.

Daí, pode-se ter $J(w) = \sum_{i=1}^N (d_i - S(\hat{y}))^2$, e expandindo a expressão, temos,

$$J(w) = \sum_{i=1}^N (d_i - S(\hat{y}))^2 = \sum_{i=1}^N \left(d_i - \frac{1}{1 + e^{-w^t \hat{x}}} \right)^2.$$

Portanto, o $J(w)$ é uma função não-convexa, dessa maneira não possui uma derivada para algum ponto pertencente a um dado intervalo.

Para resolver esse tipo de problema é utilizado o gradiente descendente, que é um método iterativo. Neste método, o vetor w é o valor a ser ajustado e onde temos que μ é o passo de otimização da função custo J no ponto $w^{(i-1)}$. A derivada parcial de J é dado por:

$$w^{(i)} = w^{(i-1)} - \mu \nabla J(w^{(i-1)}) \quad (1.7)$$

$$\nabla J(w) = \begin{bmatrix} \frac{dJ(w)}{d(w_0)} \\ \frac{dJ(w)}{d(w_1)} \\ \frac{dJ(w)}{d(w_2)} \\ \vdots \\ \frac{dJ(w)}{d(w_N)} \end{bmatrix} \quad (1.8)$$

Com base nisso, é preciso desenvolver a função de perda para explorar a saída \hat{d}_i com valores $(0, 1)$, na função de custo, sendo medido da seguinte forma para cada amostra da matriz:

$$j_i(w) = -d_i \log(\hat{d}_i) - (1 - d_i) \log(1 - \hat{d}_i) \quad (1.9)$$

Nesta função de custo para cada amostra i em $j_i(w)$ temos:

- $d_i = 0$ e $\hat{d}_i \rightarrow 0 \rightarrow j_i(w) \rightarrow 0$
- $d_i = 1$ e $\hat{d}_i \rightarrow 1 \rightarrow j_i(w) \rightarrow 0$
- Um erro de classificação ocorrem nos casos
- $d_i = 0$ e $\hat{d}_i \rightarrow 1 \rightarrow j_i(w) \rightarrow \infty$
- $d_i = 1$ e $\hat{d}_i \rightarrow 0 \rightarrow j_i(w) \rightarrow \infty$

A função apresentada em 1.9 é chamada de detector de erro de classificação. Se não há erro de classificação, a função atribui valor zero para $j_i(w)$, caso contrário, a função tende ao valor infinito com aproximação assintótica quando na etapa de treinamento acontece algum erro para o valor estimado de predição. Portanto, pode-se dispor o cálculo do critério de treinamento como a média da soma da função de custo J .

$$J(w) = \frac{1}{N} \sum_{i=1}^N j_i(w) \quad (1.10)$$

A função objetivo 1.10 é conhecida como entropia cruzada. Desta maneira, o aprendizado baseado em minimizar a entropia cruzada é expressa por: $\min_w J(w) = \frac{1}{N} \sum_{i=1}^N (-d_i \log(\hat{d}_i) - (1 - d_i) \log(1 - \hat{d}_i))$ e $\hat{d}_i = S(\hat{y}) = \frac{1}{1 + e^{-w^t \tilde{x}_i}}$. Dessa forma, para achar o w mínimo é realizado o método iterativo do gradiente descendente $w^{(i)} = w^{(i-1)} - \mu \nabla J(w^{(i-1)})$ e a função objetivo de custo fica expressa por:

$$\nabla J(w) = - \sum_{i=1}^N ((d_i - S(w^t \tilde{x}_i)) \tilde{x}_i) \quad (1.11)$$

Para evitar *overfitting* é importante abordar a regularização dos pesos ajustados, conforme em [60]. Às duas formas mais simples são a regularização $L1$ e $L2$. Essas regularizações penalizam grandes pesos, mas as formas como atuam diferem, portanto, é uma maneira de controle na fase de treinamento, em que podem aumentar ou diminuir o grau de flexibilidade da função no ajuste dos pesos [61, 27, 60].

Nesta abordagem, modificamos a função de custo não regularizada, adicionando um termo extra ao regularizador dos pesos $\min_w J(w, X, d) + \lambda R(w)$, onde $\lambda \geq 0$, é o parâmetro de regularização.

Na regularização via norma $L2$ é realizada como segue:

$$R_{L2}(w) = \|w\|_2^2 = w_0^2 + w_1^2 + \dots + w_M^2 \quad (1.12)$$

Na regularização via norma $L1$ é realizada como segue:

$$R_{L1}(w) = \|w\|_1 = |w|_0 + |w|_1 + \dots + |w|_M \quad (1.13)$$

A regularização $L1$ tende a concentrar o peso da rede em um número relativamente pequeno de conexões de alta importância, enquanto os outros pesos são direcionados para zero, e, conseqüentemente, uma propriedade importante dessa norma é levar a função custo em vetores w a qual os elementos na maioria são zeros, ou seja, vetores esparsos [60].

1.2.3.1 Saídas em perspectiva probabilística

Em um modelo matemático de regressão linear, para prever a variável de saída Y , pode ser calculada como uma combinação linear dos atributos de entrada X em $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n = \beta^t X$. Onde $X^t = [1, X_1, \dots, X_n]$ e β o vetor de coeficientes. Como ajustar os coeficientes β para encontrar uma curva que minimize a função custo, ou seja, fazer um ajuste com erro mínimo.

Para classificação, podemos ter geralmente probabilidades posteriores que se encontram no intervalo $(0, 1)$. Vamos definir as probabilidades para o caso de duas classes em $P(y = 0|X)$ e $P(y = 1|X)$.

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \text{ e } X = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}.$$

Podemos considerar a regressão logística h como uma função de distribuição de probabilidades como chances de um evento acontecer.

$$h(x) = \frac{e^x}{1 + e^x}.$$

A probabilidade de pertencer à classe 1:

$$P(y = 1|X) = \frac{e^{\beta^t x}}{1 + e^{\beta^t x}}.$$

A probabilidade de pertencer à classe 0, ou seja, a probabilidade complementar.

$$P(y = 0|X) = 1 - P(y = 1|X) = \frac{1}{1 + e^{\beta^t x}}.$$

Fazendo $P = P(y = 1|X)$ e $1 - P = P(y = 0|X)$, podemos usar o modelo de Bernouli para probabilidades $P(x = k) = P^k(1 - P)^{1-k}$, $k = [0, 1]$.

Para cada observação no treinamento temos $P^{y_i}(1 - P)^{1-y_i}$. A função de verossimilhança que a cada uma das observações são independentes.

$$\mathcal{L}(\beta) = \prod_{i=1}^n P^{y_i}(1 - P)^{1-y_i}. \quad (1.14)$$

Assim, para maximizar a verossimilhança, temos que aplicar, primeiramente, o log na função 1.14.

$$\log \mathcal{L}(\beta) = \sum_{i=1}^n (y_i \log P + (1 - y_i) \log(1 - P)). \quad (1.15)$$

Dessa forma, possuímos uma função custo, então para maximizar a verossimilhança temos:

$$\begin{aligned} \log \mathcal{L}(\beta) &= \sum_{i=1, y=1}^n \log \frac{e^{\beta^t x}}{1 + e^{\beta^t x}} + \sum_{i=1, y=0}^n \log \frac{1}{1 + e^{\beta^t x}} \\ \log \mathcal{L}(\beta) &= \sum_{i=1, y=1}^n \log e^{\beta^t x} - \log(1 + e^{\beta^t x}) + \sum_{i=1, y=0}^n -\log(1 + e^{\beta^t x}) \\ \log \mathcal{L}(\beta) &= \sum_{i=1, y=1}^n \beta^t x_i - \log(1 + e^{\beta^t x_i}) - \sum_{i=1, y=0}^n \log(1 + e^{\beta^t x_i}) \\ \frac{d \log \mathcal{L}(\beta)}{d\beta} &= \sum_{i=1}^n x_i (y_i - \frac{e^{\beta^t x_i}}{1 + e^{\beta^t x_i}}) = 0. \end{aligned}$$

Portanto, temos que resolver essa equação para acharmos o valor dos coeficientes que maximiza o log da verossimilhança. Alguns métodos podem ser empregados para isso, como o caso de gradiente descendente ou o método de Newton-Raphson.

A superfície de decisão pode ser calculada usando $P(y = 1|x) = P(y = 0|x)$ como em:

$$\frac{e^{\beta^t x}}{1 + e^{\beta^t x}} = \frac{1}{1 + e^{\beta^t x}}$$

Resolvendo essa equação temos que $z = \beta^t x = 0$, em que a solução são hiperplanos. Logo, a superfície de separação são hiperplanos lineares $z = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$. Dessa maneira, podemos dizer que uma regressão linear se torna um método mais simples e de fácil interpretação.

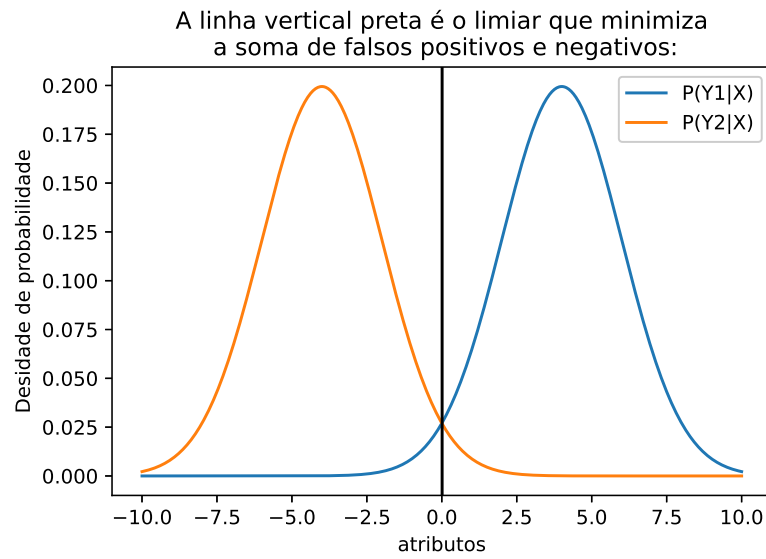


Figura 4 – Região de separação das densidades de probabilidades pelo *threshold*

1.2.4 KNN- K-Vizinhos Mais Próximos

O classificador k-vizinhos mais próximos ou do inglês *k-nearest neighbors-KNN* é um algoritmo que realiza as operações de aproximação entre variáveis para estimar as classificações através da menor distância. Dessa forma, podemos definir distâncias como uma medida de proximidade entre duas variáveis ou vetores no espaço e tem como métricas principais a *Euclidiana* e *Minkowski* [62, 63].

O algoritmo KNN possui o seguinte esquema:

1. Identificar os k-vizinhos mais próximos do vetor de atributos X que se queira classificar.
2. Determinar o número de vizinhos em cada classe.
3. Classificar X como pertencente à classe com o maior número de vizinhos (a moda entre o número de classes)

A medida de tendência central, moda, está relacionada com a teoria de classificação bayesiana, ou seja, classifique de acordo com a classe mais provável. Dessa maneira, podemos extrair uma probabilidade condicional a posteriori desse classificador.

$$P(y = j|x) = \frac{1}{k} \sum_{i \in R} P\{y_i = j\}. \quad (1.16)$$

Podemos ilustrar, as disposições dos pontos em um gráfico de dispersão com três classes para k-vizinhos mais próximos igual a 5 com o objetivo de classificar o quadrado se pertence a classe das bolas pretas, brancas ou vermelhas.

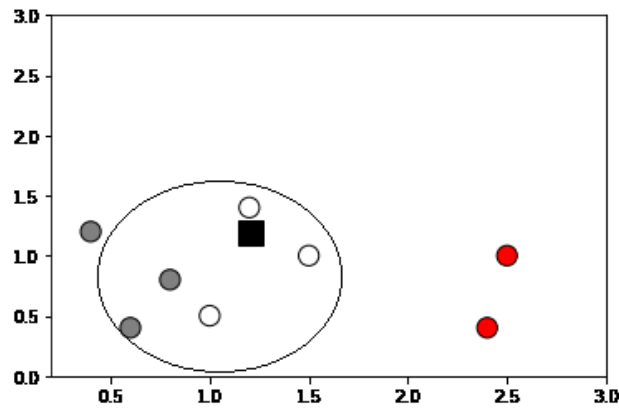


Figura 5 – localização dos pontos

A região de separação depende do valor de k . Assim, a região de decisão forma o que é chamado de telhas de *Voronoi* que é um tipo especial de decomposição de um dado espaço métrico com distâncias entre objetos ou subconjuntos no espaço. Dessa forma, dados similares tendem a estar concentrados na mesma região no espaço de dispersão dos dados.

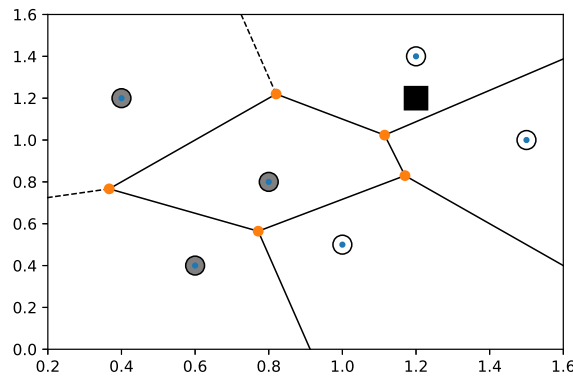


Figura 6 – Telhas de Voronoi

Para a classificação em KNN, como se trata de distâncias, precisamos normalizar ou padronizar as variáveis de entradas.

$$Z_i = \frac{X_i - \min(X)}{\max(X) - \min(X)} \rightarrow Z_i \in [0, 1]$$

$$Z_i = \frac{X_i - \mu_x}{\sigma_x} \rightarrow E[Z_i] = 0, V(Z_i) = 1$$

O algoritmo KNN possui algumas propriedades que são:

- O KNN não aprende um modelo compacto, apenas memoriza instâncias de treinamento.

- O KNN pode ser entendido como não paramétrico, dependendo apenas do número de vizinhos k .
- É um classificador não linear, não sendo restrito a regiões de separação linear.

Dado que o conjunto de treinamento seja relativamente grande, pode-se provar que o erro cometido na classificação é, no máximo, duas vezes maior do que o classificador Bayesiano, que é ótimo.

$$P(Y_j|X) = \frac{1}{k} \sum_{i \in R} I\{y_i = Y_j\}$$

$$0 \leq P_e \leq 2P_b$$

Então, para estimar as probabilidades a posteriori $P(y_i|x)$ de um conjunto de n amostras rotuladas devem ser usadas para estimar as densidades envolvidas. Suponha que coloquemos uma célula de volume V em torno de x e capturemos k amostras, k_i das quais acabamos sendo rotuladas y_i . Então, a estimativa para a probabilidade conjunta $P(x, y_i)$ é:

$$P(x \cap y_i) = \frac{k_i}{V}.$$

Então uma estimativa para a posteriori razoável para

$$P_n(y_i) = \frac{P_n(x_i \cap y_i)}{\sum_{j=1}^n P_n(x, y_j)} = \frac{k_i}{k}.$$

Ou seja, a estimativa da probabilidade a posteriori de que y_i é meramente a fração das amostras dentro da célula que são rotuladas y_i .

Na Figura 7, mostramos um exemplo da aplicação do método KNN para estimar qual a probabilidade da bola pertencer ao grupo vermelho ou azul.

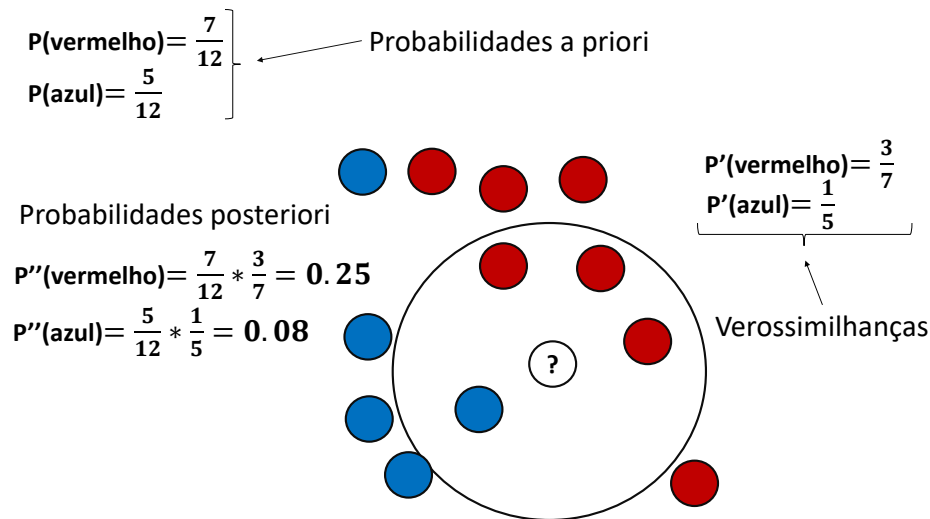


Figura 7 – Exemplo de probabilidade com KNN

1.2.5 Rede Neural Multicamadas ou Multilayer Perceptron

Nessa seção abordaremos alguns conceitos de aprendizado de máquina com o algoritmo de uma rede neural com apenas um neurônio ou com várias camadas, no inglês, Multilayer Perceptron (MLP).

1.2.5.1 Neurônio artificial

O modelo de um neurônio artificial foi concebido inicialmente por MCCULLOCH & PITTS em 1945 [64]. Um neurônio artificial é uma simplificação da realidade, com inspiração biológica de conectividade e propósito de representar aspectos relevantes de um sistema em estudo, sendo que detalhes de menor significância são descartados para viabilizar a modelagem [65, 66]. Conforme ilustrado na Figura 8.

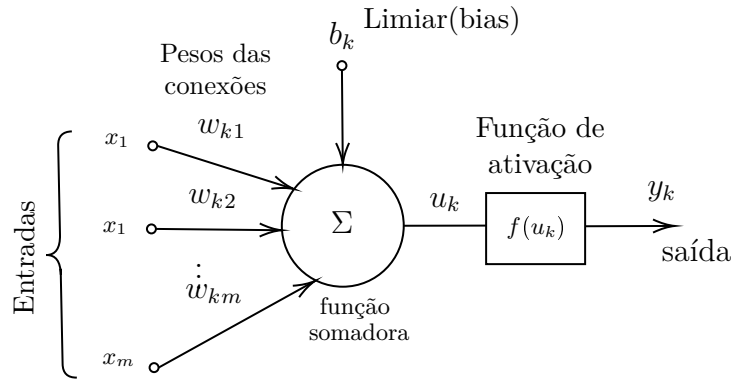


Figura 8 – Modelo matemático de um neurônio artificial

A saída do neurônio k pode ser descrita por: $y_k = f(u_k) = f\left(\sum_{j=1}^m w_{kj}x_j + b_k\right)$, definindo um sinal de entrada de valor fixo $x_0 = 1$ com peso associado $w_{k0} = b_0$, temos a seguinte expressão:

$$y_k = f(u_k) = f\left(\sum_{j=0}^m w_{kj}x_j\right) = f(\mathbf{w}^t \mathbf{x}). \quad (1.17)$$

Uma das redes neurais mais utilizadas é a *MLP- multilayer perceptron* que é conceitualmente considerada como uma combinação linear de multicamadas, em que as camadas estão dispostas em entrada, oculta e saída, as quais conseguem mapear dados de forma mais precisa em relação a apenas um neurônio ou vários de saída.

Definição 3. *Um Multi-layer Perceptron (MLP) é um algoritmo de aprendizado supervisionado que aprende uma função $f(.) := R^m \rightarrow R^o$ treinando em um conjunto de dados, onde m é o número de dimensões para entrada e o é o número de dimensões para saída [67].*

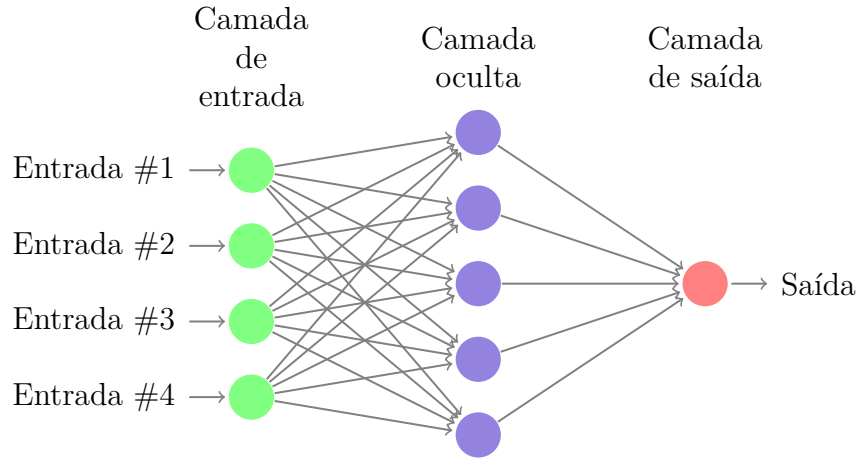


Figura 9 – Exemplo de uma Rede neural perceptron em várias camadas MLP

As MLPs utilizam, via de regra, funções de ativação sigmoidais, que possuem um perfil de “S”, ou seja, de um degrau suave. As funções de ativação identidade e tangente hiperbólica também são consideradas como funções de ativações padrões para uma MLP.

Treinando um neurônio

Considerando, como exemplo, uma rede perceptron da Figura 8 de um neurônio e m entradas e um viés, onde essas entradas representam atributos de um paciente i em um conjunto de amostras m , por exemplo, e a função de ativação f seja uma identidade, o modelo matemático pode ser dado da seguinte maneira:

$$\sum_{i=0}^m w_i^t x_{ij} = f(z) = z = y$$

Dado que m é a quantidade de amostras e o vetor peso w com o vetor de atributos das amostras i sendo dado por x .

$$w = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} \text{ e } x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

O objetivo é encontrar o peso mínimo que satisfaz a equação para o cálculo do valor w em $w^* = \arg_w [\min [D(y, f(x, w))]]$. Em que a distância mínima para esse caso será extraído pela norma-2 mínima $\|D\|_2^2$.

$$D(x, w, y) = \sum_{j=1}^m (y_j - f(x_j, w))^2 \quad (1.18)$$

Para D ser mínimo $\frac{\partial D}{\partial w} = 0$, daí vem que
$$\begin{cases} \frac{\partial D}{\partial w_0} = 0 \\ \frac{\partial D}{\partial w_1} = 0 \\ \frac{\partial D}{\partial w_2} = 0 \\ \frac{\partial D}{\partial w_3} = 0 \end{cases}$$

Portanto, o cálculo para a derivação em relação a w_0 é o que segue:

$$\frac{\partial D}{\partial w_0} = \frac{\partial}{\partial w_0} \left[\sum_{j=1}^m (y_j - f(x_j, w))^2 \right]$$

Sendo f é uma função identidade.

$$\frac{\partial D}{\partial w_0} = \frac{\partial}{\partial w_0} \left[\sum_{j=1}^m (y_j - w^t x_j)^2 \right] = \sum_{j=1}^m \frac{\partial}{\partial w_0} [(y_j - w^t x_j)^2]$$

Podemos realizar a expansão do último membro da expressão acima em:

$$\frac{\partial D}{\partial w_0} = \sum_{j=1}^m \left(\frac{\partial}{\partial w_0} (y_j - w_1 x_{1j} - w_2 x_{2j} - w_3 x_{3j} - 1w_0)^2 \right)$$

Derivando em relação a w_0 :

$$\frac{\partial D}{\partial w_0} = -2 \sum_{j=1}^m y_j + 2mw_0 + \left(\sum_{j=1}^m 2x_{1j} \right) w_1 + \left(\sum_{j=1}^m 2x_{2j} \right) w_2 + \left(\sum_{j=1}^m 2x_{3j} \right) w_3$$

Podemos realizar uma substituição nos termos da derivada parcial conforme segue.

$a = -2 \sum_{j=1}^m y_j$; $b = 2m$; $c = \left(\sum_{j=1}^m 2x_{1j} \right)$; $d = \left(\sum_{j=1}^m 2x_{2j} \right)$; $e = \left(\sum_{j=1}^m 2x_{3j} \right)$. Portanto, a representação passa a ser:

$$\frac{\partial D}{\partial w_0} = a + bw_0 + cw_1 + dw_2 + ew_3$$

Agora derivando em relação a w_1 , obtemos a seguinte expressão.

$$\begin{aligned} \frac{\partial D}{\partial w_1} = & \left(-2 \sum_{j=1}^m y_j x_{1j} \right) + \left(2 \sum_{j=1}^m x_{1j} \right) w_0 + \left(2 \sum_{j=1}^m x_{1j} x_{1j} \right) w_1 + \\ & + \left(2 \sum_{j=1}^m x_{2j} x_{1j} \right) w_2 + \left(2 \sum_{j=1}^m x_{3j} x_{1j} \right) w_3 \end{aligned}$$

Onde podemos está substituindo as expressões por: $f = \left(-2 \sum_{j=1}^m y_j x_{1j} \right)$; $g = \left(2 \sum_{j=1}^m x_{1j} \right) w_0$;

$h = \left(2 \sum_{j=1}^m x_{1j} x_{1j} \right) w_1$; $i = \left(2 \sum_{j=1}^m x_{2j} x_{1j} \right) w_2$; $j = \left(2 \sum_{j=1}^m x_{3j} x_{1j} \right) w_3$. Portanto, a representação passa a ser:

$$\frac{\partial D}{\partial w_1} = f + gw_0 + hw_1 + iw_2 + jw_3 \quad (1.19)$$

O procedimento será o mesmo para as derivadas em relação a w_2 , w_3 , portanto, temos o seguinte.

$$\frac{\partial D}{\partial w_2} = k + lw_0 + mw_1 + nw_2 + ow_3 \quad (1.20)$$

$$\frac{\partial D}{\partial w_3} = p + qw_0 + rw_1 + sw_2 + tw_3, \quad (1.21)$$

Assim, para uma rede perceptron, o vetor w de distância mínima de norma-2 pode ser encontrado resolvendo o sistema em que os valores da matriz com os coeficientes são conhecidos. Significa que podemos resolver o sistema 1.22 para encontrar o w mínimo.

$$\begin{bmatrix} a & b & c & d & e \\ f & g & h & i & j \\ k & l & m & n & o \\ p & q & r & s & t \end{bmatrix} \begin{bmatrix} 1 \\ w_0 \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (1.22)$$

Para podermos estimar a probabilidade a posteriori de uma rede neural iremos utilizar a função de ativação sigmoidal.

$$P(Y_i|X) = \frac{1}{1 + e^{\beta^t x}} \quad (1.23)$$

De modo semelhante, podemos estimar as probabilidades com uma outra função de ativação nas saídas de uma rede neural através da função softmax. A função *softmax* 1.24 mapeia as saídas em valores dentro do intervalo $[0, 1]$.

$$P(Y_i|X) = \frac{e^{w^t x}}{\sum_{k=1}^n e^{w_i^t x_i}} \quad (1.24)$$

1.2.6 Support Vector Machine- SVM

Nesta seção será apresentada uma breve abordagem do funcionamento do algoritmo de máquina de vetores de suporte. Uma máquina de vetores de suporte (Support Vector Machine-SVM, em inglês) é uma técnica de aprendizado de máquina supervisionada desenvolvida por Vapnik [46]. Esse método busca a construção de um hiperplano como fronteira de decisão, de tal maneira que a margem de separação entre os dados seja máxima e é usada para problemas de classificação e regressão [68].

A Teoria de Aprendizado Estatístico (TAE) estabelece condições matemáticas que auxiliam na escolha de um classificador particular f , a partir de um conjunto de dados de treinamento [69].

Um classificador tem a sua medida de desempenho aferida por meio de um classificador f , para que se obtenha um erro mínimo. Dado R_{emp} na Equação 1.25 como

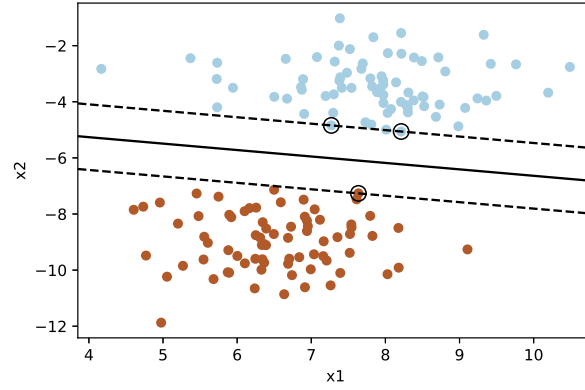


Figura 10 – Exemplo de uma separação dos dados pelo hiperplano por SVM

sendo a medida do risco empírico entre a resposta desejada e a resposta real, conforme em [70].

$$R_{emp}(f) = \frac{1}{n} \sum_{i=1}^n c(f(x_i), y_i), \quad (1.25)$$

onde $c(\cdot)$ é a função de custo relacionada à previsão $f(x_i)$ com a saída desejada y_i .

A dimensionalidade do hiperplano é obtida por meio da dimensão do conjunto de dados (n), $\dim(H) = n - 1$. Dessa forma, quando o conjunto de dados possui duas dimensões linearmente separáveis, a dimensão do hiperplano é igual a 1.

Definição 4. Seja T um conjunto de treinamento com n dados $x_i \in X$ e seus respectivos rótulos $y_i \in Y$, em que X constitui o espaço dos dados e $Y = \{-1, +1\}$. T é linearmente separável se é possível separar os dados das classes $+1$ e -1 por um hiperplano [69, 67].

Vamos verificar esta definição com um exemplo de um conjunto de dados em duas dimensões.

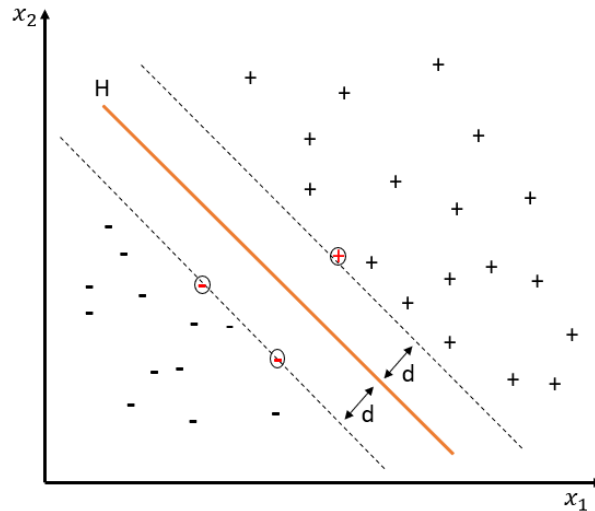


Figura 11 – Hiperplano de separação, margem máxima e vetores de suporte

Cada classe está a uma distância d do hiperplano de forma mínima [67]. Os pontos mais próximos da fronteira de separação são chamados de vetores de suporte. A distância entre os vetores de suporte e o hiperplano é chamada de margem. Nesse exemplo, o melhor hiperplano que separa as duas classes encontra-se no ponto médio.

$$f(x) = w^t x_i + b = 0 \quad (1.26)$$

Onde, na função 1.26, x é a amostra de dados, w é o vetor de pesos e b é uma constante qualquer que define o deslocamento da fronteira de decisão. Portanto, pode-se ter o seguinte:

$$g(x) = \text{sgn}(f(x)) = \begin{cases} \text{classe+} & \text{se } w^t x + b > 0 \\ \text{classe-} & \text{se } w^t x + b < 0 \end{cases} \quad (1.27)$$

Vamos convencionar que classe + e - são +1 e -1 respectivamente e sgn a para a função sinal. Define-se uma variável de classificação y que assume o valor +1 para amostras positivas (+) e o valor -1 para amostras negativas (-).

Devemos ter w e b escalados de forma que os exemplos mais próximos ao hiperplano $w^t x + b = 0$. O que implica em:

$$|w^t x_i + b| = 1 \quad (1.28)$$

Assim, com as propriedades de módulo de uma função, temos:

$$\begin{cases} w^t x_i + b \geq +1 & \text{se } y_i = +1 \\ w^t x_i + b \leq -1 & \text{se } y_i = -1 \end{cases} \quad (1.29)$$

O que implica em:

$$y_i(w^t x_i + b) - 1 > 0, i = \{1, 2, \dots, n\} \quad (1.30)$$

O cálculo das distâncias entre os dois hiperplanos pode ser feito conforme a equação 1.31, que representa a projeção ortogonal de $(x_1 - x_2)$ no hiperplano de separação, em que x_1 e x_2 são dois pontos do hiperplano $wx - b = 1$ e $wx - b = -1$, separados por $wx - b = 0$.

$$(x_1 - x_2) \left(\frac{w}{\|w\|} \frac{(x_1 - x_2)}{\|x_1 - x_2\|} \right) \quad (1.31)$$

Graficamente pode ser representado pela Figura 12. Sendo $w^t(x_1 - x_2) = 2$, pode-se substituir na Equação 1.31, vem:

$$\frac{2(x_1 - x_2)}{\|w\| \|x_1 - x_2\|} \quad (1.32)$$

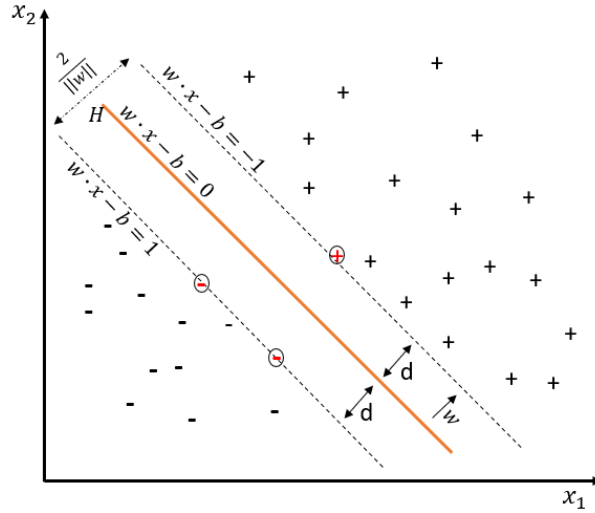


Figura 12 – Distância entre H1 e H2

Como os vetores são normalizados, obtemos $\frac{(x_1 - x_2)}{\|x_1 - x_2\|} = 1$, assim, a melhor separação é:

$$d = \frac{2}{\|w\|} \quad (1.33)$$

Dessa forma, verifica-se que a minimização da margem de separação dos dados d em relação a $w^t x + b = 0$, pode ser simplificado por $\frac{1}{\|w\|}$, pois w e b são normalizados.

Então, pode-se ter o seguinte problema de otimização.

Minimizar $\frac{1}{2}\|w\|^2$, restrito a $y_i(w^t x_i + b) - 1 > 0$, para todo $i = 1, 2, \dots, n$. [71]. A função objetivo a ser minimizada é convexa e possui um mínimo global [72, 71]. Trata-se de um problema clássico de otimização, que pode ser resolvido por multiplicadores de Lagrange [72].

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n (y_i(w^t x_i + b) - 1) \quad (1.34)$$

Onde α_i são os multiplicadores de Lagrange. Portanto, para minimizar a função Lagrangiana implica em maximizar as variáveis α_i e minimizar w e b . Obtendo um ponto de sela [73].

$$\frac{\partial L}{\partial b} = 0 \quad (1.35)$$

$$\frac{\partial L}{\partial w} = 0 \quad (1.36)$$

Daí, pelas equações anteriores obtemos o que segue:

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (1.37)$$

e

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (1.38)$$

Dessa forma, chegamos ao problema de otimização que pode ser expresso como:

$$\begin{aligned} \max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i^t x_j) \quad (1.39) \\ \left\{ \begin{array}{l} \text{Sujeito a: } \alpha_i \geq 0, \quad i = \{1, 2, \dots, n\} \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{array} \right. \end{aligned}$$

Essa é a forma dual, formulado utilizando apenas os dados de treinamento e os seus rótulos. O objetivo é determinar os valores ótimos de (w, b) .

Quando os dados não são separáveis linearmente, precisamos mapear os dados para outro espaço (geralmente de dimensão maior) e criar um classificador linear nesse novo espaço, conforme Figura 13.

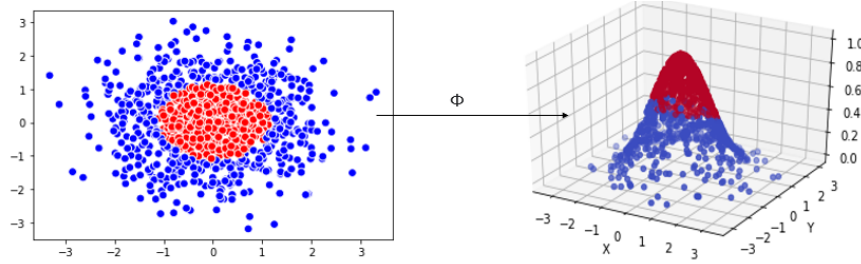


Figura 13 – Transformação do núcleo $\phi : L \rightarrow H$

Um método de produzir saídas probabilísticas de um SVM foi proposto por Wahba [74, 75]:

$$P(Y_i|X) = P(y = 1|x) = p(x) = \frac{1}{1 + e^{-f(x)}} \quad (1.40)$$

Em que f é definido na Figura 1.26, e a proposta em minimizar a verossimilhança mais um termo de penalização.

$$-\frac{1}{m} \sum_i \left(\frac{y_i + 1}{2} \log(p_i) + \frac{1 - y_i}{2} \log(1 - p_i) \right) + \lambda \|h\|_f^2 \quad (1.41)$$

Onde $p_i = p(x_i)$. A saída $p(x)$ será a probabilidade a posteriori do SVM.

1.2.7 Métodos Baseados em *Ensemble Learning*

Nesta seção apresentaremos algumas técnicas para melhorar a precisão da classificação em termos de probabilidades, agregando as previsões de vários classificadores. Essa técnica é conhecida como *ensemble* ou combinação de métodos.

O *Random forest* é um exemplo de técnica que se utiliza de manipulação das suas entradas e usa árvores de decisão como seu classificador base. Logo, esse método é

baseado em várias árvores, que manipulando dados de entradas em um subconjunto de dados, escolhidos aleatoriamente para formar um conjunto de treinamento.

Métodos baseados em *ensemble* comumente abordam a descida do gradiente para aproximação em etapas iterativas para otimização matemática nos ajustes dos pesos. O *Boosting* é uma técnica de aprendizagem de conjunto em que cada modelo tenta corrigir os erros do modelo anterior. Então, métodos que envolvem a descida do gradiente com *boosting* são uma família de técnicas poderosas de aprendizado de máquina que mostram uma redução no tempo de execução considerável em uma ampla gama de aplicações práticas.

O *Bagging*, também conhecido como agregação de *bootstrap*, é uma técnica que coleta repetidamente (com substituição) de um conjunto de dados de acordo com uma distribuição de probabilidade uniforme. Cada amostra de *bootstrap* tem o mesmo tamanho dos dados originais. E a técnica que combina previsões de resultados de diferentes saídas em termos de votos.

$$C^*(x) = \text{Voto}(C_1(x), C_2(x), \dots, C_k(x))$$

A classe pode ser obtida por maioria de votos nas previsões individuais ou ponderando cada previsão com a precisão do classificador base.

Dentre os modelos de *boosting* que podem ser explorados, como exemplos, são os *Gradient Boosting*, *XGBoost*, *Ligthboosting* e *Catboosting*. Em que as saídas desses classificadores serão dadas como a probabilidade a posteriori do modelo.

1.2.7.1 Árvore de Decisão e Floresta Aleatória

Uma árvore de decisão é um grafo induzido a partir de dados, sendo acíclico e direcionado, que apresenta um caminho único entre a raiz e qualquer outro nó do grafo, onde são geradas cadeias de regras em ordem de importância das variáveis, em que a variável de maior relevância fica no topo do grafo [70, 76].

Os elementos básicos de uma árvore de decisão são as folhas, que representam as classes de saída e os nós de decisão, que definem algum teste sobre o valor de um atributo específico, dando origem a um ramo ou sub-árvore para cada um dos valores possíveis do teste.

Ganho de informação por entropia

A impureza mede a aleatoriedade de uma variável em um dado conjunto de amostras. Uma das métricas baseadas em impureza é o Ganho de Informação, o qual usa a entropia como medida de impureza.

Em teoria da informação, a entropia é uma forma de medir o grau médio de incerteza a respeito de dados de informações, e o como resultado permite a quantificação

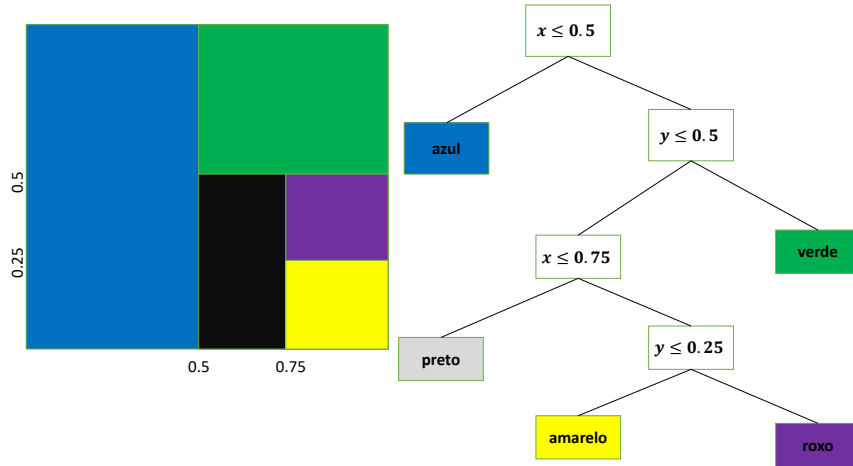


Figura 14 – Exemplo de uma árvore de decisão

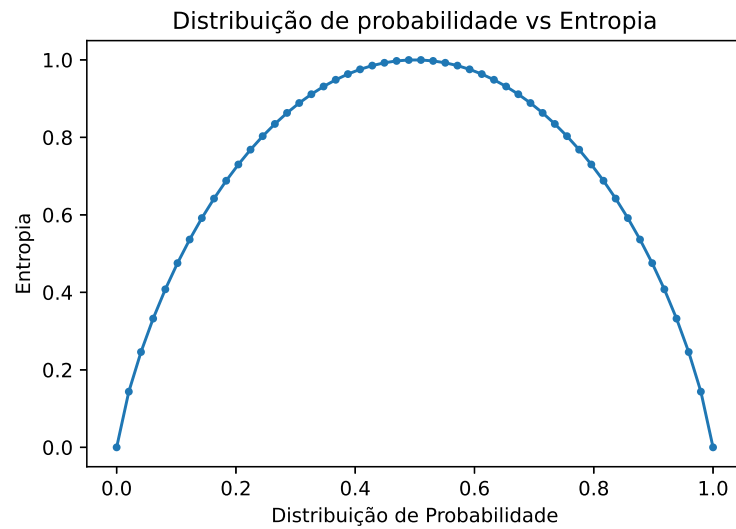


Figura 15 – Comparação entre a distribuição de probabilidade e entropia

da informação presente que tramita no sistema [77]. Quanto mais incerto o resultado de um experimento aleatório, maior é a informação que obtém-se ao observar a sua ocorrência, como no gráfico, Figura 15, pode-se notar que a entropia sempre estará variando de 0 a 1, para o problema proposto da classificação nessa pesquisa.

Para determinar o critério de início de uma árvore de decisão, faz-se o cálculo do teste das variáveis envolvidas, em que são necessários para comparar o grau de entropia do nó-pai, (antes da divisão), com o grau de entropia dos nós-filhos, (após a divisão) [78]. A variável que causar uma maior diferenciação é escolhida como condição de teste. Dessa

maneira, o ganho de informação é dado por:

$$Ganho(S, A) = Entropia(S) - \sum_{v \in Valores(A)} \frac{|S_v|}{S} Entropia(S_v), \quad (1.42)$$

onde n é o número de classe na variável saída, ou seja, o número de nós-filhos, S é o número total de objetos do nó-pai e S_v é o número de exemplos associados ao nó-filho em n . A entropia é definida pela equação a seguir:

$$Entropia(S) = \sum_{i=1}^c -p_i \log_2(p_i), \quad (1.43)$$

em que p_i =número de repetição de uma variável pela quantidade de amostras ou registros e c é o número de classes.

Como o algoritmo é recursivo, após calculado todos os ganhos de entropia, os ramos subsequentes são novamente recalculados e deve-se verificar qual variável tem o maior grau de importância. Quando uma variável tem um ganho de informação muito pequena, possivelmente, não irá ser incluída como um ramo ou nó da árvore, e pode ser excluída.

Um algoritmo derivado de árvores de decisão é o *random forest*, que foi concebida por Breiman em 2001 [79], como um comitê de máquinas composto por mais de uma árvore. Um comitê de máquinas ou comitê de classificadores retrata a agregação de mais de uma máquina de aprendizado na construção de uma solução computacional para um determinado problema [67].

A ideia de aprendizado de máquina com algoritmos em conjunto foi um passo dado para a construção de modelos de elevado desempenho que são derivados a partir de comitê (*ensemble*) de múltiplos modelos simples os quais funcionam com subconjuntos menores e aleatórios dos dados disponíveis. Dessa maneira, um modelo mais complexo é montado com a consulta a diversos especialistas para a tomada de decisão na chamada floresta randômica, ou seja, é a combinação de árvores com vetor aleatório, amostrado independentemente e segundo uma mesma distribuição de probabilidade [80].

Em se tratando de classificação, o algoritmo *random forest* utiliza o voto da maioria para dar a resposta final. Já para problemas de regressão, o cálculo é por meio da média aritmética das amostras contidas em um nó terminal/folha, e a minimização do erro quadrático médio.

A capacidade de generalização de uma *random forest* converge para um valor máximo quando o número de árvores tende ao infinito [79]. O custo computacional cai bastante em lidar com árvores distribuídas aleatoriamente. Sendo assim, é possível combiná-las para formar um único modelo de aprendizado de elevado desempenho.

Algoritmo para *random forest*

O algoritmo escolhe de forma aleatória cada variável k para a comparação da métrica de pureza/impureza (impureza de gini/entropia). Ex: sendo $k=3$, cada árvore vai ter apenas 3 variáveis com n árvores aleatórias.

Os passos lógicos do algoritmo de random forest extraídos de [81] são dados como seguem:

- Para a i -ésima árvore da floresta, faz-se uma amostragem *bootstrap* de S , denotada por $S^{(i)}$, similar ao que é feito em um *bagging*.
- Em seguida, emprega-se uma versão modificada de algum algoritmo de síntese de árvore de decisão. O algoritmo é modificado da seguinte forma: em cada nó interno da árvore, em lugar de testar todos os atributos candidatos, seleciona-se aleatoriamente um subconjunto de atributos $f \subseteq F$, onde F é o conjunto de todos os atributos de entrada, e f tem uma cardinalidade bem menor que aquela de F .
- A melhor decisão a ser tomada no nó corrente é escolhida, então, no conjunto f , e não no conjunto F . Como essa é a etapa mais custosa da técnica de construção da árvore de decisão, então o custo de síntese da árvore fica bem menor [81].

Algoritmo 1 – *Random Forest: pseudo-código 1 extraído da homepage de Matthew N. Bernstein [81]*

Entrada: A training set $S = (x_1, y_1) \dots (x_n, y_n)$, features F , and number of trees in forest B

```

Function RandomForest:  $((S, F))$ :
   $H \leftarrow 0$ ;
  para cada  $i \in 1 \dots B$  faça
     $S^{(i)} \leftarrow$  A bootstrap from  $S$ 
     $h_i \leftarrow$  RandomizedTreeLearn( $S^{(i)}, F$ )
     $H \leftarrow H \cup (h_i)$ 
  fim
  return  $H$ 
End Function

Function RandomizedTreeLearn( $((S, F))$ ):
  At each node:
     $f \leftarrow$  very small subset of  $F$ 
    Split on best feature in  $f$ 
  return The learned tree
End Function

```

O método de floresta aleatória é composto por *bagging* com reposição dos dados amostrados simultaneamente com diversificação aleatórias nas decisões pretendidas para os nós dentro das ramificações das árvores, que irá a aumentar a falta de correlação no comportamento de entrada-saída dos modelos de aprendizado [82].

Saídas em termos probabilísticos

As probabilidades podem ser calculadas pela proporção de árvores de decisão que votam em cada classe. Essa é uma abordagem mais robusta para estimar probabilidades do que usar árvores de decisão individuais. Portanto, em uma floresta aleatória, as probabilidades são dadas por $p(y_i|x) = \frac{n_a}{(n_a + n_b)}$, ou seja, o número de observações da classe A que foram ajustas por aquela folha ao longo de todo o número de observações por aquela folha.

Os rótulos das classes de saídas desse classificador podem ser disponibilizadas em forma de distribuição de probabilidade. A probabilidade do nó pode ser calculada pelo número de amostras que chegam ao nó dividido pelo número total de amostras. Quanto maior o valor, mais importante o atributo.

Denotamos por $P(y|x)$ a probabilidade estimada de um exemplo incidente em um nó x possuir classe y . Dessa forma, o $P(y|x)$ pode ser definido como a proporção de exemplos de x que possuem classe y [83, 84].

O critério de atribuição é adotar a classe de maior probabilidade estimada no nó, ou melhor, a classe que maximiza $p(y|x)$. Assim, a probabilidade estimada de um exemplo incidente sobre o nó x ser classificado erroneamente é:

$$r(x) = 1 - \max_y P(y|x) = \min_y \sum_{i \neq y}^n P(i|y) \quad (1.44)$$

Em que na Equação 1.44 $i = \{1, 2, 3, \dots, N\}$. Escolher a classe que possui a maior probabilidade significa minimizar o erro.

O custo de erro em classificar um objeto pertencente à classe y como sendo de classe i é denotado por $C(i, y)$ e satisfaz:

- $C(i, y) \rightarrow 0$ e $i \neq y$
- $C(i, y) = 0$

Se um objeto de classe desconhecido incide sobre um nó x , for classificado com classe i , então o custo esperado do erro na classificação será $\sum_y C(i, y)P(i, y)$. Dessa forma, escolher um critério natural de atribuição é escolher a classe i que minimiza a expressão acima.

$$r(x) = \min_i \sum_y C(i, y)P(y|x) \quad (1.45)$$

Para classificação em várias saídas, podemos ter como saída de uma árvore probabilística, vide [85].

$$classe(x) = (p(1|x), p(2|x), \dots, p(y|x))$$

O custo estimado em se atribuir a um objeto à classe y estando no nó x é dado pelo somatório ponderado.

$$\sum_y C(i, y)p(y|x).$$

O custo estimado de erro no nó x é:

$$r(x) = \sum_i p(i|x) \left(\sum_{i \neq y} C(i, y)p(y|x) \right) = \sum_{i,y} C(i, y)p(i|x)p(y|x).$$

No exemplo a seguir, vamos apresentar a primeira etapa através de um exemplo que leva em consideração uma simples coleta hipotética de amostras de 14 pacientes. A tarefa do algoritmo é decidir qual o sintoma possui o maior grau de importância, ou seja, aquele que ficará no nó raiz. Devemos calcular a entropia geral dos dados tabulares acima

Tabela 4 – Tabela de sintomas de pacientes com suspeita de COVID-19

Calafrio	Coriza	Febre	Cefaleia	Risco COVID-19
Baixo	Sim	Não	Forte	Alto
Médio	Sim	Não	Média	Alto
Médio	Não	Não	Média	Moderado
Médio	Não	Não	Baixa	Alto
Médio	Não	Não	Baixa	Baixo
Médio	Não	Sim	Baixa	Baixo
Baixo	Não	Não	Forte	Alto
Baixo	Não	Sim	Baixa	Moderado
Alto	Não	Não	Baixa	Baixo
Alto	Sim	Sim	Baixa	Baixo
Alto	Sim	Não	Forte	Alto
Alto	Sim	Não	Média	Moderado
Alto	Sim	Não	Baixa	Baixo
Baixo	Sim	Não	Média	Alto

com relação a coluna *Risco de COVID-19*. Assim, iremos realizar, por partes separadas, os devidos cálculos.

Para calcular o grau de entropia do sintoma Cefaleia iremos montar um grafo com as probabilidades como ilustrado em 16 e Tabela 5.

Observando as probabilidades atribuídas no grafo, podemos calcular a entropia e o ganho para o sintoma Cefaleia.

Risco de COVID-19

Alto	$P(Alto Risco) = \frac{6}{14}$
Alto	$P(Moderado Risco) = \frac{3}{14}$
Moderado	$P(Baixo Risco) = \frac{5}{14}$
Alto	$E(Risco) = -\frac{6}{14} \log_2 \left(\frac{6}{14} \right) - \frac{3}{14} \log_2 \left(\frac{3}{14} \right) -$
Baixo	$-\frac{5}{14} \log_2 \left(\frac{5}{14} \right) = \mathbf{1.53}$
Baixo	
Alto	
Moderado	
Baixo	
Baixo	
Alto	
Moderado	
Baixo	
Alto	

- $E(Baixa) = -\frac{3}{4} \log_2 \left(\frac{3}{4} \right) - 0 \log_2(0) - 0 \log_2(0) = 0$
- $E(Média) = -\frac{2}{4} \log_2 \left(\frac{2}{4} \right) - \frac{2}{4} \log_2 \left(\frac{2}{4} \right) - 0 \log_2(0) = 1$
- $E(Forte) = -\frac{1}{7} \log_2 \left(\frac{1}{7} \right) - \frac{1}{7} \log_2 \left(\frac{1}{7} \right) - \frac{5}{7} \log_2 \left(\frac{5}{7} \right) = 1.15$

Portanto, o ganho de informação é dada como:

$$G(Risco|Cefaleia) = 1.53 - \left(\frac{3}{14} \cdot 0 \right) - \left(\frac{4}{14} \cdot 1 \right) - \left(\frac{7}{14} \cdot 1.15 \right) = \mathbf{0.66}$$

A próxima etapa será o cálculo da entropia e o ganho para o atributo Coriza. O grafo que representa a distribuição de probabilidades para Coriza.

Então calculando as entropias individualmente.

- $E(Sim) = -\frac{4}{7} \log_2 \left(\frac{4}{7} \right) - \frac{1}{7} \log_2 \left(\frac{1}{7} \right) - \frac{2}{7} \log_2 \left(\frac{2}{7} \right) = 1.38$
- $E(Não) = -\frac{2}{7} \log_2 \left(\frac{2}{7} \right) - \frac{2}{7} \log_2 \left(\frac{2}{7} \right) - \frac{3}{7} \log_2 \left(\frac{3}{7} \right) = 1.56$

Tabela 5 – Tabela de Risco de Cefaleia

Cefaleia	Risco COVID-19
Forte	Alto
Média	Alto
Média	Moderado
Baixa	Alto
Baixa	Baixo
Baixa	Baixo
Forte	Alto
Baixa	Moderado
Baixa	Baixo
Baixa	Baixo
Forte	Alto
Média	Moderado
Baixa	Baixo
Média	Alto

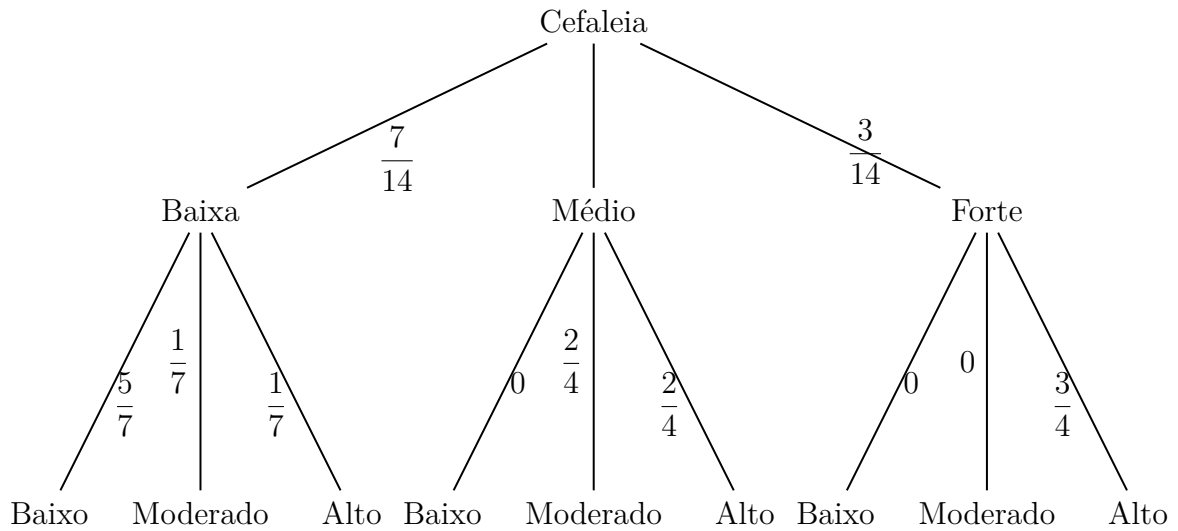


Figura 16 – Grafo para o sintoma Cefaleia

Portanto, tendo em vista as entropias, temos que o ganho de informação do atributo Coriza conforme o Grafo 17 será:

$$G(Risco|Coriza) = 1.53 - \left(\frac{7}{14} \cdot 1.38\right) - \left(\frac{7}{14} \cdot 1.56\right) = \mathbf{0.06}$$

Agora, pretendemos realizar o cálculo para o atributo, ou o sintoma Febre no Grafo 18. Do grafo com as probabilidades, temos:

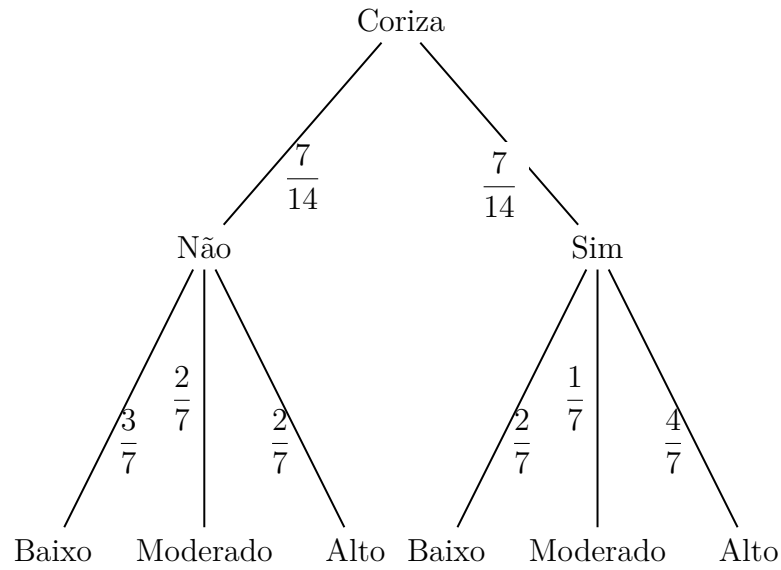


Figura 17 – Grafo para o sintoma de Coriza

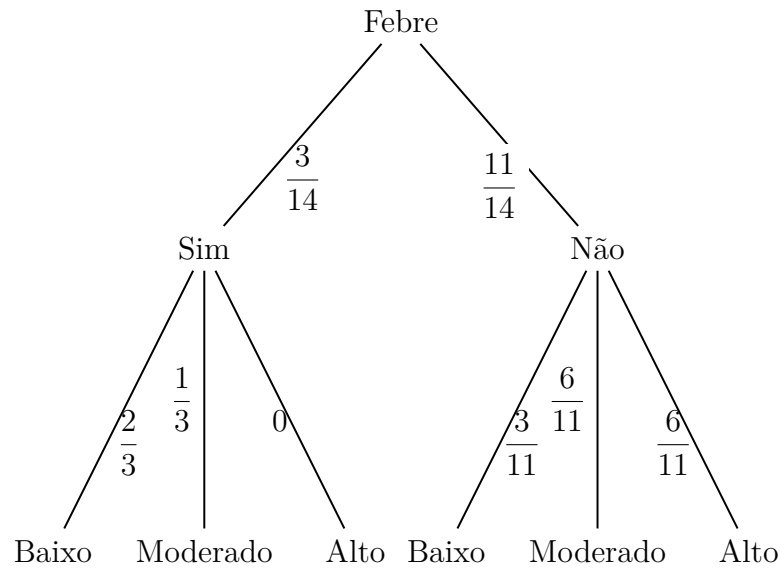


Figura 18 – Grafo para o sintoma de Febre

Então calculando as entropias individualmente.

- $E(Não) = -\frac{6}{11} \log_2 \left(\frac{6}{11} \right) - \frac{2}{11} \log_2 \left(\frac{2}{11} \right) - \frac{3}{11} \log_2 \left(\frac{3}{11} \right) = 1.44$
- $E(Sim) = -0 \log_2(0) - \frac{1}{3} \log_2 \left(\frac{1}{3} \right) - \frac{2}{3} \log_2 \left(\frac{2}{3} \right) = 0.92$

Portanto, tendo em vista as entropias, temos que o ganho de informação do atributo Coriza será:

$$G(Risco|Febre) = 1.53 - \left(\frac{11}{14} \cdot 1.44 \right) - \left(\frac{3}{14} \cdot 0.92 \right) = \mathbf{0.20}$$

O cálculo das entropias para o atributo Calafrio. Então calculando as entropias

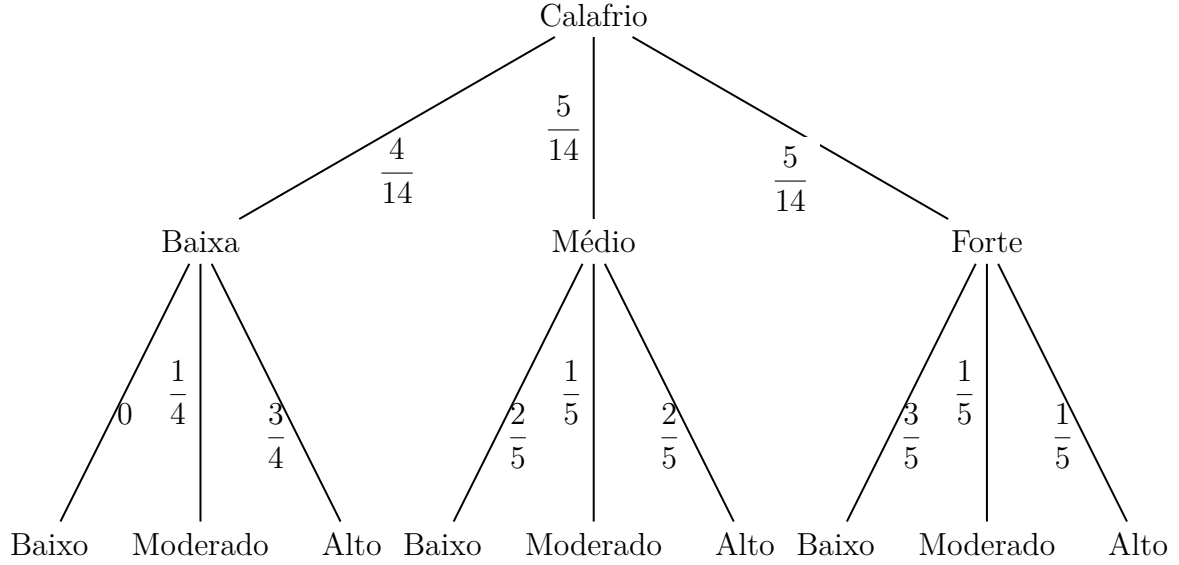


Figura 19 – Grafo para o sintoma Calafrio

individualmente.

- $E(Forte) = -\frac{1}{5} \log_2 \left(\frac{1}{5} \right) - \frac{1}{5} \log_2 \left(\frac{1}{5} \right) - \frac{3}{5} \log_2 \left(\frac{3}{5} \right) = 1.37$
- $E(Médio) = -\frac{2}{5} \log_2 \left(\frac{2}{5} \right) - \frac{1}{5} \log_2 \left(\frac{1}{5} \right) - \frac{2}{5} \log_2 \left(\frac{2}{5} \right) = 1.52$
- $E(Baixa) = -\frac{3}{4} \log_2 \left(\frac{3}{4} \right) - \frac{1}{4} \log_2 \left(\frac{1}{4} \right) - 0 \cdot \log_2(0) = 0.81$

Portanto, tendo em vista as entropias, temos que o ganho de informação do atributo Calafrio será:

$$G(Risco|Calafrio) = 1.53 - \left(\frac{5}{14} \cdot 1.37 \right) - \left(\frac{5}{14} \cdot 1.52 \right) - \left(\frac{4}{14} \cdot 0.81 \right) = \mathbf{0.26}$$

Então o nó raiz da árvore, por meio dos cálculos feitos acima, será conforme a Figura 20.

Para o cálculo de uma árvore completa, precisamos realizar os mesmos procedimentos para os outros ramos de forma recursiva.

Perceba que na Figura 20, trata-se de uma árvore de decisão com profundidade 1 e com o nó raiz sendo Cefaleia, ou seja, possui maior risco para ser diagnóstico positivo. Dessa forma, em uma primeira análise, percorrendo no sentido positivo para a variável Cefaleia, pode-se inferir as chances altas para o verdadeiro positivo da doença. Continuando o olhar para a árvore do exemplo, observamos que os outros dois ramos são risco Médio e risco Baixo, o que pode indicar uma situação menos provável de prognóstico positivo para a saída positiva.

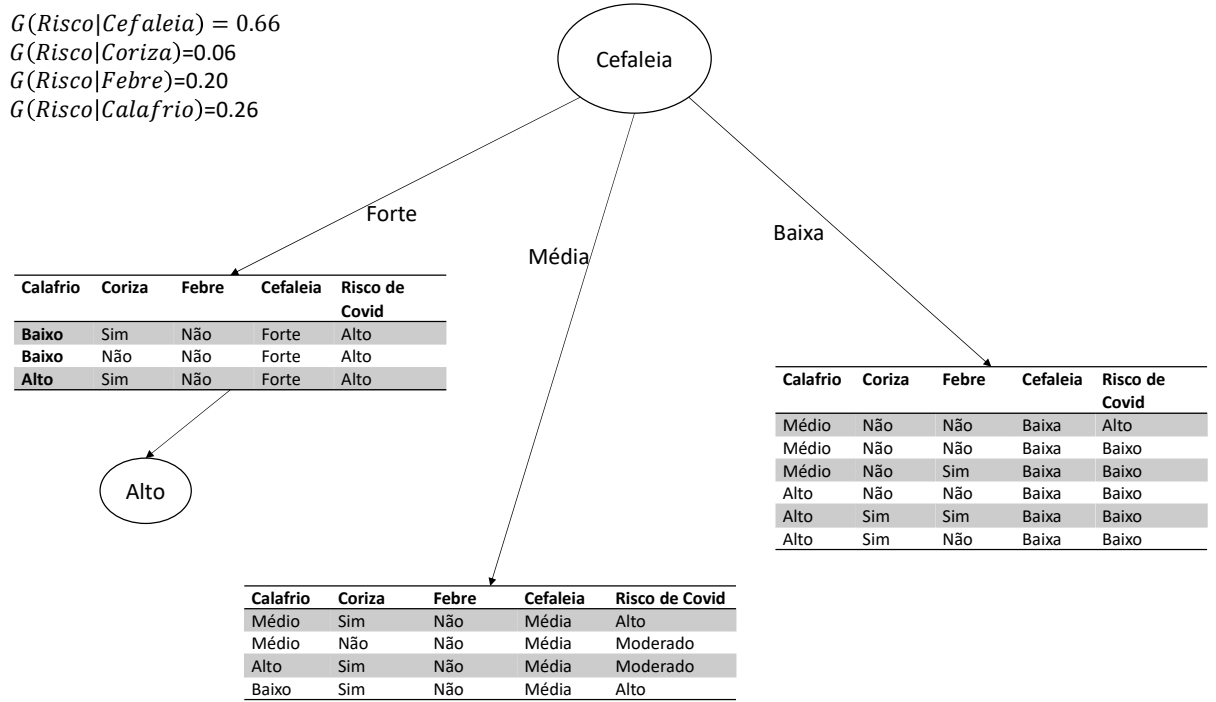


Figura 20 – Primeiro ramo da árvore de decisão construída

1.2.7.2 Gradient Boosting Machine-GBM

Dentre as categoriais de algoritmos que fazem o uso do *gradient boosting* como núcleo de seu algoritmos são o XGBoost, LigthBoost, Catboost, que serão aplicados neste trabalho [86, 87].

O objetivo do boosting é resolver o seguinte problema de otimização:

$$\min_f \sum_{i=1}^N L(y_i, f(x_i)) \quad (1.46)$$

Em que $L(y, \hat{y})$ é a função de perda, que depende do modelo inicial e dos valores de y e $f(x_i) = \hat{y}$. Se usamos quadrados mínimos, o cálculo da estimativa é dada pelo erro E :

$$f^*(x) = \arg \min_{f(x)} E_{y|x}[(Y - f(x))^2] = E[Y|x] \quad (1.47)$$

O *boosting* tentará aproximar essa expectativa condicional. Para classificação binária, a perda óbvia é entre $[0, 1]$, mas isso não é diferenciável. Em vez disso, é comum usar *logloss*, que é um limite superior convexo na perda de 0 e 1. Nesse caso, pode-se mostrar que a estimativa ótima é dada por:

$$f^*(x) = \frac{1}{2} \log \frac{P(\tilde{y} = 1|x)}{P(\tilde{y} = -1|x)} \quad (1.48)$$

O algoritmo XGBoost significa “*Extreme Gradient Boosting*”, onde o termo “*Gradient Boosting*” origina-se do artigo Greedy Function Approximation: A Gradient Boosting Machine, de Friedman [86]. O XGBoost foi o resultado de uma pesquisa de Tianqi Chen, Ph.D. na Universidade de Washington [61].

Seja o seguinte a função objetivo obj dada por $obj = \sum_{i=1}^n l(y_i, \hat{y}_i^t) + \sum_{i=1}^t \Omega(f_i)$ onde l é a função de perda e Ω a função de regularização.

1.2.7.3 Comitê de Classificadores

Comitê de máquinas (do inglês *committee machine*), é um tipo de *ensemble*, que visam produzir ganhos de desempenho frente a propostas de soluções isoladas, ou seja, que não formam comitês [88, 89].

Um comitê de máquinas é uma forma de aprendizado de máquina que pode ser aplicada tanto para problemas de classificação quanto regressão [58].

Para comitês de máquinas, no que se trata de *ensembles*, visam obter desempenho tão maior quanto melhor for à execução individual de cada máquina e tanto mais aumenta-se a performance desses componentes do *ensemble*, o classificador terá uma boa eficiência [90]. A utilização de comitês de máquinas pode trazer os seguintes benefícios [67, 62]:

- Redução na variância do modelo final;
- Melhora na tolerância a ruído nos dados;
- Melhora na capacidade de generalização.

O termo *ensemble* é normalmente utilizado para a combinação de um conjunto de modelos ou classificadores redundantes, no sentido de que cada modelo obtém uma solução para a tarefa [91].

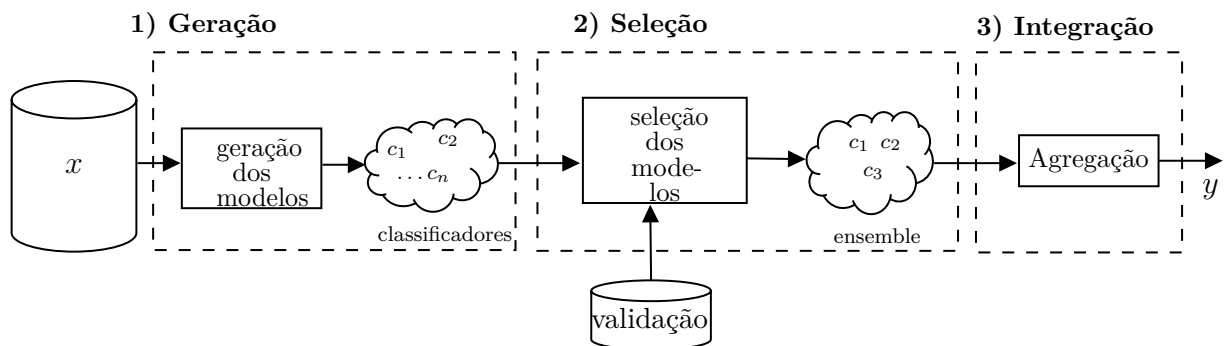


Figura 21 – Etapas de um sistema de múltiplos classificadores. [Adaptada de Cruz et al [1].]

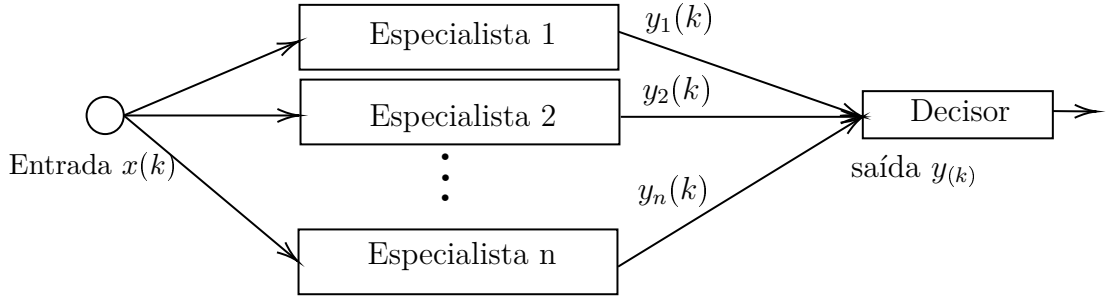


Figura 22 – Exemplo de um combinador

Na Figura 21 consiste em três etapas: a geração dos modelos, seleção dos melhores modelos classificadores e agregação dos resultados da tomada de decisão. Portanto, a escolha dos melhores classificadores permite que a saída obtenha a melhor resposta a qual pretende-se obter. A Figura 22 representa um combinador de vários especialistas ou classificadores diferentes conforme descrito por [92].

1.2.7.3.1 Combinador Linear

Para combinar as predições, utilizamos voto majoritário, para o qual há basicamente dois modos de implementação: *hard* e *soft*, os quais são baseados nos discriminantes Δ_{qc} , onde $q = 1, 2, \dots, Q$ indexa as máquinas, e $c = 1, 2, \dots, Q$, indexa as classes do conjunto $\Lambda = \lambda_1, \lambda_2, \dots, \lambda_C$. Portanto, para uma classe fixa λ_K , com $K = 1, 2, \dots, C$, o discriminante fornece segundo a Equação 1.49 [93, 94, 93, 95].

$$\Delta_{qc}(x_n) = \begin{cases} 1, & \text{se } h_q(x_n, \beta) = \lambda_K \\ 0, & \text{caso contrário} \end{cases} \quad (1.49)$$

Com $q = 1, 2, \dots, Q$ e onde $h_q(x_n, \beta)$ é um modelo ajustado com os parâmetros β na etapa de treinamento [95].

Assim, a classe $\lambda \in \Delta$ é predita considerando os pesos de cada máquina w_q .

Para os dois tipos de classificação pode ser realizado por [96]:

Hard que utiliza a quantidade de predições corretas

$$\lambda_h = \underset{\lambda_K \in \Lambda}{\operatorname{argmax}} \left\{ \sum_{q=1}^Q w_q \Delta_{qK}(x_n) \right\}.$$

Soft que emprega as probabilidades.

$$\lambda_s = \underset{\lambda_K \in \Lambda}{\operatorname{argmax}} \left\{ \sum_{q=1}^Q w_q p_q(x_n | \lambda_K) \right\}.$$

Onde $p_q(x_n | \lambda_K)$ é a probabilidade da q –ésima máquina prever o valor λ_K dado o padrão x_n . O método *soft* será empregado neste trabalho.

1.2.7.3.2 Pesos das Máquinas

Há diversas maneiras para obtenção dos pesos, a mais simples baseada na acurácia de cada máquina, sensibilidade, precisão, AUC entre outros como uma média. Dessa forma, obtemos um vetor de pesos w para o problema de classificação utilizando o comitê de máquinas.

1.3 Métodos de Apoio à Decisão Multicritérios

Nesta seção traremos técnicas em métodos de apoio à decisão multicritérios, com o propósito de fornecer componentes de saídas em termos de *scores*. Os *scores* são pontuações agregada conforme a função de pontuação e o peso de cada critério com indicação da classificação. E, logo em seguida, os *scores* são transformados em probabilidades [97].

Apresentaremos o método aditivo linear, TOPSIS e Integral de Choquet que possuem características em que cada critério possui um conjunto de pesos que pode ser atribuído por um especialista ou ser ajustado por otimização. Uma vantagem da utilização desses métodos se deve à fácil interpretabilidade na aplicação. Esses métodos irão compor o outro ramo do *framework* do C3PO.

1.3.1 Método Aditivo Linear

O método *Simple Additive Weighting (SAW)* [97] é um procedimento multi-atributo baseado no conceito de um somatório ponderado, dessa maneira, é considerado um método aditivo linear ponderado. O sistema procurará uma soma ponderada da classificação do desempenho de cada amostra em todos os critérios especificados como *scores*. Esse método pode ser usado para compor o sistema do C3PO. O Método Aditivo Linear ou *SAW* terá como saída um range de *scores* ponderados e posteriormente transformado em probabilidade de forma que se pode estabelecer um ponto de corte para a classificação ou *threshold*.

A utilidade do conceito básico do método *SAW* é encontrar o número de classificações de desempenho ponderadas para cada alternativa em todos os atributos. Assim, ele pode ser utilizado em conjunto com outros métodos como reforço adicional para o resultado de interesse. Abaixo, ilustramos, na Figura 23, um fluxograma para o método *SAW* realizado neste estudo.

Para evidenciar o método *SAW*, ou seja, o método aditivo de decisão multicritério, exporemos um exemplo.

Para os resultados que serão apresentados, as siglas, por simplificação, são dadas da seguinte forma: o sintoma, s_1 para Febre, s_2 para Tosse, s_3 para Dor no peito,

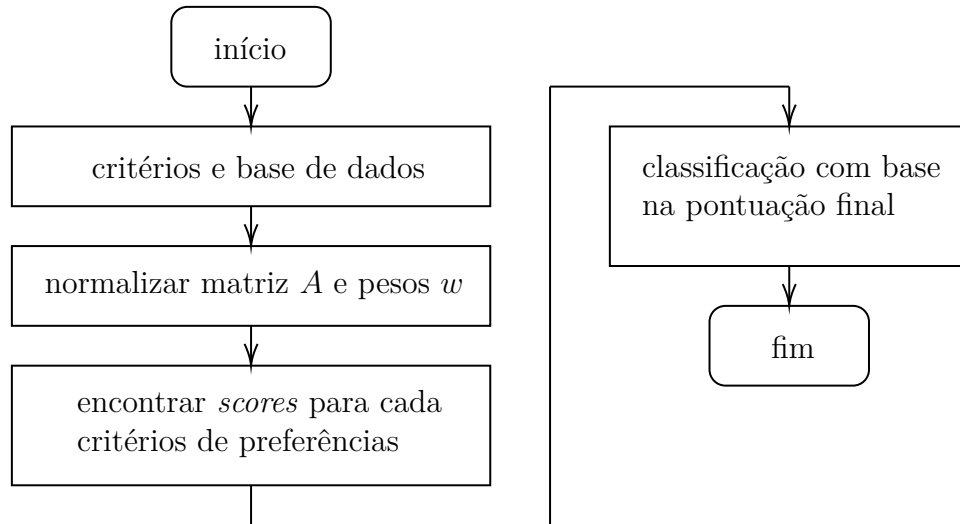


Figura 23 – Etapas do Método Aditivo Linear

s_4 para Saturação 95%. E na tabela de resultados, a coluna *scores* com os resultados do método e, logo em seguida, com as probabilidades dadas através de uma regressão logística.

Tabela 6 – Método aditivo

pesos	0.2	0.2	0.3	0.3			
sintomas.	s_1	s_2	s_3	s_4	<i>scores</i>	<i>Probs(Aditivo)</i>	<i>Resut.</i>
p_1	1	1	0	0	0.4	0.38	0
p_2	1	0	1	1	0.8	0.9	1
p_3	0	1	0	1	0.5	0.5	0
p_4	0	0	1	1	0.6	0.55	1

No resultado apresentado na Tabela 6, os valores em que as probabilidades ficaram acima de um ponto de corte preestabelecido de 0.5 tiveram classificação igual a 1.

1.3.2 Método TOPSIS

O método de análise de decisão multicritério TOPSIS foi originalmente desenvolvido por Ching-Lai Hwang e Yoon em 1981, passando atualizações de desenvolvimentos por Yoon em 1987, e Hwang, Lai e Liu em 1993 [98, 99].

Por meio da ideia principal desta técnica, a alternativa preferida é aquela que estiver mais perto da solução ideal positiva e mais longe da solução ideal negativa. A solução ideal positiva é formada pela combinação dos melhores pontos de cada critério, conforme mostrado na Figura 24. A solução ideal negativa é uma combinação dos piores pontos de cada critério. Esta técnica só pode ser aplicada para o conjunto de dados numéricos onde os pesos de importância do critério são conhecidos ou definidos com base na opinião de especialistas numericamente. E os resultados do *ranking* podem ser obtidos correspondendo aos pesos de importância dos critérios definidos [100].

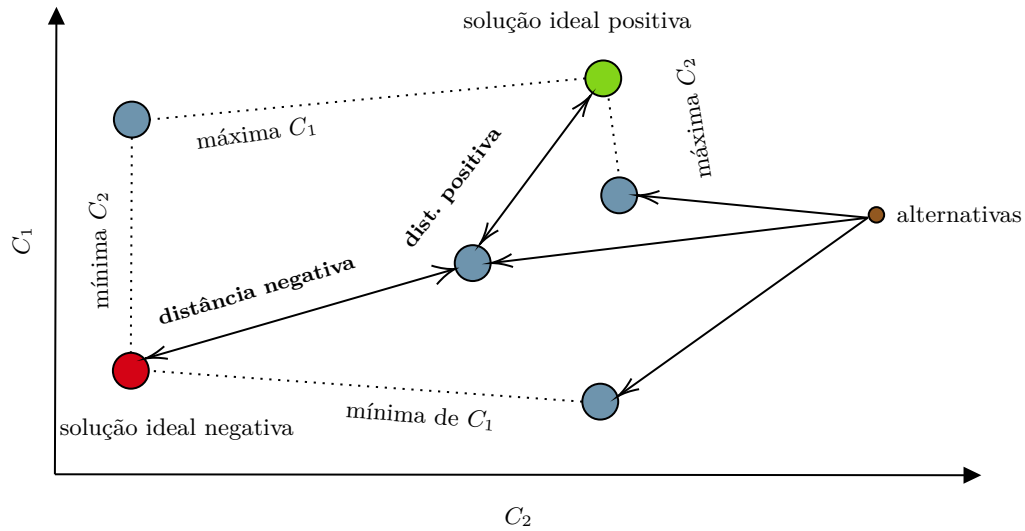


Figura 24 – Solução ideal para o método TOPSIS

O algoritmo TOPSIS, será dado conforme as etapas mostradas no esquema exemplificado no Algoritmo 2. Com base na estrutura do pseudocódigo, tornaremos-lo, o TOPSIS, como parte do bloco da abordagem do C3PO na tomada de decisão multicritério e sempre, como estudo de caso, tomando um limiar de 0.5, após a transformação dos *scores* de saídas em probabilidade.

Vamos ilustrar uma aplicação para o método de tomada de decisão multimétricos TOPSIS. Para isso, utilizaremos a mesma matriz de decisão do exemplo anterior, Tabela 6, para o método SAW.

No método TOPSIS, primeiramente, realizamos a normalização da matriz de decisão que contém os sinais e sintomas da Tabela 6. Logo em seguida, é realizado os cálculos das distâncias de separação pela norma-2. Os *scores* vêm do resultado de classificação da alternativa da solução ideal positiva.

No resultado apresentado na Tabela 7, obtivemos os valores de probabilidades através da regressão logística, caso do exemplo, o ponto de corte de diagnóstico está fixo com o valor de 0.5. No entanto, esse valor pode ser modificado com valores na faixa de zero a um em situação de triagem hospitalar ou diagnósticos clínicos.

Portanto, um dos propósitos do uso do método TOPSIS, neste trabalho, foi o fato dessa técnica possuir baixa complexidade e custo computacional [99].

Na Tabela 7, a coluna de *scores* possuem valores obtidos das distâncias das alternativas ideais positivas e negativas, e logo em seguida foram transformadas em probabilidades aplicando-se uma regressão logística com um ponto de corte de 0.5 para a classificação, os pesos foram obtidos de forma *Ad hoc*.

Algoritmo 2 – Etapas do método TOPSIS em formato de pseudocódigo

1. Construção da matriz de decisão normalizada, ou seja, $r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}}$,
para todo r_{ij} e x_{ij} elementos de normalização da matriz de decisão original respectivamente.
2. Construção da matriz de decisão normalizada ponderada $v_{ij} = r_{ij}w_{ij}$
onde w_j é o peso atribuído ao atributo j
3. Determinar a solução ideal (A^+) e a solução ideal negativa (A^-)
 $A^+ = \{(max_j v_{ij} | i \in I), (min_j v_{ij} | i \in I'); \forall j\} = \{v_1^+, v_2^+, \dots, v_n^+\}$
 $A^- = \{(min_j v_{ij} | i \in I), (max_j v_{ij} | i \in I'); \forall j\} = \{v_1^-, v_2^-, \dots, v_n^-\}$
 onde I e I' estão associados aos atributos de benefício e custo, respectivamente.
4. Cálculo da medida de separação, ou seja,

$$S_i^+ = \sqrt{\sum_{j=1}^n (v_{ij} - v_i^+)^2}$$

$$S_i^- = \sqrt{\sum_{j=1}^n (v_{ij} - v_i^-)^2}$$
5. Cálculo da proximidade relativa da solução ideal, ou seja,

$$D_j^+ = \frac{S_j^+}{S_j^+ + S_j^-}$$

$$D_j^- = \frac{S_j^-}{S_j^+ + S_j^-}$$
6. Classificação de alternativas com base no valores de D_j^+ ou de D_j^- .

Tabela 7 – Método TOPSIS

pesos	0.2	0.2	0.3	0.3					
sintomas	s_1	s_2	s_3	s_4	d_+	d_-	score	Probs.	Result.
p_1	0.2121	0.0707	0	0	0.2707	0.2236	0.4522	0.48	0
p_2	0.2121	0	0.1414	0.2309	0.0707	0.3439	0.8294	0.85	1
p_3	0	0.0707	0	0.2309	0.2549	0.2415	0.4864	0.5	0
p_4	0	0	0.1414	0.2309	0.2236	0.2708	0.5477	0.6	1
p_+	0.2121	0.0707	0.1414	0.2394	0	0.3511	1		
p_-	0	0	0	0	0.3511	0	0		

1.3.3 Integral de Choquet

Nesta seção iremos fazer uma breve revisão e trazer algumas definições sobre a integral de Choquet para a tomada de decisão multicritério. A Integral de Choquet é intensamente utilizada em tomada de decisão com múltiplos critérios como operador não-linear de agregação [101].

Na abordagem multicritério baseada em integrais de Choquet, é possível modelar sinergia e/ou redundância entre critérios, que poderá ser aplicada para considerar informações adicionais sobre os dados, evitando resultados enviesados [102]. Logo, a proeminência da integral de Choquet ocorre devido ao seu modelo em respeitar as preferências e inseguranças de decisões para cada atributo a ser agregado, bem como suas interações.

A integral de Choquet é uma função de agregação de utilidade que mede a influência da importância dos critérios individuais e a importância das relações entre os critérios [103].

Uma importante característica notória da integral de Choquet é o conceito de capacidade de Choquet (ou medida *fuzzy*) [104]. Nas circunstâncias das funções de agregação, a medida *fuzzy* representa o grau de relação entre os elementos a serem agregados [105].

Definição 5. (*Medida Fuzzy*) Seja $N = \{1, 2, \dots, n\}$. Uma medida fuzzy discreta é uma função $\mu : 2^N \rightarrow [0, 1]$ que é monótona (i.e. $\mu(A) \leq \mu(B)$ nem sempre $A \subset B$) e satisfaz $\mu(\emptyset) = 0$ e $\mu(N) = 1$ (condição de fronteira).

A integral de Choquet discreta do inglês (*Choquet Integral-CI*) em uma amostra i , pode ser definida com respeito a uma medida monotonicidade de μ como sendo dada pela definição a seguir [104]:

Definição 6. Seja μ uma medida fuzzy em um conjunto N cujos elementos são denotado como $P_{i1}, P_{i2}, \dots, P_{im}$. A integral discreta de Choquet de uma função $f : X \rightarrow \mathbb{R}^+$ em relação a μ é dada por:

$$CI_i = \sum_{j=1}^m (P_{i(j)} - P_{i(j-1)})\mu(D_{(j)}) \quad (1.50)$$

Onde, P_{ij} é o valor de uma alternativa i para o critério j . O operador de índice representa uma permutação com a seguinte ordem: $P_{i(1)} \leq P_{i(2)} \leq \dots \leq P_{i(m)}$ e $D_{(j)} = \{C_{(t)}, C_{(t+2)}, \dots, C_{(m)}\}$ e o m é o número de critérios de uma determinada amostra i e os elementos do conjunto C representam os critérios da alternativa ou amostra. Devido ao caráter combinatório dessa medida, para um problema com m critérios, teremos 2^m elementos que a compõem.

A propriedade de monotonicidade, significa que o procedimento de se adicionar um novo critério em uma coalizão não pode diminuir o valor da capacidade para aquela coalizão.

Seja o exemplo com três critérios [104] $C = \{C_1, C_2, C_3\}$. Sendo $\mu(\cdot)$ a medida de capacidade de Choquet de um modelo de interação de critérios e relacionada com a importância da coalizão entre os critérios do conjunto C é dada por:

- $\mu(C_1, C_2) < \mu(C_1) + \mu(C_2)$ (Redundância)
- $\mu(C_1, C_2) > \mu(C_1) + \mu(C_2)$ (Sinergia)
- $\mu(C_1, C_2) = \mu(C_1) + \mu(C_2)$ (Ausência de interação)

Onde a capacidade de Choquet é uma função μ que satisfaz as seguintes propriedades:

- $\mu(C_1) \leq \mu(C_1, C_2)$;
- $\mu(C_2) \leq \mu(C_1, C_2)$;
- $\mu(C_1, C_2) \leq \mu(C_1, C_2, C_3)$;
- ...
- $\mu(C_1, C_2, C_3) = 1$ valor máximo;
- $\mu(\emptyset) = 0$ valor mínimo.

Exemplo 1. *Vamos realizar uma aplicação para o problema em determinar se uma gestante possui COVID-19 de acordo com a pontuação de três sintomas apresentados. Quanto maior o escore de pontuação, maiores são as chances da paciente está com a doença. Modelar considerando a relação de redundância entre os sintomas s_1 e s_2 .*

Tabela 8 – Sintomas de pacientes

	s_1	s_2	s_3
P_1	0.30	0.60	0.95
P_2	0.45	0.40	0.85
P_3	0.55	0.95	0.10

Fazendo, $\mu(c_1) = \frac{1}{3}$, $\mu(c_2) = \frac{1}{3}$, $\mu(c_3) = \frac{1}{3}$, $\mu(c_1, c_2) = \frac{2}{3}$, $\mu(c_1, c_3) = \frac{2}{3}$, $\mu(c_2, c_3) = \frac{1}{3}$, $\mu(\emptyset) = 0$ e $\mu(c_1, c_2, c_3) = 1$.

Para entender melhor a ideia sobre a integral de Choquet, vamos demonstrar graficamente uma situação com um paciente P_1 sendo avaliada para três critérios.

Observe que $\mu(c_2, c_3) = \frac{1}{3} < \mu(c_2) + \mu(c_3) = \frac{2}{3}$.

Para o cálculo do escore para o paciente 1, podemos proceder da seguinte forma.

Conforme os critérios estão disposto, iremos ordenar como mostrado abaixo.

$$s_1 < s_2 < s_3$$

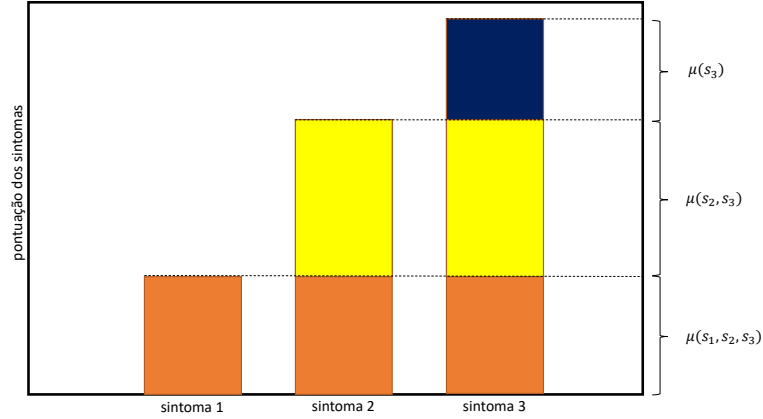


Figura 25 – Representação gráfica para três sintomas do paciente P_1

Assim, $x_{1(1)} = x_{11} = 0.3 < x_{1(2)} = x_{12} = 0.6 < x_{1(3)} = 0.95$

Dessa forma, o *score* dado pela integral de Choquet disponibilizado ao paciente 1 é:

$$CI_1 = 0.3 \cdot 1 + (0.6 - 0.3) \cdot \frac{1}{3} + (0.95 - 0.6) \cdot \frac{1}{3} = 0.5167$$

Para calcular a integral de Choquet do paciente 2, temos, $s_2 < s_1 < s_3$.

Então, vamos considerar a seguinte ordenação $x_{2(1)} = x_{22} = 0.40 < x_{2(2)} = x_{21} = 0.45 < x_{2(3)} = 0.85$, assim, o *score* gerado pelo paciente 2, vem a ser:

$$\begin{aligned} CI_2 &= x_{2(1)}\mu(c_1, c_2, c_3) + (x_{2(1)} - x_{2(2)})\mu(c_1, c_2) + (x_{2(3)} - x_{2(1)})\mu(c_3) \\ &= 0.40 \cdot 1 + (0.45 - 0.40) \cdot \frac{2}{3} + (0.85 - 0.45) \cdot \frac{1}{3} = 0.5667. \end{aligned}$$

Então, prosseguindo para o paciente P_3 , temos.

$$CI_3 = 0.1 \cdot 1 + 0.55\mu(c_1, c_2) + 0.95 \cdot \mu(2) = 0.1 + 0.55 \cdot \frac{2}{3} + 0.95 \cdot \frac{1}{3} = 0.7833$$

Dessa maneira, temos que os escores dos pacientes podem ser classificados como $P_1 = 0.5167$, $P_2 = 0.5667$, $P_3 = 0.7833$.

Portanto, para um prognóstico de doença, o médico poderá estabelecer, com base na experiência clínica, um limiar para um ponto de corte em que manterá uma classificação entre os casos positivos e negativos.

A determinação das medidas de capacidades de Choquet podem consistir em um problema de otimização quadrática com restrições, onde a função custo, sem ou com o termo de regularização, é dada por uma distância entre as saídas da integral de Choquet e os valores que explicam as preferências do tomador de decisão, onde é assumido o conhecimento da matriz de decisão e do vetor a ser estimado [106]. Dessa maneira, uma

forma de busca desses valores está baseada em problemas com abordagem com métodos supervisionados.

Para o propósito desta tese, a integral de Choquet empregada para demonstração dos resultados, será do caso da capacidade 2-aditivo apresentada no trabalho de [102].

1.3.4 Métricas de Desempenho

Nesta tese foram escolhidas algumas métricas para avaliar a qualidade da classificação de cada algoritmo de aprendizado de máquina: acurácia, sensibilidade, especificidade, *F1-Score*, acurácia balanceada, precisão e área sob a curva (AUC). A acurácia é dada como o número de amostras classificadas corretamente pelo total de amostras. Pode não ser adequada para dados desbalanceados. As métricas de sensibilidade e especificidade podem ser calculadas matematicamente como sendo a taxa de acerto para predições corretas de verdadeiros positivos e de verdadeiro negativo. O *F1-Score* é considerada a média ponderada entre a taxa de precisão e a sensibilidade, muito utilizada em conjunto de dados desbalanceados, onde F1 atinge seu melhor valor em 1 e a pior pontuação em 0. A acurácia balanceada é utilizada para lidar com conjuntos de dados desbalanceados e é definida como a média entre a taxa de sensibilidade e especificidade. A precisão é porcentagem de exemplos classificados como positivos que são realmente positivos.

A área sob a curva (AUC), pode ser visualizada com o auxílio do gráfico da curva ROC, que evidencia o desempenho do algoritmo com o critério de mudança na curva dependendo da taxa de verdadeiro positivo e falso positivo. A AUC representa o grau ou medida de separabilidade de classificação, quanto maior o AUC, melhor será o modelo na tarefa de realizar as predições.

As principais métricas apuradas para avaliação dos resultados desta tese foram a **Especificidade**, **Sensibilidade**, **Precisão**, **F1-Score** e **AUC**, pois apresentam maior grau de importância no auxílio em diagnósticos médicos a qual é estabelecida um ponto de corte que será ajustado conforme especificações ou experiências do profissional de saúde.

Acurácia	$ACC = \frac{TP + TN}{TP + TN + FP + FN}$
Sensibilidade	$TVP = \frac{TP}{TP + FN}$
Especificidade	$TVN = \frac{TN}{TN + FP}$
Precisão	$P = \frac{TP}{TP + FP}$
F1-Score	$F_1 = \frac{2TP}{2TP + FP + FN}$
Área sob a curva	$AUC = \int_0^1 ROC(t)dt, t \in [0, 1]$

2 O Método C3PO

Neste capítulo apresentamos a principal contribuição metodológica desta tese, que é sintetizada pelo método denominado C3PO¹ - *Clinical Probability Outcome*. O C3PO foi desenvolvido com o propósito de conciliar dois paradigmas de apoio à decisão para a predição de risco e diagnósticos na área médica: decisões baseadas em protocolos clínicos e decisões orientadas a dados.

Historicamente, a área médica utiliza *scores* a partir de ensaios clínicos específicos, que passam por várias etapas de validação [107]. Denominaremos essas ferramentas de "orientadas a protocolos". Por outro lado, na última década, temos acompanhado o desenvolvimento de ferramentas baseadas em aprendizado de máquina, que utilizam grandes volumes de dados para treinar algoritmos computacionais capazes de prever riscos e/ou diagnósticos relacionados aos mais variados desfechos clínicos [108]. Denominaremos tais ferramentas de "orientadas a dados".

No entanto, apesar de ferramentas orientadas a dados alcançarem altas taxas de acurácia, na prática, essa abordagem ainda encontra dificuldades de se estabelecer como ferramentas de apoio à decisão clínica. Em nossa visão, parte dessa dificuldade pode ser explicada por uma aparente dicotomia entre o raciocínio clínico e a saída de modelos de classificação supervisionados, preponderantes na modelagem de predição de risco e diagnósticos médicos.

Outra característica relevante sobre o desenvolvimento de ferramentas orientadas a dados é a grande variedade de métodos, abordagens e algoritmos para resolver um mesmo problema. Graças ao volume crescente de dados e o aumento da capacidade computacional, inúmeros métodos e suas variações têm sido propostos na literatura, em geral, competindo entre eles para obter métricas de desempenho cada vez melhores. Nesse sentido, o uso de técnicas de *ensemble*, que permitem combinar resultados de vários métodos têm se mostrado uma alternativa interessante que possibilita ao decisor conciliar estimativas advindas de diferentes fontes.

No entanto, um desafio ainda pouco explorado na literatura é a junção de métodos baseados em protocolos, com métodos orientados a dados, em particular, para aplicações em saúde. O desenvolvimento do Método C3PO contribui com a literatura científica nesse aspecto.

¹ O C-3PO é um androide do mundo fictício de Star Wars (Guerra nas Estrelas). É um "droide de protocolo"(intérprete e relações-sociais) fluente em 6 milhões de formas de comunicação.

2.1 Fundamentação do Método C3PO

Diversos trabalhos na literatura, como [109, 110], têm mostrado que o raciocínio clínico é fortemente baseado em evidências e no pensamento Bayesiano, mesmo que de forma intuitiva. Ao examinar um paciente, o médico coleta um conjunto X de evidências (sintomas, sinais, resultados de exames, histórico do paciente, etc) e combina seus conhecimentos de medicina com heurísticas mentais para estimar quais os diagnósticos mais prováveis, dado a ocorrência de X [111, 112, 113].

Para cada possível diagnóstico Y , o médico estima chance de Y ser positivo (ou negativo) frente ao conjunto de evidências X . Podemos denotar essa chance por $P(Y|X)$, o que é denominado de probabilidade a posteriori. A probabilidade a posteriori não pode ser medida diretamente, mas pode ser estimada em função de dois fatores: a incidência (a priori) de Y na população, que denotamos por $P(Y)$, e a verossimilhança do conjunto de evidências X nos pacientes com diagnóstico positivo, ou seja, a probabilidade de um paciente que tem diagnóstico positivo Y ter o conjunto de evidências X . No âmbito da estatística Bayesiana, essa probabilidade é denotada por $P(X|Y)$ [109, 110].

Com base nessas duas estimativas, o Teorema de Bayes [53] estabelece que:

$$P(Y|X) = \frac{P(Y)P(X|Y)}{P(X)} \quad (2.1)$$

Portanto, o diagnóstico médico é, em essência, probabilístico e dependente do conjunto de informações prévias que o médico conhece. Assim, médicos mais experientes são capazes de estimar de forma mais precisa as probabilidades $P(Y)$ e $P(X|Y)$, em função do conjunto de informações prévias (histórico de pacientes) a que foram submetidos.

É justamente neste ponto, que a análise de um grande volume de dados pode auxiliar no diagnóstico médico. Por exemplo, um médico que atua em um pequeno hospital de uma cidade do interior, poderia ter acesso às informações sistematizadas de todos os pacientes de um grande hospital, aumentando assim sua capacidade de estimar $P(Y)$ e $P(X|Y)$, sem a necessidade de ter examinado um grande número de pacientes.

Em função do raciocínio clínico ser probabilístico, é esperado que haja um natural desconforto por parte da classe médica quando confrontados com modelos de classificação que fornecem recomendações estritas, ou determinísticas, principalmente se forem utilizados métodos que dificultam a interpretabilidade dos resultados. Um exemplo disso é o uso de algoritmos tipo “caixa-preta” para a previsão de diagnóstico com base em modelos de classificação supervisionados. Em geral, o resultado desses modelos é expresso de forma determinística, quase sempre binária, pois tais métodos mapeiam um vetor de dados de entrada (as evidências) em uma das possíveis classes de resultados (0=negativo; 1=positivo) [114].

Como mencionamos na introdução desse capítulo, grande parte das soluções de aprendizagem de máquina não incorporam apropriadamente informações advindas de *scores* de risco, protocolos clínicos ou mesmo a preferência dos médicos.

O método C3PO busca contornar esses problemas por meio da combinação de técnicas de aprendizado de máquina e métodos de apoio à decisão multicritério, com o objetivo de mapear evidências em probabilidades.

2.1.1 Dois Paradigmas de Apoio à Decisão Clínica

O C3PO utiliza duas abordagens complementares que podem atuar em paralelo ou de formas independentes:

Orientação à dados históricos: essa abordagem utiliza um conjunto de dados históricos de pacientes, para os quais se conhece as evidências X e os respectivos diagnósticos Y , para treinar algoritmos de aprendizado de máquina supervisionados com o objetivo de estimar $P(Y|X)$.

O C3PO admite o uso individual ou a combinação de algoritmos de aprendizado de máquina, cuja saída é sempre expressa em termos de probabilidade. Ou seja, para cada classe é assumida uma função de densidade de probabilidade que estima $P(Y|X)$.

Desta forma, o médico pode interpretar essa probabilidade como um mapeamento das evidências de um paciente específico frente ao conjunto de dados históricos que o modelo teve acesso no treinamento. Assim, quanto maior e mais consistente for o conjunto de treinamento dos modelos, maior será a acurácia (sensibilidade, especificidade) do processo.

A escolha de um, ou mais, modelos de classificação deve considerar a natureza e o volume de dados históricos. Modelos como *Naive Bayes* ou técnicas derivadas de Árvore de Decisão podem oferecer boas estimativas para $P(Y|X)$ com algumas poucas centenas de dados históricos, enquanto modelos com estruturas mais complexas requerem mais dados [87]. Uma boa alternativa é a composição de modelos por meio de Comitês de Máquinas (*Ensembles*) [115].

O resultado na forma de probabilidade é uma característica essencial do C3PO, que o difere dos modelos determinísticos de diagnóstico. A saída em termos probabilísticos permite que o médico defina seu próprio limiar para a classificação em função da situação clínica que o interessa. Em um cenário de triagem, o médico pode estabelecer um limiar mais baixo, visando alta sensibilidade. Em um cenário de diagnóstico, pode-se estabelecer um limiar mais elevado, visando alta especificidade conforme são os ajustes do valor de limiar no intervalo $[0, 1]$ [116]. Na prática, isso significa que pode-se adotar um limiar diferente do usual 0,5 para estabelecer a regra de decisão que associa a probabilidade $P(Y|X)$ à classe. Numa situação de triagem, pode ser desejável configurar uma regra de

decisão que torne o método mais sensível. Assim, o médico pode optar por associar $Y = 1$, se $P(Y|X) < 0.3$, por exemplo.

Dessa maneira, o C3PO visa se aproximar do modelo mental utilizado por médico para compor um diagnóstico e servir como um método de apoio à decisão fundamentado em dados observados, protocolos clínicos e experiência do especialista. Para tanto, utiliza-se de uma perspectiva bayesiana para descrever os resultados em termos probabilísticos.

Orientação a protocolos: Em complemento à abordagem orientada a dados, o C3PO permite a incorporação de *scores* ou calculadoras de risco usualmente utilizadas na área médica. Além disso, propomos o uso de métodos de apoio multicritério (MCDM) para a inclusão de critérios clínicos apropriados para cada diagnóstico.

Para cada sintoma s_i o médico pode atribuir um peso w_i que representa a importância de s_i para o diagnóstico Y . Dessa forma, métodos clássicos de agregação determinísticas como *SAW* (*Simple Additive Weighting*), *TOPSIS* (*Technique for Order of Preference by Similarity to Ideal Solution*) ou Integral de Choquet podem ser utilizados para calcular *scores*.

Embora o médico possa utilizar o *score* multicritério como um agregador de evidências, o C3PO propõe que tal indicador seja convertido para probabilidade por meio da utilização de um método de aprendizado de máquina supervisionado, como regressão logística. Dessa forma, as saídas fornecidas pelos módulos de tomadas de decisão multicritérios ao especialista, em *scores*, podem ser quantificadas usando probabilidades de classificação.

2.1.2 Estrutura do C3PO

O método C3PO visa ser de fácil interpretabilidade, demandando blocos que irão ser apresentados pela abordagem orientado a dados e orientação pôr protocolos.

Dentre o conjunto de algoritmos de aprendizado de máquinas, podemos destacar os que possuem saídas em termos de probabilidades que são: Regressão Logística, Redes Neurais, *Gaussian Naive Bayes*, *Support Vector Machine*, *K-Nearest Neighbor*, *Gradient Boosting Classifier*, *Light Gbm*, *XGBoost Classifier*, *Random Forest Classifier* como visto no Capítulo 1 1.2.

Para os métodos de apoio à decisão multicritério para a modelagem de preferência, podemos destacar algumas técnicas: Métodos aditivos de agregação determinística como o SAW (*Simple Additive Weighting*), TOPSIS (*Technique for Order of Preference by Similarity to Ideal Solution*), que são métodos aditivos, dessa forma, a nossa pretensão é fazer a aplicação do método não-aditivo de decisão multicritérios.

Para tanto, precisamos determinar um parâmetro para as ponderações dos escores, dessa forma, a consulta aos especialistas são necessários para determinar essas constantes. O método C3PO é, portanto, um *framework* composto por dois módulos que permite integrar diferentes calculadoras de risco, métodos multicritérios e algoritmos de aprendizagem de máquina. Como saída o C3PO apresenta duas probabilidades derivadas dos respectivos módulos. A Figura 26 apresenta um esquema do método C3PO.

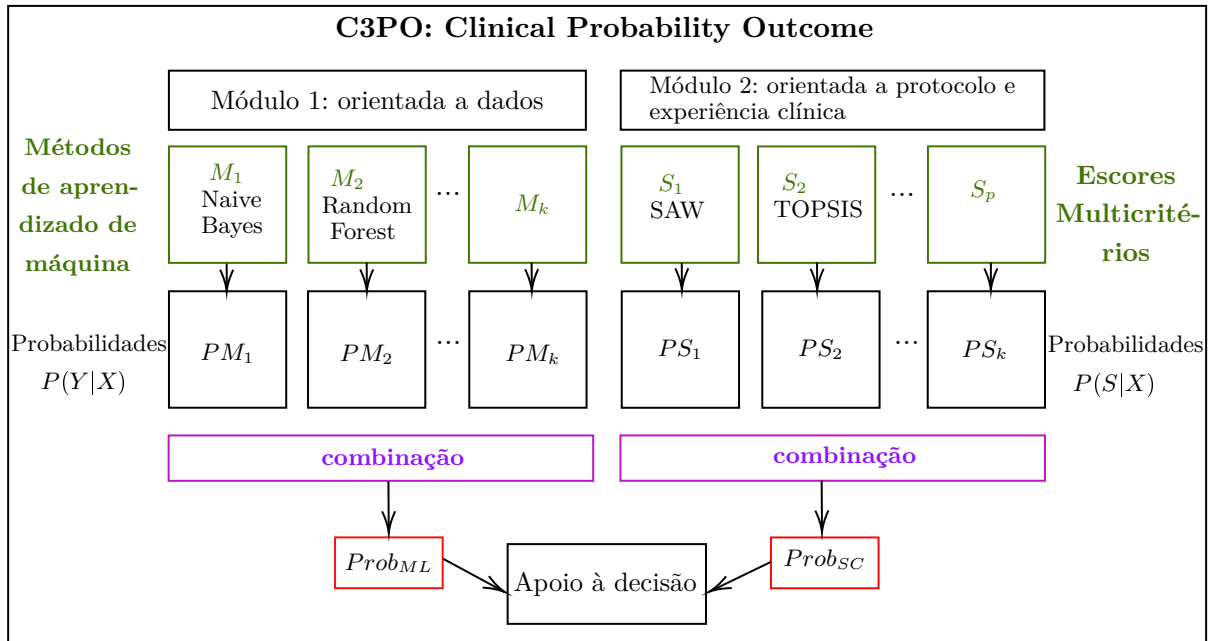


Figura 26 – Estrutura do C3PO

Uma vez escolhido um método de classificação, a estrutura do C3PO irá apresentar a saída correspondente em probabilidades, que poderá ser utilizada individualmente ou combinada com a saída de outros métodos, como representado na parte final do Módulo 1 da Figura 26.

O método C3PO utiliza como dados de entrada sinais, sintomas e dados laboratoriais dos pacientes. Para o Módulo 1 e para abordagem supervisionado do Módulo 2 também é necessário o conhecimento do desfecho real, ou seja, do diagnóstico associado a cada paciente. Tais informações devem ser oferecidas ao método no formato de dados numéricos tabulares. A Figura 27 apresenta um esquema do Módulo 1 do C3PO, com destaque para os dados de entrada.

Na Figura 28, tem-se um esquema do Módulo 2. Em um primeiro passo uma abordagem multicritério, ou calculadoras de risco, são utilizadas para a obtenção dos *scores* de risco que, em um segundo passo, podem ser convertidos em probabilidades por meio do uso da abordagem supervisionada. Dessa forma, para cada método de MCDA, podemos ter um *score* que posteriormente pode ser convertido em probabilidade.

Portanto, o método C3PO é um *framework* geral de apoio à decisão ao diagnóstico ou triagem médica, que combina diferentes métodos de aprendizagem de máquina,

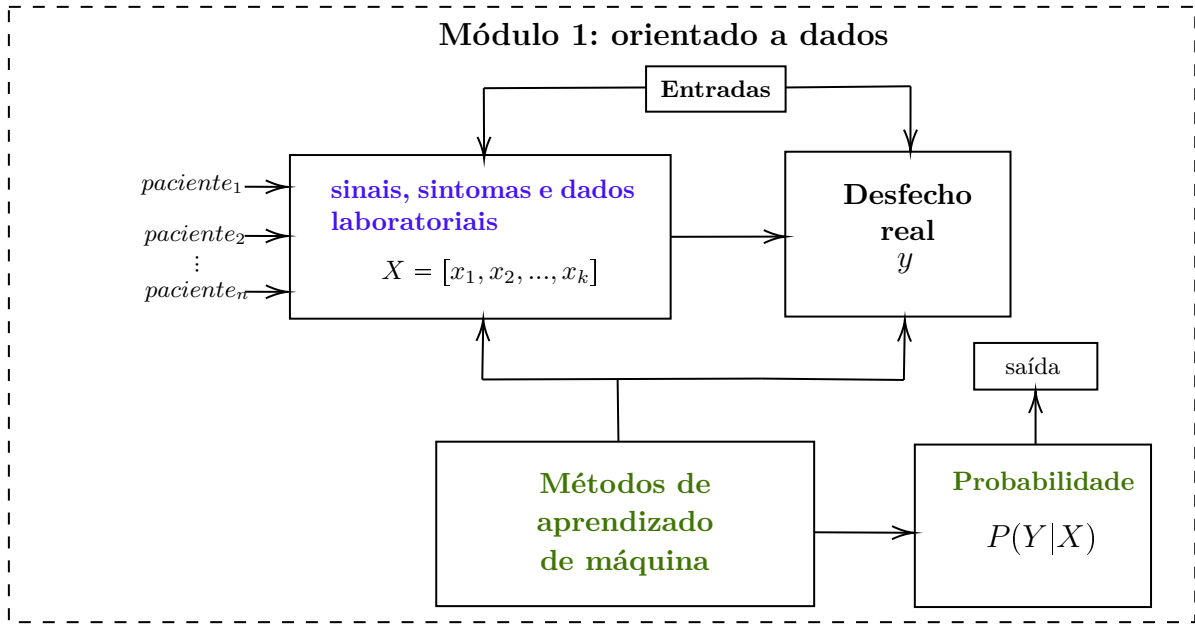


Figura 27 – Esquema do Módulo 1 do C3PO.

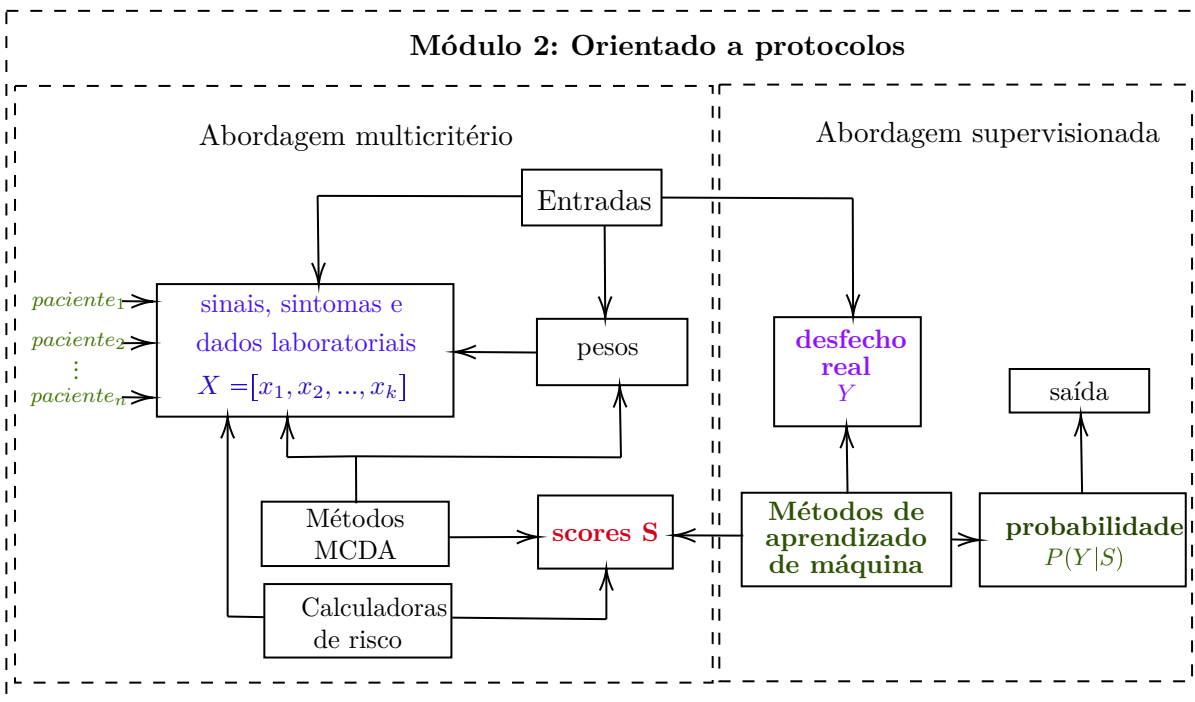


Figura 28 – Esquema do Módulo 2 do C3PO.

calculadoras de risco baseadas em protocolos clínicos e abordagem multicritério em forma de *ensemble*. A saída do C3PO é sempre expressa em termos de probabilidades, o que permite ao médico estabelecer seus próprios limiares de decisão, inclusive utilizando tais probabilidades como informação adicional para compor seu diagnóstico.

Em termos de interpretabilidade, o Módulo 2 do C3PO permite que os resultados sejam interpretáveis em termos dos pesos dos critérios ou das calculadoras de risco utilizadas. No Módulo 1 podem ser utilizados métodos interpretáveis, como regressão logística, ou

métodos baseados em regras. Por fim, é possível realizar um estudo futuro relacionando a probabilidade de saída com os sinais e sintomas de entrada, com intuito de calcular a importância de variável de entrada para o resultado final.

Na próxima seção apresentamos um novo método multilinear, que pode ser utilizado como uma alternativa para o cálculo de *scores* do Módulo 2 do C3PO, que permite interpretar os resultados por meio de pesos para cada variável de entrada e para todos os pares de coalizões multiplicativas.

2.2 Um Novo Método Interpretável: Método Multilinear por Coalizões Multiplicativas

Nesta seção apresentamos a segunda contribuição metodológica desta tese, que consiste em um método multicritério, que considera interação par-a-par das variáveis de entrada, permitindo o cálculo de um *score* aditivo que considera pesos individuais para cada variável e para suas coalizões multiplicativas.

Embora métodos como a integral de Choquet [102] ou modelos multilineares baseados em teoria de jogos [117] considerem a interação entre critérios (variáveis), tais métodos impõem restrições ao conjunto de pesos (ou capacidades) para atender hipóteses como monotonicidade e independência preferencial, que são típicas da área de apoio multicritério. No entanto, tais hipóteses não são necessariamente úteis. Por exemplo, um paciente que possui todos os possíveis sintomas em uma base de dados não é necessariamente o mais provável de ter diagnóstico positivo, visto que a ocorrência simultânea de muitos sintomas pode, inclusive, descartar um diagnóstico positivo.

O modelo introduzido a seguir preserva características presentes na versão 2-aditiva da integral de Choquet e flexibiliza as restrições do espaço de peso do modelo [117].

Assim como no modelo aditivo linear *SAW*, a importância de cada variável será representada por um peso ω_i , com $0 \leq \omega_i \leq 1$. A interação entre pares de sintomas s_i e s_j será representada por pesos α_{ij} , com $-1 \leq \alpha_{ij} \leq 1$, que podem modelar uma relação de sinergia ou redundância. Assim, caso a ocorrência simultânea de dois sintomas, s_i e s_j aumente a chance de um diagnóstico $Y = 1$, um valor $\alpha_{ij} > 0$ pode representar essa intensidade. Por outro lado, se a ocorrência simultânea desses sintomas reduz a chance do diagnóstico $Y = 1$, um valor $\alpha_{ij} < 0$ pode representar essa situação. Caso a interação dos sintomas não interfira no diagnóstico, um valor de $\alpha_{ij} = 0$ pode ser atribuído.

Assim, considerando um conjunto $s = \{s_1, s_2, \dots, s_m\}$ de variáveis que representam sinais, sintomas e resultados de exames ou dados laboratoriais utilizados para modelar um diagnóstico, e dois conjuntos de pesos, $\omega = \{\omega_1, \omega_2, \dots, \omega_m\}$ e $\alpha = \{\alpha_{12}, \alpha_{13}, \dots, \alpha_{m-1\ m}\}$,

o *score* de risco do Modelo Multilinear por Coalizões Multiplicativas (M2CM) é calculado da seguinte forma:

$$M2CM = F(s) = \sum \omega_i \cdot s_i + \sum \alpha_{ij} \cdot s_i \cdot s_j, \quad (2.2)$$

onde:

- $0 \leq \omega_i \leq 1$ representa o peso do i –ésimo sintoma
- $-1 \leq \alpha_{ij} \leq 1$ representa os pesos da interação entre os sintomas i e j
- $\omega_i + \omega_j \geq \alpha_{i,j}$ representa a interação entre dois sintomas i e j é menor do que a soma entre os dois sintomas isolados.

Os conjuntos de pesos ω e α podem ser atribuídos diretamente pelo médico, com base em sua experiência clínica ou literatura científica, ou serem estimados com base em histórico de pacientes através de um problema de programação matemática, como descrito a seguir.

2.2.1 Programação Quadrática para a Estimação de Pesos com Base em Dados em Abordagem Supervisionada

Programação quadrática são particularidade de problemas de otimização numérica que possui variadas aplicações, como ajuste de curvas em estatística, aprendizado de máquina para máquinas de vetores de suporte (SVMs) entre outros [118, 119].

Vamos considerar uma abordagem supervisionada cujo objetivo é minimizar o erro quadrático médio entre as avaliações obtidas pela estimação na função $F(s)$ e os valores reais de saída em y_i .

Pretendemos minimizar o erro cometido ao realizar a estimativa para os valores de y_i , através de $\min \sum (y_i - F(s))^2$, obedecendo as seguintes restrições $0 \leq w_i \leq 1$ e $-1 \leq \alpha_{ij} \leq 1$.

A partir da matriz que contém as entradas, chamada de M com m o número de colunas e n a quantidade de linhas dessa mesma matriz, iremos obter a sua matriz de expansão B com entradas sendo o produto dos elementos de M , ou seja, tomamos a coalizão entre pares de critérios. Dessa forma, o número de colunas da matriz expandida será $\frac{m \cdot (m - 1)}{2}$.

Em termos de representação na forma matricial, as entradas da matriz de sintomas, ou, matriz com os coeficientes de decisão é representada por,

$$M = \begin{bmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,m} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n,1} & s_{n,2} & \cdots & s_{n,m} \end{bmatrix}.$$

A matriz B , de coeficientes multiplicativos, é construída da seguinte maneira

$$B = \begin{bmatrix} s_{1,1}s_{1,2} & s_{1,1}s_{1,3} & \cdots & s_{1,1}s_{1,m} & \cdots & s_{1,2}s_{1,3} & \cdots & s_{1,2}s_{1,m} & \cdots & s_{1,m-1}s_{1,m} \\ s_{2,1}s_{2,2} & s_{2,1}s_{2,3} & \cdots & s_{2,1}s_{2,m} & \cdots & s_{2,2}s_{2,3} & \cdots & s_{2,2}s_{2,m} & \cdots & s_{2,m-1}s_{2,m} \\ \vdots & \vdots & \ddots & & & & & & & \vdots \\ s_{n,1}s_{n,2} & s_{n,1}s_{n,3} & \cdots & s_{n,1}s_{n,m} & \cdots & s_{n,2}s_{n,3} & \cdots & s_{n,2}s_{n,m} & \cdots & s_{n,m-1}s_{n,m} \end{bmatrix}.$$

Seja $P = [1|M|B]$ a matriz obtida pela concatenação da entrada, M , com a matriz expandida, B , e cuja a primeira coluna é um vetor de uns, que representa o intercepto do modelo de ajuste linear. A inclusão desta coluna é importante para modelar uma translação linear que representa o viés (*bias*) do modelo de aprendizagem. Com isso será necessária a inclusão de um peso adicional ω_0 correspondente a essa nova coluna. A Figura 29 apresenta esquematicamente a matriz P .

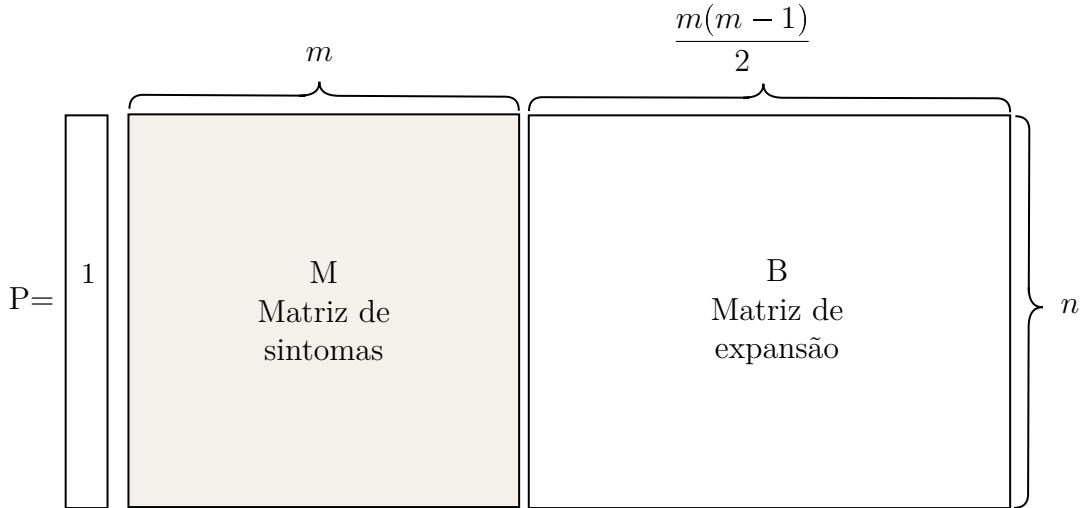


Figura 29 – A matriz P

Seja $x = [\omega_0, \omega, \alpha] = [\omega_0, \omega_1, \omega_2, \cdots, \omega_m, \alpha_{12}, \alpha_{13}, \cdots, \alpha_{m-1m}]$, o vetor de pesos, onde ω_0 representa o coeficiente do intercepto, ou *bias*. O problema de encontrar valores de ω e α que melhor se ajustam ao modelo pode ser escrito como um problema de otimização quadrática da seguinte forma:

$$\begin{aligned} &\underset{x}{\text{minimizar}} \quad \sum (y_i - F(s))^2 \\ &\text{sujeito a:} \quad 0 \leq \omega_i \leq 1 \\ &\quad \quad \quad -1 \leq \alpha_{ij} \leq 1. \end{aligned} \tag{2.3}$$

A função $\sum (y_i - F(s))^2$ pode ser expressa por

$$\begin{aligned}
 \|Px - y\|_2^2 &= (Px - y)^T (Px - y) \\
 &= x^T P^T Px - x^T P^T y - y^T Px + y^T y \\
 &\propto (1/2)x^T P^T Px - (1/2)x^T P^T y - (1/2)y^T Px \\
 &= (1/2)x^T (P^T P)x + (-P^T y)^T x.
 \end{aligned} \tag{2.4}$$

Dessa forma, fazendo $Q = P^T P$ e $q = -P^T y$, o problema 2.3 pode ser escrito na forma padrão de um problema de programação quadrática:

$$\begin{aligned}
 &\underset{x}{\text{minimizar}} \quad (1/2)x^T Qx + q^T x \\
 &\text{sujeito a:} \quad Gx \leq h
 \end{aligned} \tag{2.5}$$

Neste caso, a matriz Q é simétrica e semi-definida positiva, a matriz G e o vetor h devem ser adequadamente construídos para representar as restrições do problema 2.3. A matriz de restrições G possui $\frac{m(m+1)}{2} + 1$ colunas. Por exemplo, em um cenário com a matriz M , $m = 3$ colunas ou critérios e $n = 3$ amostras, temos:

$$G = \begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & -1 & -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & -1 \end{bmatrix},$$

e o vetor $h = [0, 0, 0, 0, 1, 1, 1, 0, 0, 0]^t$.

Em muitas aplicações é desejável normalizar o *score* resultante, $F(s) = Px$. Podemos impor, por exemplo, a restrição $0 \leq Px \leq 1$, que limitam os *scores* no intervalo $[0, 1]$. Neste caso, o problema de otimização é dado por

$$\begin{aligned}
 &\underset{x}{\text{minimizar}} \quad (1/2)x^T Qx + q^T x \\
 &\text{sujeito a:} \quad Gx \leq h \\
 &\quad \quad \quad 0 \leq Px \leq 1.
 \end{aligned} \tag{2.6}$$

Nos experimentos que realizaremos nesta tese com o método M2CM, sempre utilizaremos o problema de otimização 2.6.

Exemplo 2. *Vamos realizar uma aplicação para o problema em determinar se uma gestante contraiu COVID-19 de acordo com a indicação para três sintomas apresentados. A indicação dos sintomas estão conforme a paciente possui o sintoma (1) ou não (0). Realizar a predição.*

Tabela 9 – Sintomas de pacientes

	s_1	s_2	s_3	y
P_1	1	1	0	1
P_2	0	1	1	0
P_3	1	0	1	1

Por sua vez, a matriz de expansão B será dada como o produto das entradas da matriz extraída da Tabela 9.

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Dessa forma, a matriz expandida P é dada como a junção entre a matriz de coeficientes de entradas M com a matriz B . A matriz B possui três colunas de entradas para as variáveis, então a matriz G de restrições terá dimensão 10 linhas e 7 colunas, conforme já explicado anteriormente.

$$P = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Aplicando a otimização com suas restrições, obtemos como resultado os valores de pesos para os três sintomas avaliados $w = [0.36; 0.05; -0.42; 0.86; 1.34; 0.65]$, os três primeiros pesos são referentes aos sintomas individuais, os outros restantes estão relacionados com as coalizões. Utilizando o método multilinear, aplicando sobre a matriz ampliada P , encontramos como resultados para os *scores* dos pacientes os valores $scores = [0.99; 0.00; 0.99]$. Com base nessas pontuações de escores o especialista poderá estabelecer um ponto de corte para classificar pacientes.

Nos próximos capítulos utilizaremos o método C3PO para realizar alguns estudos de casos, com base em dados reais, para diagnósticos clínicos.

3 Aplicação do C3PO: Predição de COVID-19 em Mulheres Gestantes

Neste capítulo iremos apresentar alguns resultados relacionados a aplicação em um estudo de caso baseado em dados reais. Para exemplificar o funcionamento do *framework* C3PO em testes diagnósticos com abordagem na saúde da mulher gestante, tendo como entrada dados em formas de sinais e sintomas das bases de dados da REBRACO (Rede BRAsileira em estudos do COVID-19 em Obstetrícia) um estudo realizado por profissionais especialistas na área da saúde obstétrica. Nas apresentações dos resultados, utilizaremos os Módulo 1 e 2 da abordagem metodológica do Capítulo 2.

A base de dados da REBRACO tem como fonte um estudo que tem por objetivo identificar as principais complicações da infecção pelo corona vírus em mulheres gestantes. Esse estudo é prospectivo multicêntrico, com a finalidade de abordar os efeitos da pandemia de COVID-19 na população obstétrica e os desafios relacionados no combate à pandemia [24].

Nesse estudo, as gestantes que não apresentaram sintomas gripais ou assintomáticas foram excluídas da análise. Para a detecção da doença, foram realizadas o teste RT-qPCR nas gestantes que apresentavam determinados sintomas [24]. O conjunto de dados conta com um total de (1102) amostras (pacientes), sendo 387 (35,11%) com resultado positivo e 715 (64,89%) com resultado negativo.

Originalmente, a base da REBRACO possui setenta variáveis, das quais foram selecionadas treze variáveis através de conversas com os especialistas na área obstétrica e no uso de técnica seleção de variáveis. Dessa forma, as variáveis preditoras são as entradas: Dificuldade para deglutir, Febre38, Tosse, Congestao nasal, Dor de Cabeça, Mialgia, Dispneia, Vômito, Diarreia, hiposia/cheiro, Perda do paladar, Calafrios, Oxigenio95. A variável alvo a ser predita é se a mulher gestante foi diagnosticada com resultado da doença COVID-19, cujo valor está em termos binários como já dito na introdução dessa seção e indica se a gestante foi positiva ou negativa para a infecção pela doença.

O método C3PO, nesse primeiro estudo de caso, terá como objetivo auxiliar a decisão clínica por meio das suas saídas do Módulo 1 e 2. Todos os resultados foram simulados obedecendo ao processo decisório com $threshold=0.5$, no entanto, esse ponto de corte pode ser alterado conforme necessidade do profissional. Os resultados são apresentados através de tabelas, onde os algoritmos descritos no capítulo 1.2 são referenciados como siglas dos modelos de algoritmos. As siglas são dadas como *logist* para a Regressão Logística, *gbm* para *Gradient Boost Machine*, *xgbm* para *XGBoost*, *rf* para o método *Random Forest*,

mlp para *Multilayer Perceptron*, *ada* para *Adaptive Boosting*, *svm* para *Support Vector Machine*, *nb* para *Naive Bayes* e *knn* para *K-Vizinhos mais Próximos*. Para os resultados, conforme o Módulo 2, serão apresentados como Aditivo Linear, TOPSIS, Integral de Choquet e M2CM.

3.1 Apresentação dos Resultados

Iremos apresentar os resultados da aplicação do método C3PO na base da REBRACO. Cada resultado de saída para as gestantes, conforme o módulo de métodos é aplicado, foram transformados em probabilidades, dessa forma, para cada saída, temos probabilidades diferentes, o que possibilita ao decisor final associar em qual ponto de corte atribuir uma tomada de decisão.

Previamente, os resultados serão dados pelo Módulo 1 do C3PO, da abordagem orientada a dados por métricas padrões como AUC, sensibilidade, especificidade. Esses resultados estão dispostos da seguinte maneira:

- saídas de cada algoritmo de aprendizado em probabilidades;
- agregando todas as saídas em uma única por regressão logística;
- agregando as saídas par-a-par e extraíndo a probabilidade geral por regressão logística, somente pelo limiar de corte (sem retreinamento da regressão logística);
- e por fim, a aplicação de um comitê de máquina com os melhores preditores.

Em todos os processos, aplicamos cinco iterações de divisão da base de dados em treinamento e teste de forma aleatória.

Posteriormente, teremos os resultados para o Módulo 2 da orientação a protocolos. Esses resultados estão dispostos da seguinte maneira: aditivo linear (*SAW*), TOPSIS, Integral de Choquet e Método Multilinear por Coalizões Multiplicativas (M2CM). Para o método M2CM, vamos realizar a comparação com alguns exemplos de métodos de aprendizado de máquina e de apoio multicritério por meio da visualização da mediana.

3.2 Resultado para o Módulo 1

As variáveis de entradas foram separadas em um conjunto de dados para treinamento e teste. Vale ressaltar que a base de dados pode ser dividida em formas diferentes aleatoriamente. No caso do presente estudo, definimos para treino e teste 70% e 30% respectivamente. Assim, a Tabela 10 mostra a base de saídas do Módulo 1 de probabilidades de treinamento e a Tabela 11 a base saídas de probabilidades de teste,

segundo cada algoritmo de aprendizagem de máquina em uma das etapas de separação da base.

Tabela 10 – Tabela de probabilidades gerada para o treino na base da REBRACO

logist	gbm	lgbm	xgbm	rf	mlp	ada	svm	nb	knn	COVID-19
0.89	0.49	0.68	0.59	0.47	0.48	0.50	0.96	1	0.4	1
0.41	0.73	0.76	0.73	0.73	0.73	0.49	0.29	0.00	0.8	1
0.33	0.27	0.27	0.29	0.26	0.25	0.49	0.29	0.00	0.2	0
0.20	0.20	0.20	0.19	0.20	0.20	0.49	0.29	0.00	0.2	0
...
0.22	0.33	0.24	0.25	0.24	0.35	0.40	0.42	0.15	0.26	0

Tabela 11 – Tabela de probabilidades gerada para o teste na base da REBRACO

logist	gbm	lgbm	xgbm	rf	mlp	ada	svm	nb	knn	COVID-19
0.78	0.47	0.90	0.93	0.83	0.99	0.50	0.77	0.47	0.6	1
0.20	0.20	0.20	0.19	0.20	0.20	0.49	0.29	0.00	0.2	0
0.47	0.12	0.19	0.24	0.12	0.14	0.49	0.29	0.11	0.00	0
0.34	0.75	0.71	0.60	0.75	0.77	0.49	0.77	0.04	0.8	1
...
0.36	0.42	0.85	0.65	0.78	0.78	0.55	0.84	0.25	0.85	1

3.2.1 Resultados Individuais para cada saída do C3PO

Os primeiros resultados, Tabelas 12, 13 e 14 levam em consideração as saídas em probabilidade dos modelos de aprendizado de máquina. E essas saídas são retreinadas com o algoritmo de regressão logística, como uma forma de agregação de saídas, estabelecendo um limiar de corte de 0.5 em cinco iterações de separação da base de treinamento e teste de forma aleatória.

Tabela 12 – Resultados dos valores de AUC para cinco divisões da base

iteração	logist	gbm	lgbm	xgbm	rf	mlp	ada	svm	nb	knn
1 ^a	0.83	0.85	0.86	0.86	0.84	0.86	0.83	0.84	0.81	0.78
2 ^a	0.86	0.87	0.87	0.87	0.87	0.87	0.86	0.87	0.85	0.80
3 ^a	0.87	0.81	0.85	0.86	0.82	0.82	0.82	0.82	0.84	0.77
4 ^a	0.88	0.83	0.86	0.87	0.85	0.85	0.88	0.87	0.88	0.80
5 ^a	0.85	0.86	0.86	0.86	0.85	0.84	0.85	0.87	0.86	0.78
Média	0.86	0.84	0.86	0.87	0.84	0.85	0.85	0.85	0.85	0.79

Para as taxas de verdadeiros positivos e verdadeiros negativos, verificamos os seguintes valores, que foram encontrados através da realização de cinco iterações de

Tabela 13 – Resultados dos valores de taxas de verdadeiros positivos das saídas individuais

iteração	logist	gbm	lgbm	xgbm	rf	mlp	ada	svm	nb	knn
1 ^a	0.70	0.67	0.74	0.72	0.67	0.70	0.00	0.67	0.67	0.51
2 ^a	0.70	0.80	0.77	0.74	0.77	0.79	0.00	0.69	0.66	0.59
3 ^a	0.65	0.70	0.72	0.71	0.71	0.68	0.00	0.62	0.66	0.62
4 ^a	0.72	0.71	0.72	0.76	0.72	0.73	0.00	0.67	0.64	0.51
5 ^a	0.78	0.79	0.79	0.79	0.79	0.78	0.28	0.80	0.66	0.63
Média	0.71	0.73	0.75	0.74	0.73	0.73	0.05	0.69	0.66	0.57

separação da base de dados de forma aleatória, e com regressão logística como segue na Tabela 13 e 14

Para as taxas de verdadeiros negativos, temos os seguintes resultados como ilustrado na Tabela 14.

Tabela 14 – Resultados dos valores de taxas de verdadeiros negativos da saída individuais

iteração	logist	gbm	lgbm	xgbm	rf	mlp	ada	svm	nb	knn
1 ^a	0.88	0.92	0.92	0.90	0.92	0.92	1.00	0.93	0.84	0.90
2 ^a	0.87	0.81	0.84	0.82	0.83	0.82	1.00	0.92	0.89	0.92
3 ^a	0.88	0.86	0.88	0.88	0.86	0.87	1.00	0.88	0.88	0.91
4 ^a	0.94	0.92	0.92	0.91	0.92	0.92	1.00	0.94	0.95	0.92
5 ^a	0.79	0.86	0.81	0.82	0.86	0.87	0.98	0.83	0.94	0.87
Média	0.87	0.87	0.87	0.86	0.87	0.88	0.99	0.9	0.9	0.9

3.2.2 Agregando todas as saídas

Agregando os algoritmos de aprendizado de máquina em um conjunto de saídas probabilísticas com uma regressão logística, encontramos como valor máximo para a AUC de **0.85** e taxa de verdadeiro positivo de **0.76** e para verdadeiro negativo de **0.89**.

3.2.3 Agregando as Saídas em Pares

Outra forma de avaliar as combinações de saídas do método C3PO para a predição de COVID-19 em gestantes, é utilizar a combinação par-a-par entre as saídas e agregar os resultados por uma regressão logística. Dessa forma, teremos dez saídas em termos de pares de probabilidades. Na combinação par-a-par, a matriz triangular superior de probabilidades que possui dimensão de $[10, 10]$, com dez algoritmos e 45 combinações de resultados. Os resultados de AUC para a combinação par-a-par das saídas são como segue na Tabela 15 e as taxas de TVP e TVN conforme as Tabelas 16 e 17.

Observamos que a média de AUC na Tabela 15 está acima de 85%. Para as métricas de taxa de verdadeiro positivo e negativo, a média fica em torno de 72% e 90%.

Diante disso, combinando em pares as saídas, podemos antever melhora ao resultado, isso ocorre devido a alguns métodos terem melhor sensibilidade, enquanto os outros possuem especificidade aumentada.

Tabela 15 – Resultados para AUC

[illegible]

Tabela 16 – Taxa de verdadeiro positivo para combinação par-a-par

[illegible]

Tabela 17 – Taxa de verdadeiro negativo para a combinação par-a-par

[illegible]

3.2.4 Controle Somente pelo Limiar de Corte

Da mesma maneira, uma forma semelhante de avaliar as respostas e apurar qual melhor performa os resultados, podemos proceder apenas pelo controle de corte do *threshold* ou limiar, sem realizar a agregação de saídas com um algoritmo de aprendizado como processo de retreinamento por regressão logística como estávamos fazendo anteriormente, assim, observamos os seguintes resultados conforme a Tabela 18 para os resultados de AUC, TVP (Taxa de Verdadeiro Positivo) e TVN (Taxa de Verdadeiro Negativo).

Tabela 18 – Resultado para as saídas individuais, controlando pelo *threshold* igual a 0.5

	AUC	TVP	TVN
logist	0.88	0.75	0.94
gbm	0.84	0.77	0.92
lgbm	0.87	0.79	0.92
xgbm	0.87	0.79	0.91
rf	0.85	0.78	0.92
mlp	0.85	0.78	0.92
ada	0.88	0.76	0.91
svm	0.87	0.8	0.94
nb	0.88	0.64	0.96
knn	0.81	0.63	0.92

Note que a taxa de verdadeiro negativo alcança valores acima de 0.9, o que a torna com a taxa de AUC um fator que demonstra uma alta taxa de acertos em diagnósticos de COVID-19 em gestantes. A Tabela 18 leva apenas em consideração as probabilidades de saídas da base de teste para realizar a classificação e considerando o ponto de corte de 0.5.

3.2.5 Comitê de Máquina

Um comitê de máquina, é um sistema que realiza uma combinação de modelos de aprendizado de máquina com o princípio de obter, através de procedimentos de votos, o melhor desempenho de resultados. Da mesma maneira, consideramos o limiar de corte ou *threshold* para os votos em 0.5 da probabilidade de saída, ou seja, facilmente poderá ser alterada conforme a necessidade de diagnóstico ou triagem médica.

Portanto, alcançamos, depois de algumas parametrizações e ajustes do modelo de comitê de máquina, em que agregamos os quatro algoritmos que trouxeram a melhor AUC.

- Taxa de verdadeiro positivo: 0.86
- Taxa de Verdadeiro negativo: 0.86

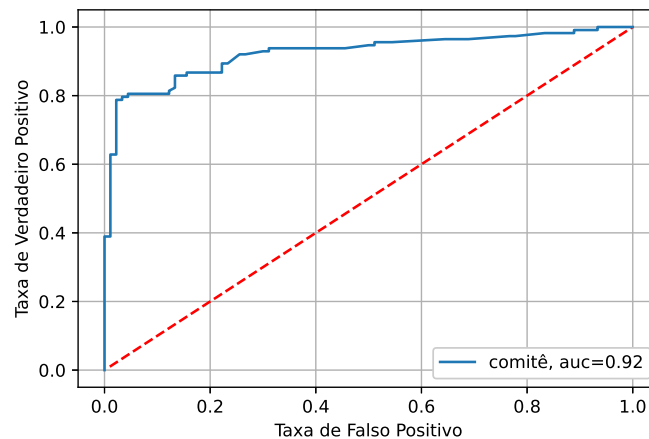


Figura 30 – Curva ROC com AUC para combinação dos melhores classificadores

A curva ROC obtida em 30 mostra, a confiabilidade do método de predição de COVID-19 em gestantes por meio da base de dados da REBRACO. Além da AUC (Área sob a curva), as taxas de verdadeiro positivo e verdadeiro negativo com valores de 0.86 ambas as medidas. Outras métricas que podemos mencionar para complementar o estudo é a acurácia e f1-score que tiveram valores de 0.85 e 0.87 respectivamente.

Portanto, para o Módulo 1 do C3PO obtemos métricas de desempenhos do método considerando um parâmetro de limiar de 0.5, assim, conforme a probabilidade entregue ao especialista, fica a seu próprio critério decidir pelo melhor controle desse parâmetro de ajuste.

3.3 Resultado para o Módulo 2

Nesta seção iremos estudar os resultados das saídas pertinentes ao campo relacionado com as tomadas de decisões por meio do Módulo 2, que consiste na orientação a protocolos clínicos para o *framework* do C3PO. Nesse estudo de caso, os métodos baseados em protocolos clínicos são aplicados para decisões em diagnósticos de COVID-19 em gestantes utilizando um conjunto de valores de pesos da seguinte forma: importâncias das variáveis de sintomas nas gestantes, obtidas no processo de rotinas de aprendizado de máquina, triagem clínica fornecidas pelos médicos especialistas, e pesos calculados dos sintomas atribuídos pela literatura médica. Outra maneira de se obter os pesos é o procedimento de ajuste por decida do gradiente, em que nesta seção será chamada apenas de pesos w . Com posse desses dados de pesos sistematizados em uma matriz tabular, e normalizado pela *norma-2*, aplicamos o processo através do Módulo 2 do C3PO na base de dados da REBRACO, como serão apresentados nesta seção.

Para os resultados são aplicados, em primeiro momento, os métodos *SAW*, *TOPSIS*, na base de dados da REBRACO, pois permitem uma melhor interpretabilidade do

mecanismo interno de cálculos. Na Tabela 19, ilustramos uma parte dos *scores* multicritérios.

Tabela 19 – Tabela de scores de preferências

Importância	Triagem clínica	Pesos Calculados	w
2	1	2,00	0,0227
10	1	2,67	0,1962
5	1	2,67	0,1739
2	0	1,67	0,0632
5	1	2,40	0,0078
...

A Tabela 19 possui três colunas que representam os pesos, contendo 13 linhas correspondentes a cada critério de sinais, sintomas e dados laboratoriais. Dessa forma, para cada critério apresentado na base de dados da REBRACO, podemos ter um vetor de pesos conforme a coluna da tabela escolhida ou a sua combinação.

Como visto na Tabela 19, observamos que a primeira coluna caracteriza o grau de importância em que os sintomas ou sinais aparecem nos métodos visto em orientação de dados. A coluna Triagem clínica representa a seleção dos pacientes mais prováveis a doença consoante o sintoma apresentado e, a coluna Pesos Calculados como uma ponderação do grau clínico atribuído pelo médico especialista em literaturas médicas.

Para realizarmos uma ponderação com melhor desempenho de classificação para os métodos multicritérios que serão abordados, encontramos um vetor de pesos ajustados a base de dados da REBRACO através do gradiente descendente linear. Esse vetor irá compor a quarta coluna da Tabela 19.

3.3.1 Método Aditivo Linear

O método de agregação *SAW* ou método aditivo linear é a primeira técnica empregada no processo de cálculo dos *scores* do estudo de caso para classificação probabilística de COVID-19 em gestantes. Esse método é baseado no conceito de somatório ponderado, em que o conjunto finito de amostras, avaliadas de forma determinística em relação a um conjunto finito de critérios. Dessa forma, dados os pesos da Tabela 19, iremos proceder os seguintes cálculos de *scores*. Como os pesos w_1, w_2, w_3, w_4 são constantes escalares associados aos critérios, vamos aplicar o procedimento multiplicativo $s(x) = \sum_{i=1}^n w_i s_i$ e com os valores de cada *scores* calculados por $\sum_{i=1}^n w_i = 1$, com n o número de amostras.

Na Tabela 20, apresentamos os resultados do processo multiplicativo e nomeamos as colunas como o conjunto de $score_1, score_2, score_3, score_4$, e teste da doença

COVID-19 (binário).

Tabela 20 – Os *scores* extraídos do processo multiplicativo

$score_1$	$score_2$	$score_3$	$score_4$	$Covid - 19$
0.0327	0.1	0.1128	0.0411	0
0.0327	0.1	0.0698	0.0486	0
0.0327	0.0	0.1128	0.0411	0
...
0.9672	0.1	0.9301	0.9513	1

Com posse dos resultados, podemos realizar uma agregação dos *scores* individualmente ou em combinação, ambos os procedimentos através de uma agregação por regressão logística. As probabilidades geradas pelas saídas de *scores* por método aditivo linear é apresentada conforme a Tabela 21 e a correlação entre as probabilidades geradas por cada *score* e a variável dependente y são representadas na Figura 31.

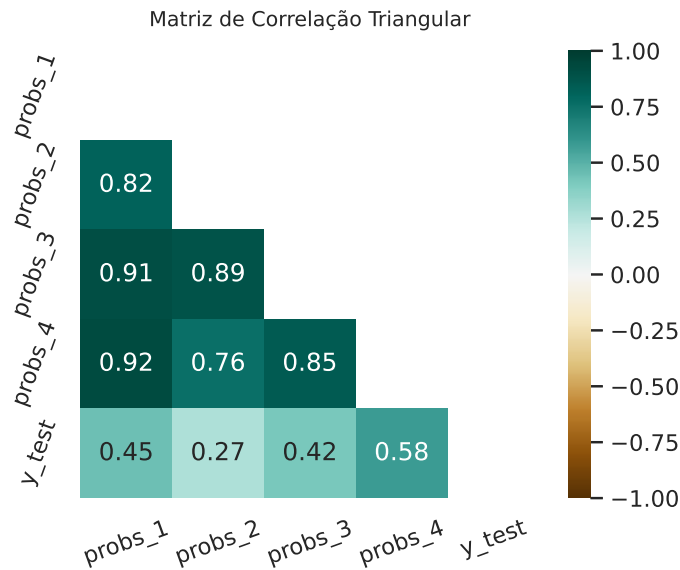


Figura 31 – Correlação das saídas em probabilidades do método aditivo

Tabela 21 – Probabilidades geradas pelo método SAW

$probs_1$	$probs_2$	$probs_3$	$probs_4$	$probs_{agreg}$	$COVID_{Test}$
0.4535	0.5329	0.5550	0.3359	0.3364	1
0.7313	0.6243	0.6947	0.7972	0.8056	1
0.4720	0.6243	0.4684	0.4608	0.4489	0
...
0.3471	0.3498	0.3727	0.2807	0.2738	0

Na etapa de regressão logística, para agregar o método aditivo linear, realizamos a separação dos valores constante na Tabela 20 para treinamento do algoritmo em 70%.

Para cada coluna de probabilidade ($probs_1, probs_2, probs_3$) foram calculadas por um modelo de regressão logística, permitindo uma transformação de *scores* em probabilidades. Por fim, a última coluna $probs_{agreg}$ considera a aplicação da regressão logística em todas as colunas da Tabela 20.

Variando o ponto de corte ou limiar no intervalo de $[0, 1]$, obtemos diferentes valores para as métricas de taxas preditivas. Então, estabelecendo um limiar de corte de 0.5, obtemos, como resultado para a taxa de verdadeiro positivo, verdadeiro negativo e área sob a curva (AUC), valores que se encontram na Tabela 22.

Para melhor ordenamento dos resultados, primeiro realizamos o cálculo de predição olhando para os *scores*, conforme a primeira parte do Módulo 2 na Figura 28 do Capítulo 2, e conforme os valores constantes na Tabela 21, individualmente, ou seja, o primeiro resultado será atribuído para a coluna com o $probs_1$, usando os pesos das importâncias das variáveis e extraímos uma nova probabilidade aplicando a regressão logística. As demais avaliações seguem os mesmos passos de procedimentos já realizados para cálculo das métricas.

Tabela 22 – Resultado para o método aditivo linear

	TVP	TVN	AUC	ACC
$Probs_1$	0.71	0.61	0.76	0.67
$Probs_2$	0.78	0.3	0.64	0.58
$Probs_3$	0.72	0.67	0.74	0.70
$Probs_4$	0.72	0.81	0.84	0.76
Probs	0.73	0.79	0.83	0.76

Com base na tabela de resultados para o método SAW 22, observamos que as probabilidades de saídas, $Probs_4$, e $Probs$ obtiveram melhor desempenho em relação à área sob a curva AUC e a acurácia (ACC) de ambos apresentam valores acima de 75%.

3.3.2 Método TOPSIS

O outro método que iremos aplicar no Módulo 2 é o TOPSIS. E, no caso da presente aplicação, usará a mesma matriz de pesos da introdução aos resultados do Módulo 2, seção 3.3. No método TOPSIS a base de dados da REBRACO, sem a variável objetivo, será normalizada pela norma euclidiana em cada colunas de critérios, realizando todas as etapas empregadas no algoritmo e retornando os *scores* das amostras.

No conjunto de amostras de pacientes gestantes, no método TOPSIS, chamaremos alternativas, dessa maneira, podemos calcular a alternativa ideal positiva A^+ e negativa A^- e as distâncias d^+ e d^- de cada paciente gestante para as alternativas ideais e calculando *scores*, Tabela 23.

Tabela 23 – Scores do método TOPSIS

$score_{c1}$	$score_{c2}$	$score_{c3}$	$score_{c4}$	$Teste_{Covid}$
0.0868	0.2202	0.2779	0.0957	0
0.0777	0.0000	0.1674	0.1000	0
...
0.9222	1.000	0.8325	0.8999	1

Logo em seguida, calculando as probabilidades de classificação de cada amostra com a aplicação do método de regressão logística sobre a base da Tabela 23, utilizando um limiar de 0.5, encontramos o desempenho do método TOPSIS.

Com a probabilidade para cada critério como variável preditiva, podemos verificar o grau de relacionamento entre cada entrada par-a-par por uma matriz de correlação de Pearson, Figura 32.

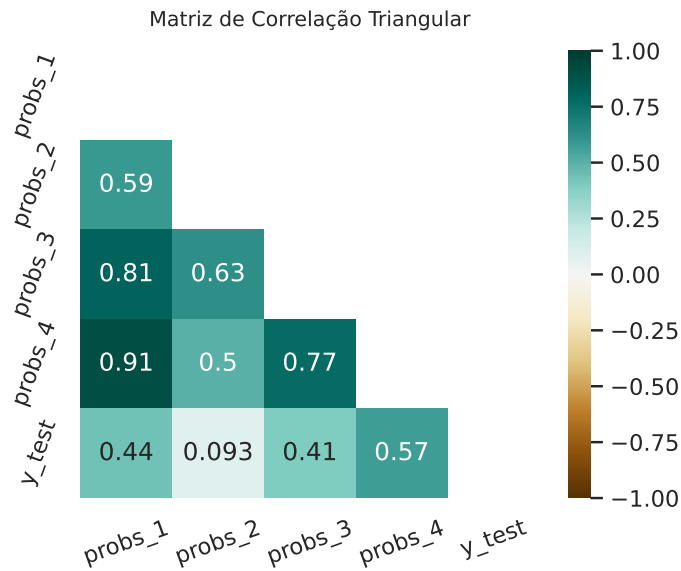


Figura 32 – Matriz de correlação entre as probabilidades do método TOPSIS

Tabela 24 – Probabilidades dos scores por regressão logística

$probs_1$	$probs_2$	$probs_3$	$probs_4$	$probs$	$COVID$
0.4292	0.5640	0.5623	0.3010	0.3075	1
0.7069	0.5803	0.6565	0.7025	0.7327	1
0.4874	0.6680	0.4665	0.4577	0.4318	0
0.7531	0.6396	0.6661	0.8516	0.8568	1
...
0.5298	0.3221	0.3698	0.4992	0.4983	1
0.4131	0.3221	0.4888	0.2759	0.2925	0

Cada coluna na Tabela 24 representa as saídas de probabilidades considerando

os quatro diferentes pesos, a coluna *probs*, representa a combinação das outras três colunas de probabilidades.

A Figura 33 mostra a distribuição de probabilidades no método TOPSIS para o critério de coluna *probs* da Tabela 24.

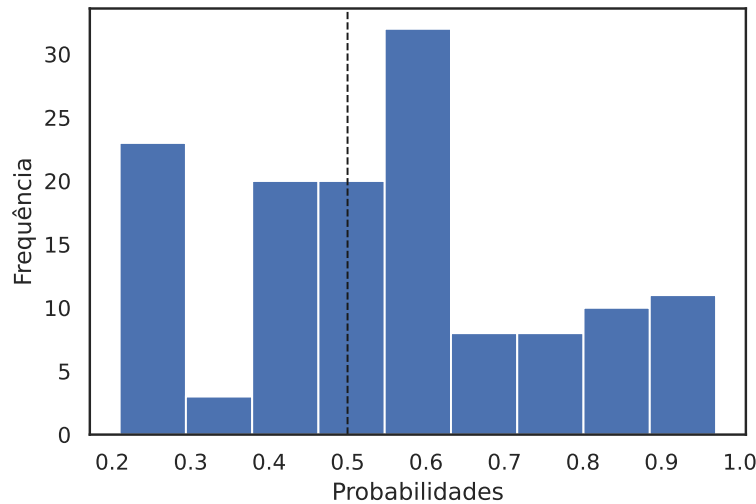


Figura 33 – Distribuição de frequência de probabilidades

Com essa configuração, temos como resultado, utilizando as saídas probabilísticas e mantendo o ponto de limiar de 0.5, conforme as métricas de desempenho apresentadas na Tabela 25.

Tabela 25 – Resultados para o método TOPSIS

	TVP	TVN	ACC	AUC
$Probs_1$	0.74	0.61	0.69	0.76
$Probs_2$	0.82	0.12	0.53	0.55
$Probs_3$	0.79	0.49	0.67	0.74
$Probs_4$	0.78	0.63	0.72	0.84
$Probs$	0.79	0.63	0.73	0.82

Os valores, que estão em negrito mostram os melhores desempenhos para as métricas investigadas. Observamos que os resultados referentes a probabilidade 4, ou seja, $Probs_4$, gerada pelos pesos ajustados do decida do gradiente linear, apresenta melhor desempenho para o método TOPSIS.

A Figura 34 representa as disposições da variação do limiar e a influência para a regulação da taxa de verdadeiro positivo através da curva ROC. Na curva ROC, observamos que a área sob a curva, AUC, alcança valores próximos aos encontrado pelo método SAW.

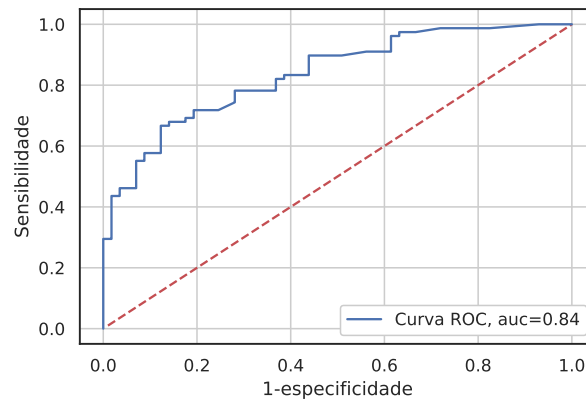


Figura 34 – Curva ROC para um preditor com TOPSIS

3.4 C3PO: Aplicação do Modelo Multilinear por Coalizão Multiplicativa

Nesta seção iremos aplicar, para o cálculo dos *scores* de risco, o Modelo Multilinear por Coalizões Multiplicativas (M2CM) concebido na seção 2.2 para estudo da REBRACO na predição de COVID-19 em gestantes e realizaremos as comparações com os outros modelos já aplicados nesta tese.

Para verificar o desempenho do modelo multilinear (M2CM), vamos comparar a medianas dos 1001 resultados de treinamento e teste separados aleatoriamente do Módulo 1 do C3PO, a saber: Regressão Logística, XGBoost, e os do Módulo 2 da abordagem multicritérios orientada a protocolos clínicos: integral de Choquet, Aditivo Linear e TOPSIS. Após, será extraída as métricas de acurácia, especificidade, sensibilidade e área sob a curva-AUC.

Nos resultados apresentados, vamos analisar em dois cenários, o primeiro com a base de dados da REBRACO desbalanceada em suas classes de saídas (*a*), e o segundo com a base balanceada (*b*).

A técnica de tornar a base balanceada é baseada na subamostragem com base no algoritmo de edição de amostragem dos vizinhos mais próximos. O conjunto de dados é removido de amostras que não concordam “o suficiente” com sua vizinhança. Para cada amostra da classe a ser subamostrada, são computados os vizinhos mais próximos e se o critério de seleção não for atendido, a amostra é removida [63].

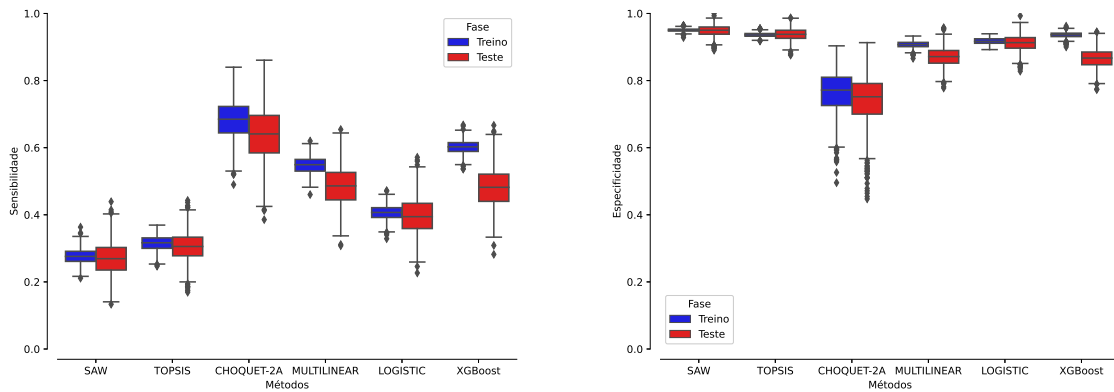
3.4.1 Resultados

Para os primeiros resultados, vamos considerar a base de dados da REBRACO de forma em que as classes não estão distribuídas de maneira equilibrada adequadamente, ou seja, desbalanceadas. Assim, aplicamos a técnica mencionada anteriormente para

compararmos dois contextos de cenários (a) e (b).

Na Figura 35(a), de forma comparativa, podemos observar a sensibilidade de treino e teste e os seus respectivos métodos. Para o modelo M2CM com a base de dados da REBRACO sem a estratégia de balanceamento, os resultados estão dentro da faixa comparáveis aos modelos de aprendizado de máquina. Os métodos SAW e TOPISIS possuem baixa sensibilidade se comparado com a generalização em treino e teste. A integral de Choquet obteve maior sensibilidade.

Na mesma linha, na Figura 35(b), mostramos os resultados em relação à especificidade dos modelos, verificamos que o método da integral de Choquet não obteve desempenho conforme os demais métodos. Os métodos, TOPSIS, SAW, Multilinear, Logística e XGboost lograram mediana de especificidade altas, situação causada pelo desbalanceamento da base.



((a)) Sensibilidade

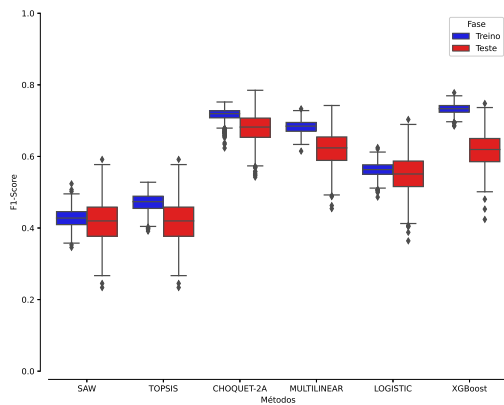
((b)) Especificidade

Figura 35 – Representação em bloxplot para o primeiro cenário

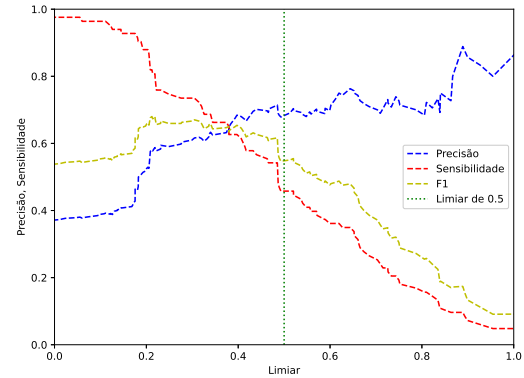
A precisão e a sensibilidade sofrem variações conforme vamos controlando o limiar de decisão. Na Figura 36(b) observamos que um limiar igual a 0.38 estabelece uma igual precisão e sensibilidade. Para maiores sensibilidades são encontradas em limiares menores do que 0.38. Vamos analisar o segundo cenário (b), onde a base é balanceada, ou seja, retiramos algumas amostras que possuem dados ruidosos. Dados esses são editados ou removidos das amostras através de um método que realiza *undersampling*.

Os primeiros resultados para o cenário (b) representado na Figura 37(a), são mostrados os resultados para a sensibilidade dos modelos. Novamente, as medianas entre integral de Choquet e M2CM são semelhantes. A Figura 37(b), em que mostramos os resultados em relação à especificidade dos modelos. Observamos, que o TOPSIS, e SAW possuem uma amplitude de diferença entre a especificidade e sensibilidade alta, isso mostra que ambos os modelos estão sujeitos ao sobreajuste, outra característica.

Para a métrica de *F1-Score*, o gráfico da Figura 36(a) evidencia o comporta-

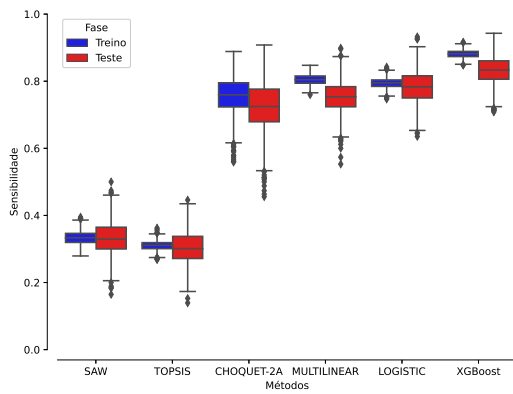


((a)) Boxplot para a métrica de F1-Score

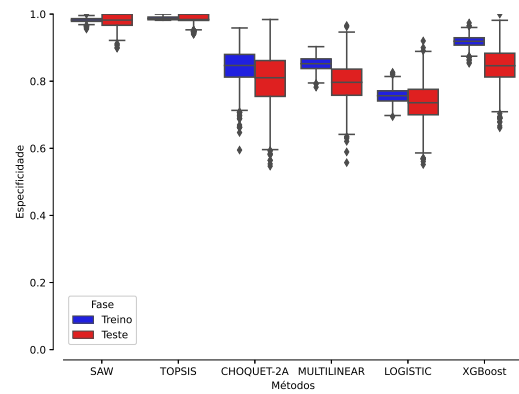


((b)) Gráfico para a Precisão, Sensibilidade e controle pelo Limiar de decisão

Figura 36 – Boxplot para F1-Score e para a representação do controle pelo limiar de decisão



((a)) Sensibilidade

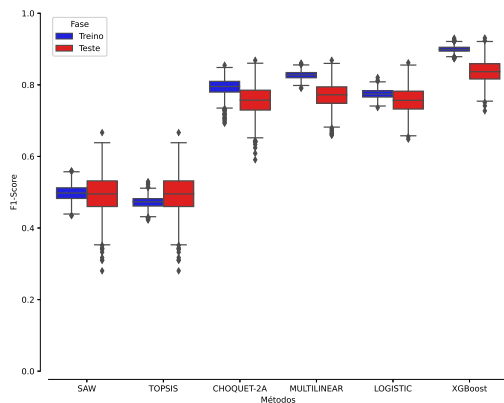


((b)) Especificidade

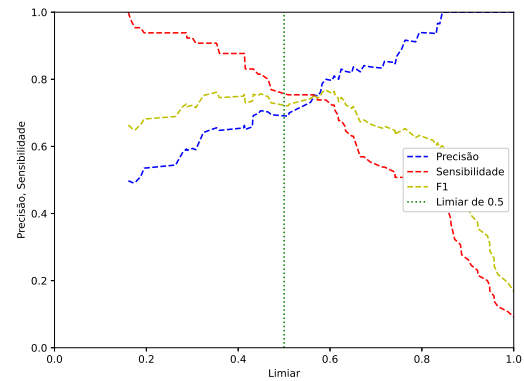
Figura 37 – Representação em boxplot das métricas para o segundo cenário

mento médio entre a precisão e a sensibilidade. Nesse cenário, o método M2CM obteve um desempenho inferior ao exibido pela integral de Choquet, mas um pouco maior do que o representado nos métodos de Regressão Logística e XGBoost.

A métrica de *F1-Score* permite verificarmos o quanto a precisão e a sensibilidade podem influenciar no resultado conforme a variação do valor de limiar. Na Figura 38(a) a disposição da mediana para a métrica de *F1-Score*. Nesse caso, a integral de Choquet, o M2CM, Regressão Logística e XGBoost mostraram-se com desempenho próximo de 0.8, no entanto, os métodos SAW e TOPSIS ficaram abaixo de 0.60. Na Figura 38(b), observamos que a precisão e sensibilidade variam conforme o limiar de decisão. Esse gráfico foi gerado com base nas medidas extraídas para o modelo M2CM. Diferente do primeiro cenário, o ponto de intersecção entre as medidas estão acima do valor de limiar de 0.6.



((a)) Boxplot para a métrica de F1-Score



((b)) Gráfico para a Precisão, Sensibilidade e controle pelo Limiar de decisão

Figura 38 – Boxplot para F1-Score e para a representação do controle pelo limiar de decisão no segundo cenário

3.4.2 Conclusões para Modelo Multilinear por Coalizões Multiplicativas (M2CM)

Os resultados foram comparados através de métricas padrões como acurácia, sensibilidade, especificidade e curva ROC. Assim, os resultados obtidos com este estudo mostram que as coalizões de sintomas aumentaram a precisão de todas as métricas de desempenho. Nos cenários analisados; os resultados obtidos em (a) e (b) foram equivalentes, o modelo multilinear proposto supera os modelos *SAW* e *TOPSIS* e tem desempenho semelhante à integral de Choquet e é próximo aos modelos de aprendizado de máquinas, com a vantagem de ser facilmente interpretável.

O modelo multilinear apresenta resultados de desempenhos em intervalos próximos aos modelos estáveis consolidados de aprendizado de máquina e de tomadas de decisão multicritério. Uma contribuição interessante para a nossa proposta no M2CM ocorre do fato da visualização gráfica da interação entre os sintomas.

Logo, podemos verificar, conforme a Figura 39, em que é uma matriz triangular superior, a diagonal representa os pesos individuais, enquanto os outros elementos da matriz dos pesos das coalizões. Notamos que o sintoma *Tosse* apresenta maior peso seguido por hiposia/cheiro, ou seja, para paciente com tais sintomas são indicativos para o diagnóstico positivo para a doença. Na condição de interação entre sintomas, temos que a combinação de *calafrios* e *dific_deglutir* possuem um maior peso.

O método M2CM obteve precisão superior a 0.85 com um limiar de decisão abaixo de 0.4. Esse valor pode ser configurado para métricas de precisão com objetivos diferentes,

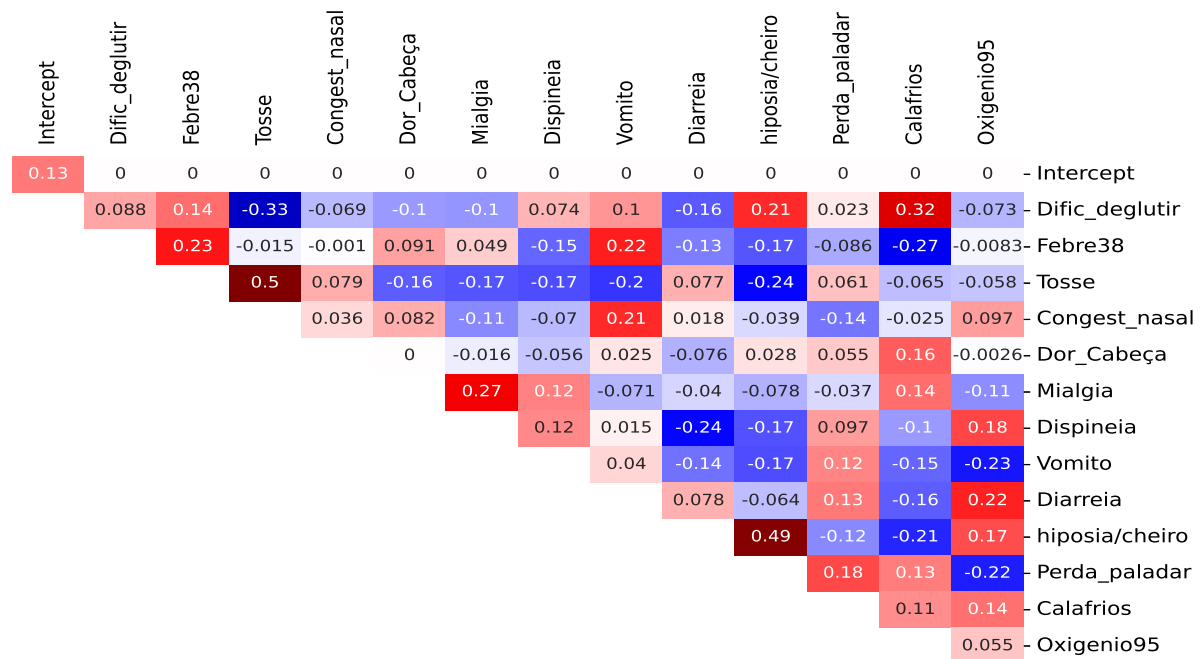


Figura 39 – Disposição dos pesos calculados pelo modelo multilinear

3.4.3 Conclusão

As avaliações das métricas de saída do Módulo 1 na abordagem orientada a dados com os algoritmos de aprendizado de máquina e no Módulo 2 com a abordagem multicritérios, trouxeram resultados que podem auxiliar na tomada de decisão do especialista. Em ambos os módulos do C3PO obtemos uma medida em termos de taxas como AUC, acurácias, sensibilidade e especificidade, dessa maneira, dados novos ou pacientes novos em fase de gestação podem ser inseridos com objetivos de realizar predições de COVID-19 bem como oferecer suporte administrativo hospitalar em questões referentes a gestão administrativa hospitalares.

Na abordagem do método proposto M2CM, observamos o poder da interpretabilidade oferecida do processo da interação multiplicativa em pares entre os critérios.

No estudo de caso da base da REBRACO, a variável *Tosse*, *hiposia/cheiro*, possuem valores de 0.5 e 0.486 respectivamente. As interações entre os sintomas *Calafrios* com *dific_deglutir* e *Vômito* com *Dific_deglutir* possuem pesos de 0.316 e 0.221. Esses pesos individuais ou das interações par-a-par representam um grau de importância alto para o caso positivo no diagnóstico. Algumas interações possuem valores próximos a zero, nesse caso seria interessante aplicar uma regularização com o objetivo de zerar esses termos.

4 Aplicação do C3PO: Predição de Desfechos para Mulheres Gestantes

Neste capítulo apresentaremos a aplicação do método C3PO na base de dados da Rede Brasileira de Morbidade Materna Grave (RNMMG), para fins de predição de alocação em leitos em UTIs por gestantes. A primeira seção descreverá os resultados referentes a abordagem orientada a dados, e na seção seguinte a abordagem com a aplicação de métodos de tomadas de decisão por preferência Aditivo Linear e o TOPSIS.

Durante os anos de 2009 e 2010, o Departamento de Tocoginecologia da Faculdade de Ciências Médicas da UNICAMP desenvolveu um estudo sobre morbidade materna grave. O estudo, denominado REDE Brasileira de Vigilância de Morbidade Materna Grave (RNMMG) contou com uma avaliação de 82.388 partos, conduzidos em 27 maternidades de referência nas cinco macrorregiões brasileiras [21, 29]. Nesse período, 9.555 mulheres tiveram complicações graves na gravidez, com 770 casos de *near miss*, ou seja, quase morte materna, e 140 mortes maternas [21, 120].

Esse banco de morbidade materna grave será utilizado para a construção dos modelos de aprendizado de máquina para a predição do uso de UTIs e desfecho materno grave (*near miss* e morte materna)[29].

As variáveis sociais podem ser descritas como: idade, cor da pele, escolaridade, estado civil, IMC, acompanhamento pré-natal, condição de internação, tempo de gestação, parto cesárea ou não, quantidade de filhos e quantidade de pré-natais realizados. Essas variáveis sociais foram inseridas no processo de treinamento, pois são atributos importantes como variáveis de entrada dos métodos de aprendizado de máquina pelo motivo das suas informações serem coletadas bem antes do desfecho.

Originalmente, a base possui variáveis preditoras com caráter socioeconômico e clínicos. Dessa maneira, de forma análoga à feita na subseção anterior 3.1, será realizado o mesmo processo da aplicação dos módulos do esquema constante no C3PO. Neste caso, para a predição de leitos em UTIs.

Na base de gestantes com morbidade materna grave, depois de realizado a seleção de atributos e pré-processamento, obtemos, de modo mínimo, 20 variáveis preditoras, que são registros antecedentes ao evento do desfecho alvo, em que temos as seguintes caracterizações: idade da mãe, raça cor da mãe, escolaridade, estado civil, índice de massa corpórea-IMC, tipo de gestação, quantidade de partos que a mulher já teve, quantidade de abortos passados, quantidade de cesarianas ocorridas anteriormente, número de filhos vivos, quantidade de pré-natais, idade gestacional, condições de morbidade da gestante,

hipertensão grave, cardiopatia, sulfato de magnésio, histerectomia/laparotomia, cianose, condições laboratoriais, condições de demora para a procura ao atendimento hospitalar, e a variável objetivo é *UtilizouUTI*. Essas variáveis foram selecionadas conforme extensas conversas mantidas em reuniões com especialistas na área de obstetrícia do hospital universitário CAISM da UNICAMP.

O atributo *UtilizouUTI* é uma variável binária que indica se a paciente esteve ou não internada em leito de UTI na fase de gestação, ou puerpera por complicações decorrentes a fatores relacionados às morbidades maternas, ou complicações de saúde prévias durante a gestação e necessitando utilizar um leito de UTI [121]. Uma previsão acurada desta variável pode ser utilizada para antecipar uma internação e preparar um leito de UTI de forma preventiva, auxiliando na gestão hospitalar e melhorando os cuidados com a saúde da mulher. [122].

4.1 Apresentação dos Resultados

Iremos apresentar os resultados empíricos da aplicação do método C3PO na base RNMMG. Então, para cada módulo do C3PO, teremos as saídas em forma de probabilidade de ocorrência do evento, conforme serão apresentados nas próximas seções.

Inicialmente, os resultados serão dados pelo Módulo 1 do C3PO, da abordagem orientada a dados, por métricas padrões. Posteriormente, teremos os resultados para o Módulo 2 do C3PO da abordagem orientada a protocolos e, por fim, resultados para o método M2CM, seguindo os mesmos passos apresentados no capítulo 4.

4.1.1 Resultado para o Módulo 1

As variáveis de entradas foram separadas em um conjunto de dados para treinamento e teste. Apresentamos uma parte da base de treino em com suas saídas em probabilidades, segundo especificações de cada algoritmo.

As tabelas mostram exemplos da separação da base, sendo 70% e 30%, treino e teste. Na Tabela 26 temos uma parte da base de treino, e a base de teste em termos de probabilidade conforme o exemplo na Tabela 27. Uma relação que podemos observar é o grau de correlação entre variáveis de testes e a variável dependente, tomando como base a correlação de Pearson conforme mostrado na Figura 40.

Notamos que as saídas dos modelos KNN e ADA possuem um grau de correlação menor com a saída.

Tabela 26 – Probabilidades de treinamento

logist	gbm	lgbm	xgbm	rf	mlp	ada	svm	nb	knn	Covid-19
0.088	0.092	0.087	0.157	0.131	0.104	0.493	0.101	0.000	0.20	0
0.145	0.059	0.049	0.144	0.067	0.036	0.493	0.160	0.000	0.00	0
0.968	0.998	0.991	0.964	0.990	1.000	0.509	0.948	0.000	0.800	1
...
0.850	0.920	0.892	0.965	0.884	0.999	0.655	0.985	0.000	0.815	1

Tabela 27 – Probabilidades de teste

logist	gbm	lgbm	xgbm	rf	mlp	ada	svm	nb	knn	Covid-19
0.973	0.990	0.993	0.943	0.934	0.966	0.595	0.966	0.00	0.4	1
0.917	0.985	0.981	0.900	0.829	0.965	0.505	0.925	0.00	0.4	1
0.534	0.133	0.189	0.251	0.261	0.397	0.499	0.433	0.00	0.4	1
...
0.255	0.152	0.192	0.565	0.125	0.123	0.323	0.125	0.00	0.120	0

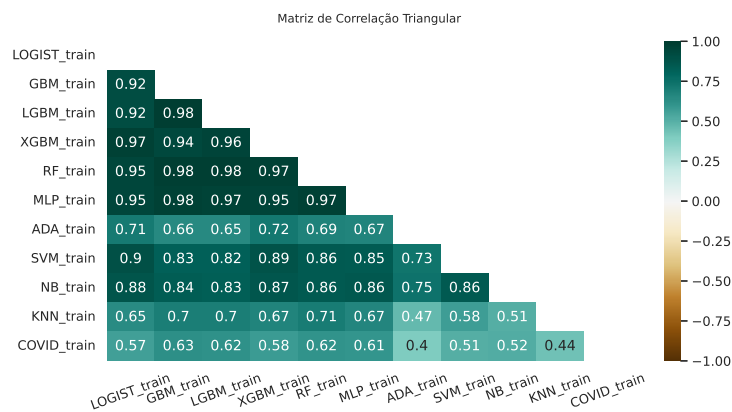


Figura 40 – Matriz de correlação das saídas de teste

4.1.2 Resultados Individuais para cada saída do C3PO

Estabelecendo um limiar de corte de 0.5, temos os seguintes resultados para a predição de leitos em UTI utilizando o Módulo 1 do *framework* do C3PO. Vale lembrar que os resultados são provenientes das saídas dos dez algoritmos exemplificados em forma de probabilidades em que cada uma são retreinadas com uma agregação de regressão logística, e o resultados estão apresentados nas Tabelas 28, 29 e 30.

Nos resultados apresentados na Tabela 28, os algoritmos LGBM e XGBM alcançaram uma taxa 0.88 de AUC. No entanto, percebemos que em relação as outras saídas, não houve uma variação significativa a menos das apresentadas por Naive Bayes e KNN.

Para os resultados de taxa de verdadeiro positivo e negativos, a partir de um limiar de 0.5 podem ser ajustados conforme especificações da probabilidade a priori de

Tabela 28 – Resultados para as saídas individuais da REDE com regressão logística-AUC

iterações	log	gbm	lgbm	xgbm	rf	mlp	ada	svm	nb	knn
1 ^a	0.87	0.87	0.88	0.88	0.86	0.88	0.87	0.87	0.68	0.77
2 ^a	0.86	0.87	0.88	0.87	0.86	0.87	0.86	0.86	0.66	0.75
3 ^a	0.86	0.85	0.86	0.86	0.83	0.86	0.86	0.86	0.68	0.75
4 ^a	0.87	0.88	0.89	0.89	0.87	0.88	0.87	0.87	0.68	0.76
5 ^a	0.87	0.87	0.88	0.88	0.85	0.88	0.87	0.87	0.66	0.76
Média	0.87	0.87	0.88	0.88	0.85	0.87	0.87	0.87	0.67	0.76

Tabela 29 – Resultados para as saídas individuais da REDE com regressão logística-TPR

iterações	log	gbm	lgbm	xgbm	rf	mlp	ada	svm	nb	knn
1 ^a	0.74	0.77	0.76	0.74	0.76	0.75	0.23	0.74	0.36	0.75
2 ^a	0.74	0.76	0.76	0.72	0.77	0.72	0.21	0.74	0.34	0.68
3 ^a	0.73	0.77	0.76	0.73	0.74	0.73	0.21	0.73	0.34	0.7
4 ^a	0.72	0.77	0.77	0.74	0.78	0.72	0.17	0.76	0.32	0.71
5 ^a	0.75	0.77	0.77	0.74	0.77	0.76	0.23	0.74	0.34	0.73
Média	0.74	0.77	0.76	0.73	0.76	0.74	0.21	0.74	0.34	0.71

Tabela 30 – Resultados para as saídas individuais da REDE com regressão logística-TNR

iterações	log	gbm	lgbm	xgbm	rf	mlp	ada	svm	nb	knn
1 ^a	0.83	0.79	0.81	0.86	0.76	0.84	0.99	0.84	0.98	0.65
2 ^a	0.82	0.83	0.81	0.86	0.78	0.82	1.00	0.83	0.97	0.68
3 ^a	0.82	0.77	0.78	0.84	0.75	0.83	1.00	0.83	0.98	0.68
4 ^a	0.84	0.82	0.82	0.86	0.78	0.87	0.99	0.81	0.99	0.68
5 ^a	0.84	0.80	0.82	0.85	0.75	0.84	0.99	0.86	0.97	0.65
Média	0.83	0.80	0.80	0.85	0.76	0.84	0.99	0.83	0.97	0.66

consenso do médico especialista.

4.1.3 Agregando todas as saídas

Dessa maneira, se realizarmos a combinação de todas as saídas de probabilidades e determinar uma agregação por uma logística, observamos que as taxas de TVP e TVN ficam em 0.77 e 0.78 respectivamente e 0.86 de AUC e demanda um custo computacional maior em relação aos outros tipos de combinações par-a-par ou individual.

4.1.4 Agregando as saídas em pares

Os resultados para combinação par-a-par com agregação das saídas estão apresentados nas Tabelas 31, 32 e 33.

Note que para os resultados de combinação dos classificadores par-a-par, obtemos valores que alcançam taxas de 0.9 como no caso de um XGBoost com Rede Neural.

Tabela 31 – Resultados para a AUC com combinação par-a-par por agregação logística.

AUC	logist	gbm	lgbm	xgbm	rf	mlp	ada	svm	nb	knn
logist	-	0.86	0.88	0.89	0.86	0.89	0.87	0.87	0.87	0.85
gbm	-	-	0.88	0.86	0.87	0.87	0.88	0.86	0.88	0.87
lgbm	-	-	-	0.89	0.87	0.89	0.89	0.88	0.89	0.88
xgbm	-	-	-	-	0.86	0.90	0.89	0.89	0.89	0.87
rf	-	-	-	-	-	0.86	0.87	0.86	0.87	0.87
mlp	-	-	-	-	-	-	0.89	0.89	0.89	0.87
ada	-	-	-	-	-	-	-	0.88	0.87	0.81
svm	-	-	-	-	-	-	-	-	0.87	0.85
nb	-	-	-	-	-	-	-	-	-	0.80
knn	-	-	-	-	-	-	-	-	-	-

Em contrapartida, a combinação entre Naive Bayes com KNN teve uma taxa de 0.8 de AUC.

Olhando para a tabela de correlação descrita na Figura 40, a correlação entre *XGBoost* com Rede Neural MLP possuem o coeficiente de correlação de Pearson de 0.93, sendo que ambos já têm, em relação à variável objetivo um coeficiente alto. Situação análoga podemos afirmar para o coeficiente de correlação na combinação entre *Naive Bayes* com *KNN*.

Tabela 32 – Resultados para a TVP com combinação par-a-par

TVP	logist	gbm	lgbm	xgbm	rf	mlp	ada	svm	nb	knn
logist	-	0.76	0.78	0.75	0.77	0.79	0.74	0.75	0.73	0.72
gbm	-	-	0.78	0.76	0.78	0.77	0.77	0.76	0.77	0.77
lgbm	-	-	-	0.77	0.78	0.78	0.78	0.78	0.78	0.77
xgbm	-	-	-	-	0.76	0.77	0.75	0.75	0.75	0.75
rf	-	-	-	-	-	0.78	0.78	0.77	0.78	0.78
mlp	-	-	-	-	-	-	0.78	0.78	0.78	0.77
ada	-	-	-	-	-	-	-	0.76	0.33	0.71
svm	-	-	-	-	-	-	-	-	0.76	0.71
nb	-	-	-	-	-	-	-	-	-	0.62
knn	-	-	-	-	-	-	-	-	-	-

A Tabela 32 que mostra os resultados de taxas de verdadeiro positivo e suas combinações de classificadores. As taxas oscilam em torno de 0.78. A menor taxa fica por conta da combinação entre os algoritmos de *KNN* com *Naive-Bayes*.

Por fim, verificando a Tabela 33 com os resultados de taxas de verdadeiro negativo, ou a especificidade, mostra resultados que alcançam 0.99, no entanto, os pares de classificadores *Naive-Bayes* e *ADA* possuem as características de não lidar bem com bases desbalanceadas e baixa correlação com a variável dependente, desse ponto, iremos olhar

Tabela 33 – Resultados para a TVN com combinação par-a-par com agregação por regressão logística

TVN	logist	gbm	lgbm	xgbm	rf	mlp	ada	svm	nb	knn
logist	-	0.81	0.82	0.85	0.77	0.83	0.83	0.82	0.84	0.81
gbm	-	-	0.81	0.81	0.78	0.81	0.81	0.81	0.81	0.81
lgbm	-	-	-	0.81	0.77	0.81	0.81	0.82	0.81	0.82
xgbm	-	-	-	-	0.77	0.83	0.85	0.85	0.86	0.83
rf	-	-	-	-	-	0.77	0.77	0.77	0.77	0.77
mlp	-	-	-	-	-	-	0.84	0.83	0.84	0.82
ada	-	-	-	-	-	-	-	0.81	0.99	0.68
svm	-	-	-	-	-	-	-	-	0.81	0.81
nb	-	-	-	-	-	-	-	-	-	0.87
knn	-	-	-	-	-	-	-	-	-	-

para as combinações de valores correspondentes aos algoritmos que classificam dentro de faixas mais plausíveis no intervalo de 0.82 a 0.88.

4.1.5 Controle somente pelo limiar

Para o controle apenas pelo limiar, ou seja, escolhendo um valor de 0.5 para ponto de corte nas probabilidades de saídas dos algoritmos de aprendizado de máquina, obtemos os seguintes resultados mostrados na Tabela 34. Verificamos que, para encontrar uma saída que melhor represente a predição, olhamos para as taxas de AUC que oferecem o quanto a taxa de verdadeiro positivo se desloca numa variação do limiar. Logo, o desempenho das métricas avaliadas na predição ficam em torno de 0.76 e 0.89 em AUC.

Tabela 34 – Resultados para as métricas controlando apenas pelo limiar de 0.5

	logist	gbm	lgbm	xgbm	rf	mlp	ada	svm	nb	knn
AUC	0.87	0.88	0.89	0.89	0.87	0.89	0.87	0.87	0.82	0.76
TVP	0.72	0.75	0.76	0.71	0.72	0.72	0.73	0.76	0.31	0.52
TVN	0.85	0.85	0.85	0.88	0.85	0.87	0.86	0.81	0.99	0.88

4.1.6 Comitê de Máquina

Os resultados para uma combinação de um comitê de classificadores, escolhendo às três melhores taxas de AUC, como o GBM, LBM e MLP, obtemos o resultado de 0.90 de AUC, TVP de 0.77 e TVN de 0.85, conforme podemos verificar na Figura 41.

Portanto, temos que a sensibilidade mostrada na curva ROC demonstra uma performance nas suas métricas de avaliação adequada em torno de 0.80, sugerindo um ponto de corte em 0.5.

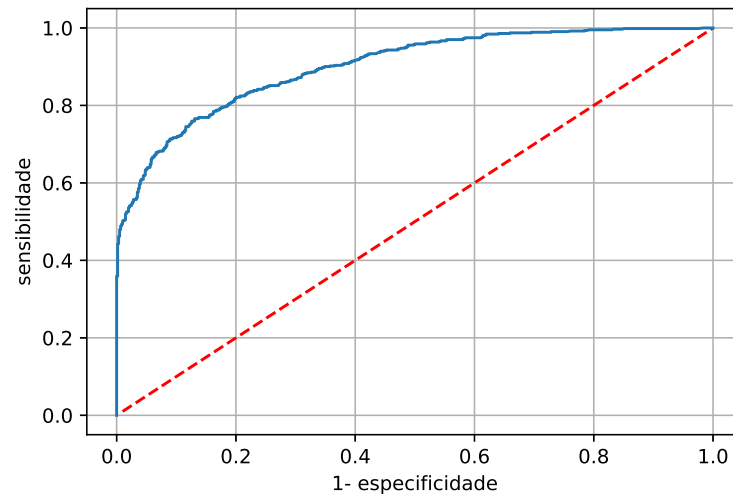


Figura 41 – Curva ROC para o comitê de máquina para predição de leitos em UTIs com AUC de 0.90

4.2 Resultado para o Módulo 2

Nesta seção iremos estudar os resultados das saídas pertinentes ao campo relacionado com as tomadas de decisões multicritérios orientado a protocolos com métodos *Multi-Attribute Value Theory* para o *framework* do C3PO para tomada de decisão em relação à alocação de leitos em UTIs para gestantes. A matriz de decisão será a extraída do estudo na seção 4.1 do Módulo 1 de predição orientada a dados. Dessa forma, aplica-se o processo através dos métodos de tomadas de decisão multicritérios como serão demonstrados nas seções seguintes.

Para os testes computacionais são aplicados o método de preferências *SAW*, *TOPSIS*, na base de dados da REDE Brasileira de Morbidade Materna Grave em que teremos as saídas em *scores*, e para obtermos valores em probabilidades, aplicamos a regressão logística.

4.2.1 Definição dos pesos dos critérios da base da REDE

Para um médico especialista na área de obstetrícia, cada critério de dados das pacientes gestantes possuem um grau de importância para os atributos que diz respeito ao prognóstico de desfecho clínico. Dessa maneira, iremos realizar um ajuste dos pesos pelo gradiente descendente linear, em que tais valores serão aplicados aos métodos do Módulo 2 do C3PO. Dessa forma, estabeleceremos um vetor com pesos ajustados, e um segundo vetor composto apenas do elemento 1 como pesos, ou seja, essa coluna não considera ponderação de pesos aos atributos para a matriz de decisão.

4.2.2 Método Aditivo Linear

Nesta aplicação do método aditivo linear, vamos analisar as saídas de probabilidades geradas pelos *scores* do primeiro grupo pesos e do segundo grupo pesos ajustados por decida do gradiente.

Tabela 35 – Probabilidades para o método aditivo linear

amostras	$probs_1$	$probs_2$	$probs$	UTI
0	0.69	0.48	0.88	1
1	0.63	0.45	0.90	1
...
882	0.30	0.48	0.20	1
883	0.40	0.48	0.37	0

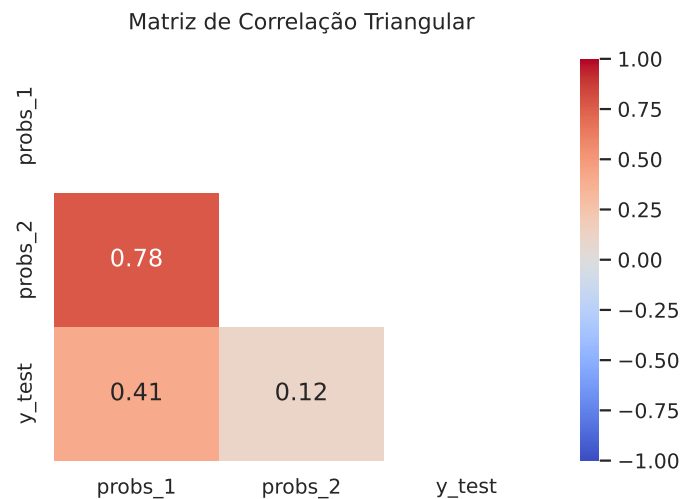


Figura 42 – Matriz de correlação das saídas de teste

A correlação entre as probabilidades juntamente com a variável dependente mostrarão o comportamento preditivo do modelo. Desta maneira, tendo como ponto de corte na faixa de probabilidades de 0.5, temos como resultado demonstrado na Tabela 36.

Tabela 36 – Resultado para o método aditivo linear

	TVP	TVN	AUC	ACC
$probs_1$	0.59	0.81	0.73	0.7
$probs_2$	0.40	0.79	0.55	0.60
$probs$	0.65	0.82	0.79	0.74

Observamos que a linha com a alternativa **probs**, composta pelas duas probabilidades a $probs_1$ e $probs_2$, apresenta melhor taxa em relação aos demais da tabela.

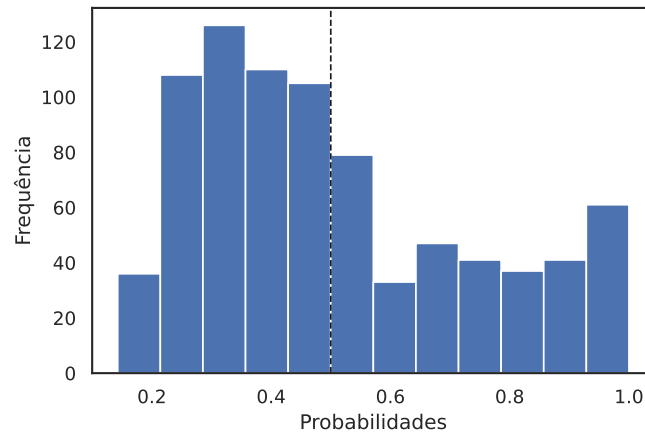


Figura 43 – Distribuição de probabilidades para o método aditivo linear

4.2.3 Método TOPSIS

A aplicação do método TOPSIS que compõe o Módulo 2 da estrutura do C3PO e usando as mesmas variáveis preditivas da seção anterior. Ao examinar a matriz de correção entre as probabilidades extraídas dos *scores* dos métodos TOPSIS. Notamos que a probabilidade *probs*, possui um coeficiente de correlação alto com a variável *probs*₁.

Em relação à variável dependente y_{test} , a variável *probs* possui a maior correlação em comparação com as outras probabilidades. Agora olhando para a saída *prob*₂ possui a menor correlação com as outras variáveis, com valor de 0.005, isso pode tornar a variável sem relevância preditiva.

A Figura 44, as probabilidades *probs*₁ e *probs* mantêm uma forte correlação, desse modo, podemos substituir algumas dessas variáveis na análise das métricas. No entanto, para fins de comparação, iremos deixá-los nos cálculos seguintes. Vamos mostrar os resultados através da Tabela 37.

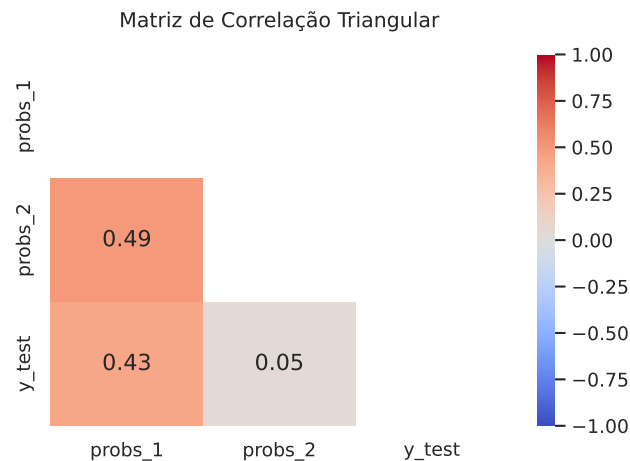


Figura 44 – Matriz de correlação entre as probabilidades de saída

Tabela 37 – Resultados para o método TOPSIS

	TVP	TVN	AUC	ACC
$probs_1$	0.58	0.85	0.75	0.71
$probs_2$	0.51	0.38	0.52	0.45
$probs$	0.59	0.87	0.77	0.73

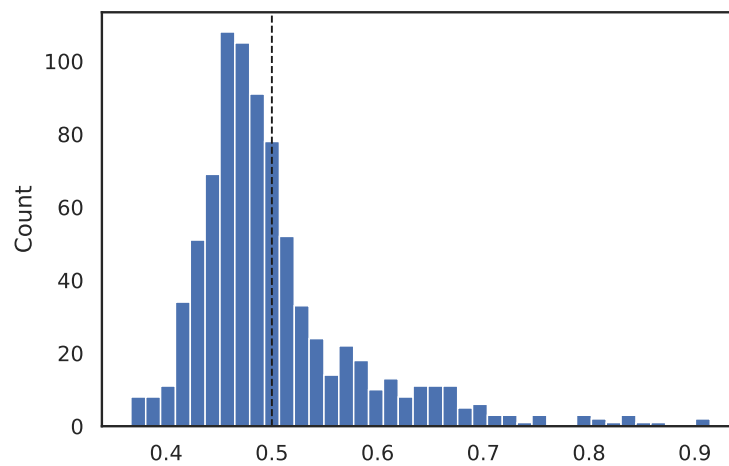


Figura 45 – Distribuição de probabilidades da REDE

Pela Tabela 37, a alternativa **probs** obteve melhor resposta. Olhando para o gráfico da Figura 45, percebemos que no ponto de corte em 0.5 é o ponto médio do intervalo $[0, 1]$, no entanto, a maioria das probabilidades são valores abaixo de 0.5, dessa forma, o ponto de corte sugerido seria ajustado a um valor menor do que a forma usual.

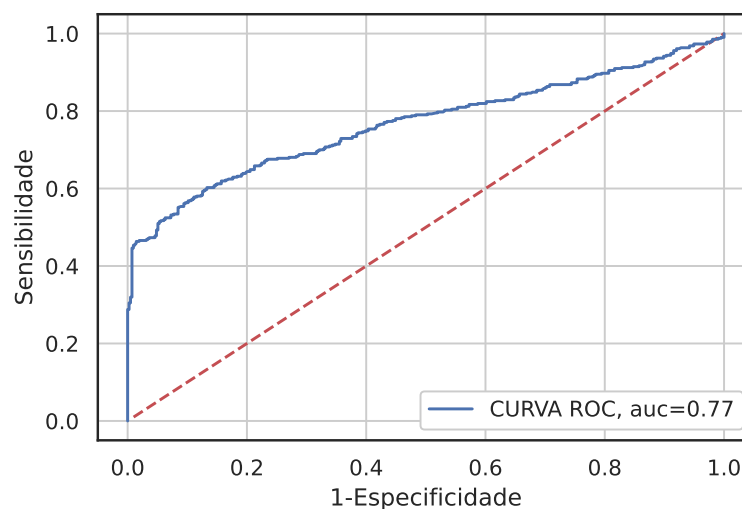


Figura 46 – Curva ROC para o método TOPSIS aplicado na base da REDE

Um outro recurso visual interessante seria observar o comportamento da variação da sensibilidade no gráfico da curva ROC Figura 46, pois, a variação na sensibilidade, implica uma variação das métricas de especificidade consoante o parâmetro do limiar setado.

4.3 C3PO: Aplicação do Modelo Multilinear por Coalizão Multiplicativa

Nesta seção iremos aplicar o modelo multilinear M2CM desenvolvido na seção 2.2 no estudo da REDE de morbilidade materna grave para a predição de uso de leitos em UTIs por gestantes e realizar as comparações com os outros modelos já aplicados nesta tese.

Para verificar o desempenho do modelo multilinear M2CM, vamos comparar a média dos 100 resultados de treinamento e testes nos modelos de aprendizado de máquina da abordagem orientado a dados: regressão logística, XGBoost, e de tomada de decisões multicritérios da abordagem orientado a protocolo: Aditivo Linear e TOPSIS e vamos extrair as métricas de acurácia, especificidade, sensibilidade e área sob a curva- AUC.

Nos resultados, vamos analisar em dois cenários que serão nomeados, o primeiro com a base de dados da REDE desbalanceada em suas classes de saídas, e o segundo, com a base balanceada. A técnica de tornar a base balanceada é baseada na mesma abordagem de balanceamento em que foi apresentada na subseção 3.4, conforme técnica de edição de amostras desenvolvida por [63].

4.3.1 Resultados

Para os primeiros resultados, vamos considerar os contextos de cenários, (a) sem balanceamento e (b) após a base de dados passar pelo processo de balanceamento. Vamos utilizar os métodos SAW, TOPSIS, M2CM, Regressão Logística e XGBoost.

Na Figura 47(a), de forma comparativa, podemos observar a sensibilidade de treino e teste e os seus respectivos métodos. Para o modelo multilinear M2CM, com a base de dados da REDE sem a estratégia de balanceamento, os resultados estão dentro da alcançada em faixa comparáveis aos modelos de aprendizado de máquina clássicos, no nosso caso, a Regressão Logística e XGboost. Observamos que a mediana ficou abaixo de 0.5, isso se deve ao fato da base está desbalanceada. No entanto, a mediana de sensibilidade se contrapõe a verificada na especificidade.

Na Figura 47(b), mostramos os resultados em relação à especificidade dos modelos. Percebemos que os valores ficaram com médias acima de 0.9, ou seja, o modelo M2CM alcançou desempenho comparável aos modelos clássicos. Esses resultados com uma

amplitudes de diferenças altas entre as métricas de sensibilidade e especificidade acontecem devido à base se encontrar desbalanceada.

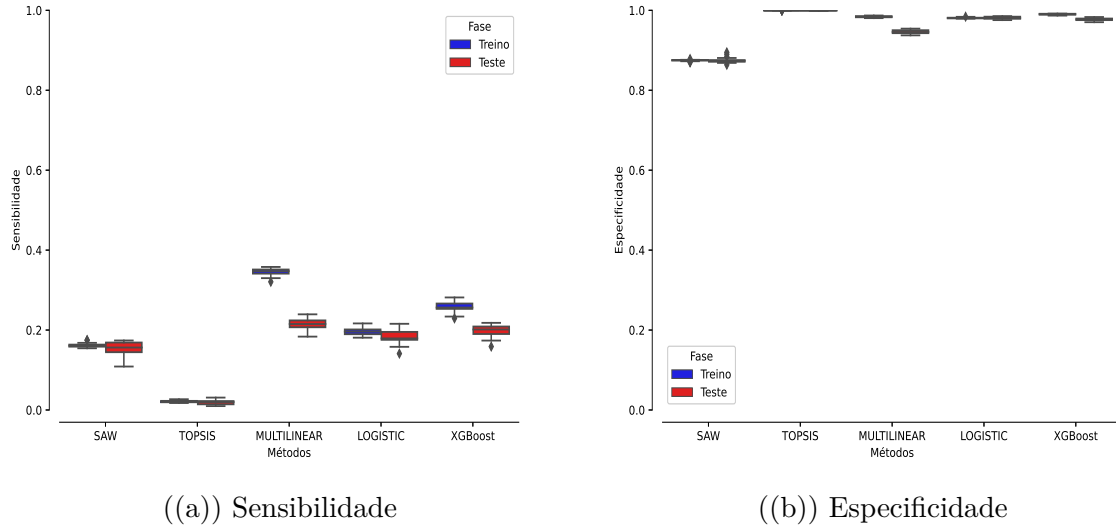


Figura 47 – Representação gráfica para o primeiro cenário

Os resultados para o $F1$ -Score dos modelos, temos conforme a Figura 48(a). E o controle pelo limiar pode ser evidenciado conforme a Figura 48(b). Em ambos os gráficos, percebemos que a medida para o $F1$ -Score não alcançou 0.5 de desempenho para treino e teste. O método M2CM mostra que acontece um sobreajuste no treino e teste. Variando o limiar para valores abaixo de 0.35, podemos atingir valores de sensibilidades acima de 0.7. A precisão da taxa de acertos dos casos positivos acontece quando o limiar de decisão aumenta. Na Figura 48(b), fica claro que o controle das métricas de desempenho podem ser ajustadas conforme objetivos de diagnóstico e triagem médica.

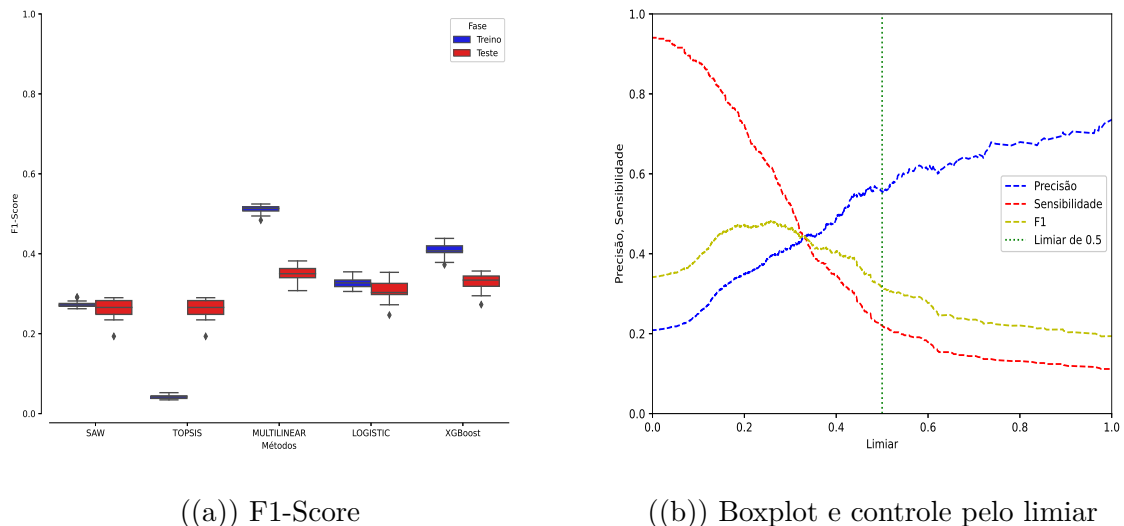


Figura 48 – Representação gráfica para o primeiro cenário

Vamos analisar o segundo cenário (b), em que iremos retirar e editar algumas amostras que possuem dados ruidosos através de um método que realiza balanceamento por *undersampling*.

O primeiro resultado que apresentamos para a sensibilidade conforme a Figura 49(a). As sensibilidades dos métodos estão com medianas baixas, e o método TOPSIS não obteve treinamento e testes significantes. Na Figura 49(b), mostramos os resultados em relação à especificidade dos modelos, a mesma análise feita para a acurácia pode ser atribuída para a Figura 49(b). Notamos, que os métodos não obtiveram boa generalização para o aprendizado no conjunto de dados da REDE, pois a sensibilidade e especificidade possuem amplitudes altas nas métricas de desempenho.

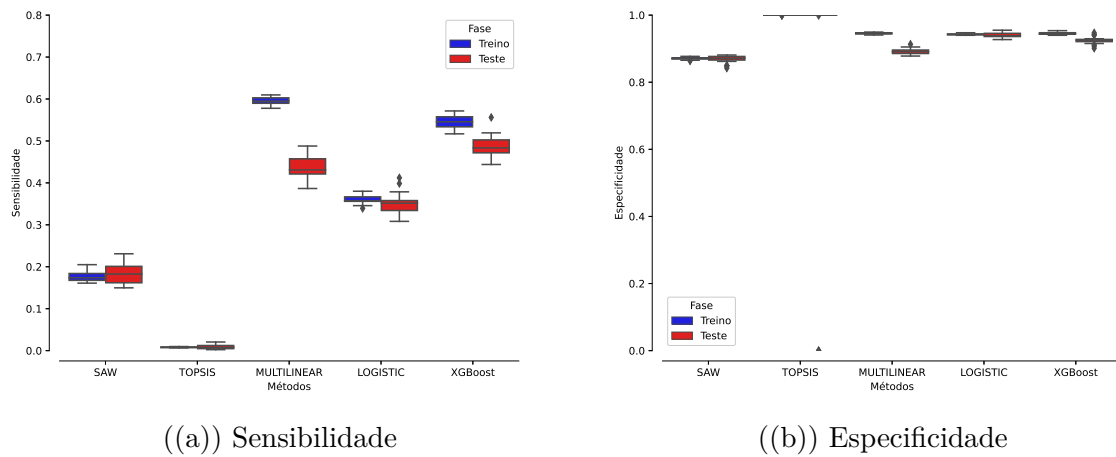


Figura 49 – Representação em boxplot das métricas para o segundo cenário

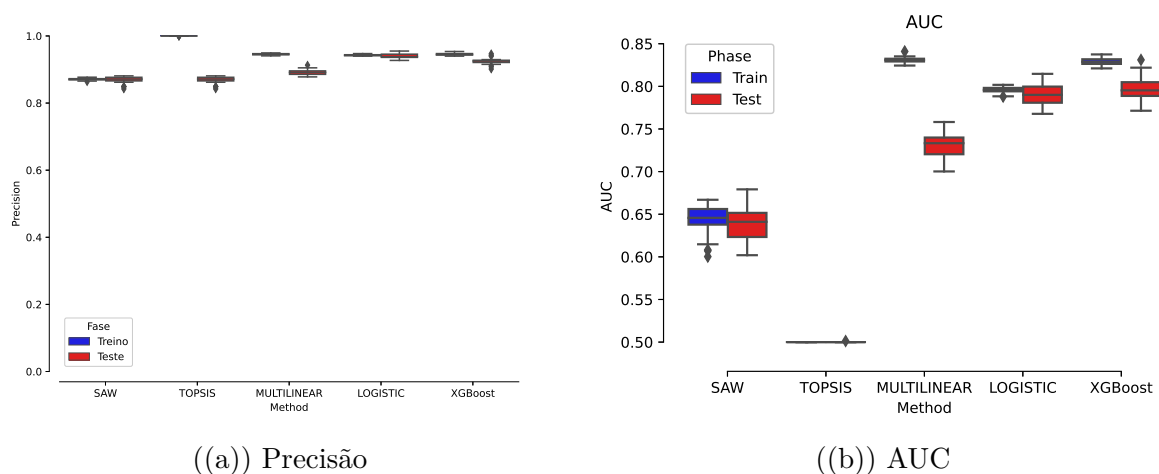


Figura 50 – Representação em boxplot das métricas para o segundo cenário

Para os resultados da métrica AUC, apresentamos na Figura 50(b). Esse resultado mostra a mediana da área sob a curva das fases da qualidade do ajuste do limiar para os diversos valores de sensibilidade pelo $(1 - \text{especificidade})$. O método TOPSIS, nessa

métrica, não conseguiu generalizar entre treino e teste. Na Figura 50(a), mostramos o resultado para a precisão dos modelos, ou seja, a capacidade de acertos dos casos positivos estão acima de 0.8.

Observamos que a AUC, na Figura 50(b), o método M2CM, obteve mediana para treinamento alta, no entanto, em seu teste foi pouco abaixo em relação ao método de Regressão Logística e XGBoost. O método M2CM tem valor de AUC acima de 0.8 no treinamento, no entanto, para o teste, acontece uma queda nas taxas. O método Aditivos Linear (SAW) medianas apresentou valor próximo a 0.65. O método TOPSIS obteve generalização para os dados com significância nula. O método XGboost possui uma particularidade da sua estrutura ser mais complexa para implementação e isso mostra uma das vantagens em que os métodos interpretáveis podem trazer pelo fato de conseguirmos identificar os seus parâmetros onde ocorre interações positivas ou negativas e evitar que tal situação ou conferir explicabilidade ao modelo.

Para os resultados mostrados através dos gráficos de *boxplots*, no primeiro cenário (a), percebemos que os métodos performam com medianas próximas de especificidades altas. No entanto, para sensibilidade, não encontramos resultados dentro da estimativa esperada. Para o cenário (b), com a técnica de *undersampling*, os resultados obtidos possuem desempenhos razoavelmente melhor aos encontrados no cenário (a), evidenciando que nessa fase o modelo pode generalizar melhor.

Além dessas métricas apresentadas, a precisão e o *F1-Score* podem fornecer informações sobre os diagnósticos positivo e estão corretas e o equilíbrio entre a precisão e sensibilidade. Na Figura 51(b) o método TOPSIS e SAW não obtiveram bons desempenhos.

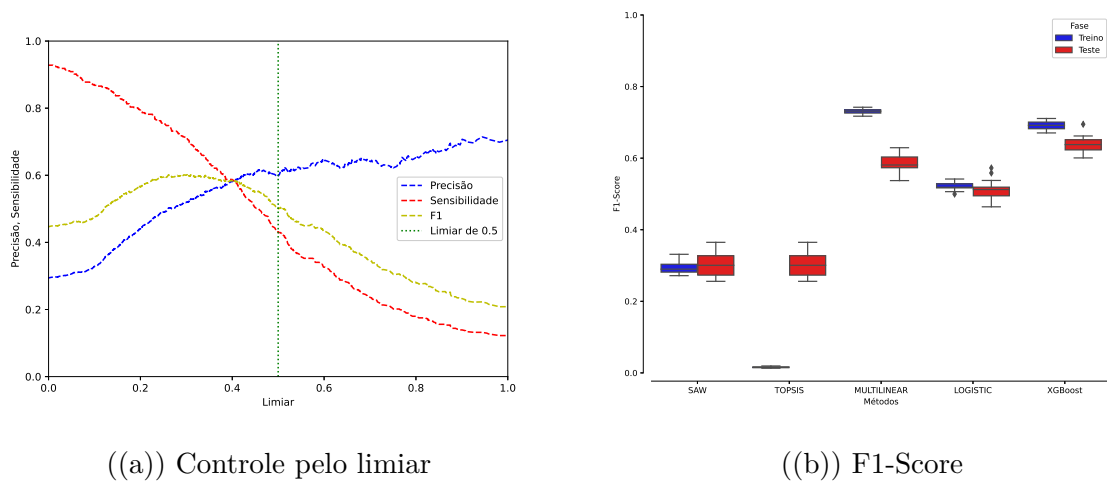


Figura 51 – Representação em boxplot das métricas para o segundo cenário

Na Figura 51(a), podemos visualizar, com melhor detalhe, o comportamento entre as três métricas ao longo do limiar de decisão. O valor do limiar padrão em 0.5

permite verificar em qual desempenho o decisor necessita para o diagnóstico ou triagem clínica. Para valores acima do ponto de corte do limiar a precisão aumenta e a sensibilidade decresce. Para um limiar de decisão próximo de 0.38 as medidas de desempenhos são iguais.

4.3.2 Conclusão

Neste capítulo constatamos a aplicação do método C3PO no estudo da Rede Brasileira de Morbidade Materna Grave. A primeira parte deste capítulo mostrou resultados referentes a aplicação dos métodos de aprendizagem de máquina no Módulo 1. Na segunda seção foi apresentada os resultados para os métodos orientados a protocolo. Ao final, apresentamos a aplicação do modelo Multilinear por Coalizões Multiplicativas (M2CM). Os experimentos computacionais realizados sugerem que o método M2CM pode atingir

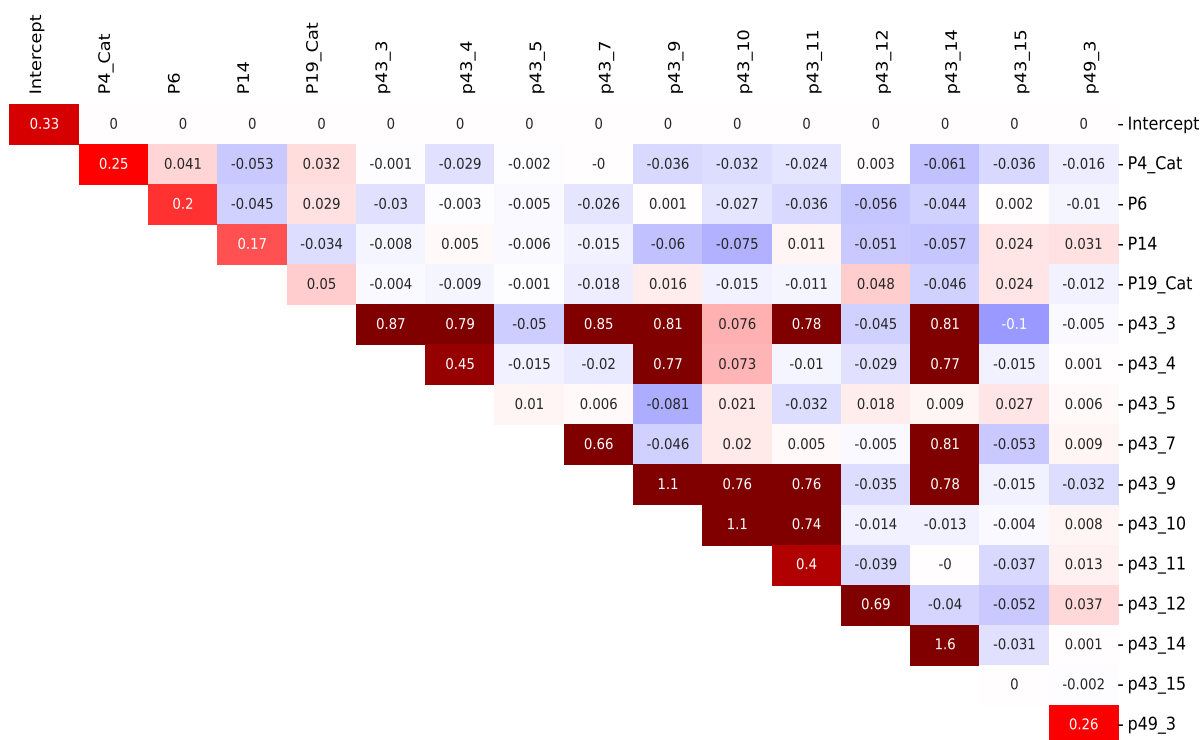


Figura 52 – Disposição dos pesos calculados pelo modelo multilinear

métricas de precisão semelhantes apresentados pelos métodos clássicos de aprendizado de máquina, dessa forma, mostramos como complemento metodológico a tais modelos pelo fato de apresentar boa interpretabilidade e, de forma prática, oferecer caminhos para verificar quais variáveis potencializam a saída em casos positivos e negativos.

Uma característica do modelo M2CM na base de dados da REDE, é o relacionado ao tempo de operação para a classificação em um número de 101 iterações para a divisão da base em treino e testes, pois se trata de um conjunto de dados com volume de informações grande.

A Figura 52, ilustra melhor as interações entre alguns sintomas selecionados da base de dados em estudo. Os sintomas/sinais que possuem maior destaque no gráfico sinalizam os indicadores que contribuem de forma significativa para os diagnósticos positivos do desfecho estimado, como no caso das variáveis $p43_9$ que indica anemia e $p43_14$, neoplasia. Na diagonal são os sintomas/sinais individuais e os demais são as interações par-a-par encontradas pelo método M2CM.

Nesse estudo de caso, para a predição da utilização de UTIs por gestantes com modelos de aprendizado de máquina e métodos multicritérios, os resultados podem oferecer um suporte com o modelo já treinado em grandes bases de dados fornecidos pelos centros clínicos-hospitalares como um sistema auxiliar em outras regiões com difícil acesso a prognósticos médicos especializados.

5 Considerações finais e perspectivas

O estudo abordado nesta tese mostra a importância do desenvolvimento de modelos matemáticos e computacionais em estruturas que possam ser interpretáveis. Trazendo transparência e confiabilidade as decisões específicas em torno de processos preditivos se tornarem menos "caixa-preta".

Apresentamos uma breve revisão da literatura sobre os métodos de aprendizado de máquina orientado a dados e de apoio multicritério orientado a protocolo, ilustrando com exemplos, sendo alguns métodos com interpretabilidade e outros com sua estrutura mais complexas. O intento foi, nos capítulos posteriores, formar uma base de conhecimentos e comparar as métricas de saídas, sendo em probabilidades e para os métodos de apoio a decisão multicritérios os *scores* e logo em seguida os transformando em probabilidades, com todos os casos controlados pelo limiar de corte.

Dentre os algoritmos de aprendizado de máquina apresentados no método C3PO, que podemos dizer interpretáveis, são a Regressão Logística, Floresta Aleatória, *Naive Bayes*. Para as saídas por métodos de orientação a protocolos multicritérios, utilizamos as abordagens em que a saída na classificação fosse comparada com os métodos de aprendizado de máquina. Portanto, para algumas saídas de método orientado a protocolos com abordagem multicritério, utilizamos a regressão logística para os *scores* fossem mapeados no intervalo entre 0 e 1 com a finalidade de se obter uma probabilidade. Assim, o esquema do C3PO, pode comportar outras categorias de métodos, tanto os orientados a dados como orientado a protocolos.

Dessa maneira, modelos interpretáveis orientado a dados e orientado a protocolos, foram inseridos no método C3PO com o propósito de oferecer melhorias nas tomadas de decisão na área médica, em especial, no que se refere a área de atendimento hospitalar e cuidados na saúde das gestantes.

Uma contribuição inerente a este estudo é o método multilinear por coalizões com a interação multiplicativa entre dois critérios em que os pesos são ajustados por programação quadrática. Dessa maneira, para o modelo multilinear desenvolvido nesta tese, verificamos que os ajustes de pesos individuais e extraídos das interações entre dois critérios, proporciona, ao tomador de decisão, ferramentas que auxiliam para indicar quais critérios ocorrem o processo com ganhos positivos e negativos ou ausência de interação. No estudo de casos, mostramos que as contribuições entre cada sintoma diferente podem levar a resultados diferentes dependendo do grau das suas interações. Outro subsídio que esta tese oferece é o suporte para a modelagem da situação operacional clínica, com aplicações, na função objetivo, de restrições nos parâmetros para o cálculo dos *scores* de pesos e,

estabelecendo, por meio da expertise médica, parâmetros que melhoram os ajustes dos resultados estimados.

Para o modelo multilinear por coalizões multiplicativas, tem como um dos objetivos, modelar o pensamento bayesiano comumente ligado aos critérios decisórios dos médicos em tomadas de decisões em diagnósticos clínicos. Dessa maneira, o C3PO propõe levar, como perspectiva, o comportamento das resoluções das expertises clínicas, como ferramenta auxiliadoras em outros contextos sociais.

Como perspectiva de continuação do presente estudo, várias práticas podem ser empregadas na elaboração dos conceitos já apresentados com o propósito de tornar mais preciso as taxas das métricas de saídas, e até realizar a abordagem por coalizões além de dois critérios.

Por cima dessa ideia, para aprimoramento dos resultados do método multilinear por coalizões multiplicativas, podemos introduzir, como perspectiva de pesquisa, a aplicação da regularização baseados na norma l_p , ou seja, para nosso propósito, as dadas pela norma l_1 ou l_2 , como uma forma de controlar os ruídos, sobreajustes provocados, de maneira a controlar os pesos e entregar as coalizões que realmente têm importância ao modelo.

Uma particularidade interessante da norma l_1 como regularização é que retorna soluções esparsas, o que significa que a solução do problema é um vetor contendo elementos nulos ou próximos de zero e mais simples em termos de complexidade do modelo. Assim, tornando o problema de otimização relacionado com abordagem conhecida como LASSO (Least Absolute Shrinkage and Selection Operator) tal qual o método de variável não-negativa [123].

Além da regularização LASSO, podemos expressar uma capacidade ou os pesos das interações de um modelo multiplicativo por coalizões em termos do índice de interação de Banzhaf como descrito em [102] e também por potência de índices de interação correspondentes ao valor de Shapley conforme descrito nos trabalhos [124, 104, 125].

Referências

- [1] R. M. Cruz, R. Sabourin, and G. D. Cavalcanti, “Dynamic classifier selection: Recent advances and perspectives,” *Information Fusion*, vol. 41, pp. 195–216, 2018. Citado 2 vezes nas páginas 9 e 54.
- [2] T. Davenport and R. Kalakota, “The potential for artificial intelligence in healthcare,” *Future Healthcare Journal*, vol. 6, pp. 94–98, 2019. Citado na página 15.
- [3] K. Bera, K. A. Schalper, D. L. Rimm, V. Velcheti, and A. Madabhushi, “Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology,” *Nature Reviews Clinical Oncology*, vol. 16, pp. 703–715, 11 2019. Citado na página 15.
- [4] S. Huang, J. Yang, S. Fong, and Q. Zhao, “Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges,” *Cancer Letters*, vol. 471, pp. 61–71, 2020. Citado na página 15.
- [5] S. Sarwar, A. Dent, K. Faust, M. Richer, U. Djuric, R. V. Ommeren, and P. Diamandis, “Physician perspectives on integration of artificial intelligence into diagnostic pathology,” *Digital Medicine*, vol. 2, pp. 1–7, 2019. Citado na página 15.
- [6] B. C. Bizzo, R. R. Almeida, M. H. Michalski, and T. K. Alkasab, “Artificial intelligence and clinical decision support for radiologists and referring providers,” *Journal of the American College of Radiology*, vol. 16, pp. 1351–1356, 2019. Citado na página 15.
- [7] C. Liu, X. Liu, F. Wu, M. Xie, Y. Feng, and C. Hu, “Using artificial intelligence (watson for oncology) for treatment recommendations amongst chinese patients with lung cancer: Feasibility study,” *Journal of Medical Internet Research*, vol. 20, pp. 1–10, 2018. Citado na página 15.
- [8] K. P. Smith, H. Wang, T. J. Durant, B. A. Mathison, S. E. Sharp, J. E. Kirby, S. W. Long, and D. D. Rhoads, “Applications of artificial intelligence in clinical microbiology diagnostic testing,” *Clinical Microbiology Newsletter*, vol. 42, pp. 61–70, 2020. Citado na página 15.
- [9] H. B. Schmitt and E. Franzen, “Utilizando técnicas de mineração de dados como ferramenta de apoio a gestão hospitalar e a tomada de decisão,” *Revista Destaques Acadêmicos*, vol. 11, pp. 200–218, 2020. Citado na página 15.

- [10] A. S. Souza and I. G. Júnior, “A utilização da inteligência artificial no direcionamento de consultas especializadas em um hospital no contexto da health 4.0 the use of artificial intelligence in directing specialized appointment in a hospital in the context of health 4 .,” *Iberoamerican Journal of Project Management*, vol. 10, pp. 1–14, 2019. Citado na página 15.
- [11] L. Davidson and M. R. Boland, “Towards deep phenotyping pregnancy: a systematic review on artificial intelligence and machine learning methods to improve pregnancy outcomes,” *Briefings in Bioinformatics*, vol. 00, pp. 1–29, 2021. Citado 2 vezes nas páginas 15 e 18.
- [12] N. dos Santos Moura, “Desenvolvimento e validação de aplicativo assistencial para predição de pré-eclâmpsia no primeiro trimestre gestacional,” 2020. Citado 2 vezes nas páginas 15 e 18.
- [13] S. Thaddeus and D. Maine, “Too to walk : Maternal mortality in,” *Social Science and Medicine*, vol. 38, pp. 1091–1110, 1994. Citado na página 15.
- [14] M. C. F. Martins, F. E. de Lucena Feitosa, A. B. V. Júnior, L. L. Correia, F. L. P. Ibiapina, R. de Carvalho Pacagnella, and F. H. C. Carvalho, “Pregnancies with an outcome of fetal death present higher risk of delays in obstetric care: A case-control study,” *PLoS ONE*, vol. 14, pp. 1–10, 2019. Citado na página 15.
- [15] M. W. Moreira, J. J. Rodrigues, V. Furtado, C. X. Mavromoustakis, N. Kumar, and I. Woungang, “Fetal birth weight estimation in high-risk pregnancies through machine learning techniques,” *IEEE International Conference on Communications*, vol. 2019-May, 2019. Citado na página 15.
- [16] P. D. Nardo, E. Gentilotti, F. Mazzaferri, E. Cremonini, P. Hansen, H. Goossens, and E. Tacconelli, “Multi-criteria decision analysis to prioritize hospital admission of patients affected by covid-19 in low-resource settings with hospital-bed shortage,” *International Journal of Infectious Diseases*, vol. 98, pp. 494–500, 2020. Citado na página 15.
- [17] G. Büyüközkan and F. Göçer, “Smart medical device selection based on intuitionistic fuzzy choquet integral,” *Soft Computing*, vol. 23, pp. 10085–10103, 2019. Citado na página 15.
- [18] M. Hasan, E. Büyüktaktın, and E. Elamin, “A multi-criteria ranking algorithm for determining breast cancer therapy,” *Omega (United Kingdom)*, vol. 82, pp. 83–101, 2019. Citado na página 15.
- [19] G. T. O’Connor and H. C. Sox, “Bayesian reasoning in medicine,” *Medical Decision Making*, vol. 11, pp. 107–111, 6 1991. Citado na página 16.

- [20] S. Kurzenhäuser and U. Hoffrage, “Teaching bayesian reasoning: an evaluation of a classroom tutorial for medical students,” *Medical Teacher*, vol. 24, pp. 516–521, 1 2002. Citado na página 16.
- [21] S. E. M. T. Haddad, “Rede nacional de vigilância de morbidade materna grave: explorando aspectos metodológicos da abordagem do near miss materno,” 2012. Citado 2 vezes nas páginas 16 e 92.
- [22] J. M. Pinho, “Morbidade materna grave e near miss materno no brasil: Revisão sistemática,” 2017. Citado na página 16.
- [23] U. E. D. E. Campinas, J. F. Cirelli, and M. S. Study, “Jessica fernandes cirelli causas indiretas de morbidade materna grave no brasil : Resultados de um estudo multicêntrico nacional de vigilância indirect causes of severe maternal morbidity in brazil : Results of a national jessica fernandes cirelli causas,” 2017. Citado na página 16.
- [24] M. L. Costa, R. T. Souza, R. C. Pacagnella, S. F. Bento, C. C. Ribeiro-Do-Valle, A. G. Luz, G. J. Lajos, S. B. Mazon, L. Bahamondes, F. G. Surita, G. M. Nobrega, T. B. Griggio, C. M. Charles, M. J. Miele, R. P. Tedesco, K. G. Fernandes, S. Martins-Costa, F. J. Peret, F. E. Feitosa, R. Mattar, E. Traina, E. V. C. Filho, J. Vettorazzi, S. M. Haddad, C. B. Andreucci, J. P. Guida, M. D. C. Junior, M. A. B. Dias, L. G. Oliveira, E. F. M. Junior, C. A. Menezes, M. G. Luz, and J. G. Cecatti, “Brazilian Network of COVID-19 during pregnancy (REBRACO: A multicentre study protocol),” *BMJ Open*, vol. 11, 12 2021. Citado 2 vezes nas páginas 16 e 75.
- [25] S. D. de Azevedo Bittencourt, E. M. Cunha, R. M. S. M. Domingues, B. A. S. Dias, M. A. B. Dias, J. A. Torres, and M. do Carmo Leal, “Nascer no brasil: continuity of care during pregnancy and postpartum period for women and newborns,” *Revista de Saúde Pública*, vol. 54, pp. 1–11, 2020. Citado na página 18.
- [26] A. Akbulut, E. Ertugrul, and V. Topcu, “Fetal health status prediction based on maternal clinical history using machine learning techniques,” *Computer Methods and Programs in Biomedicine*, vol. 163, pp. 87–100, 2018. Citado na página 18.
- [27] R. D. Shirwaikar, D. A. U, K. Makkithaya, S. M, S. Srivastava, and L. E. S. L. U, “Optimizing neural networks for medical data sets: A case study on neonatal apnea prediction,” *Artificial Intelligence in Medicine*, vol. 98, pp. 59–76, 2019. Citado 2 vezes nas páginas 18 e 30.
- [28] Y. Ye, Y. Xiong, Q. Zhou, J. Wu, X. Li, and X. Xiao, “Comparison of machine learning methods and conventional logistic regressions for predicting gestational diabetes using routine clinical data: A retrospective cohort study,” *Journal of Diabetes Research*, vol. 2020, 2020. Citado na página 18.

- [29] F. M. Soares, J. P. Guida, R. C. Pacagnella, J. P. Souza, M. Ângela Parpinelli, S. M. Haddad, and J. G. Cecatti, “Use of intensive care unit in women with severe maternal morbidity and maternal death: Results from a national multicenter study,” *Revista Brasileira de Ginecologia e Obstetricia*, vol. 42, pp. 124–132, 2020. Citado 2 vezes nas páginas 18 e 92.
- [30] Y. Liu, T. Zhao, W. Ju, S. Shi, S. Shi, and S. Shi, “Materials discovery and design using machine learning,” *Journal of Materiomics*, vol. 3, pp. 159–177, 2017. Citado na página 18.
- [31] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, “Machine learning for molecular and materials science,” *Nature*, vol. 559, pp. 547–555, 2018. Citado na página 18.
- [32] S. Brodeur, H. Terrisse, A. Pouchon, O. Godin, B. Aouizerate, V. Aubin, F. Bellivier, R. Belzeaux, T. Bougerol, P. Courtet, C. Dubertret, S. Gard, E. Haffen, C. Henry, M. Leboyer, E. Olié, P. Roux, L. Samalin, R. Schwan, B. Etain, J.-L. Bosson, and M. Polosan, “Pharmacological treatment profiles in the face-bd cohort: an unsupervised machine learning study, applied to a nationwide bipolar cohort,” *Journal of Affective Disorders*, 2021. Citado na página 18.
- [33] M. Sakai, K. Nagayasu, N. Shibui, C. Andoh, K. Takayama, H. Shirakawa, and S. Kaneko, “Prediction of pharmacological activities from chemical structures with graph convolutional neural networks,” *Scientific Reports*, vol. 11, pp. 1–15, 2021. Citado na página 18.
- [34] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, “Secure, privacy-preserving and federated machine learning in medical imaging,” *Nature Machine Intelligence*, vol. 2, pp. 305–311, 2020. Citado na página 18.
- [35] M. J. Willemink, W. A. Koszek, C. Hardell, J. Wu, D. Fleischmann, H. Harvey, L. R. Folio, R. M. Summers, D. L. Rubin, and M. P. Lungren, “Preparing medical imaging data for machine learning,” *Radiology*, vol. 295, pp. 4–15, 2020. Citado na página 18.
- [36] R. Sharma, S. S. Kamble, A. Gunasekaran, V. Kumar, and A. Kumar, “A systematic literature review on machine learning applications for sustainable agriculture supply chain performance,” *Computers and Operations Research*, vol. 119, p. 104926, 2020. Citado na página 18.
- [37] Y. Fu, W. Gou, W. Hu, and Y. Mao, “Life risk factors for future overweight or obesity among preterm infants: A prospective birth cohort,” p. 988, 2013. Citado na página 18.

- [38] E. Díaz, C. Fernández-Plaza, I. Abad, A. Alonso, C. González, and I. Díaz, “Machine learning as a tool to study the influence of chronodisruption in preterm births,” *Journal of Ambient Intelligence and Humanized Computing*, 2021. Citado na página 18.
- [39] O. M. da Saúde (OMS), “Avaliação da qualidade do cuidado nas complicações graves da gestação: A abordagem do near miss da OMS para a saúde materna,” 2011. Citado na página 19.
- [40] Y. Ding, Y. Wang, and D. Zhou, “Mortality prediction for icu patients combining just-in-time learning and extreme learning machine,” *Neurocomputing*, vol. 281, pp. 12–19, 2018. Citado na página 19.
- [41] H. Sufriyana, Y. W. Wu, and E. C. Y. Su, “Artificial intelligence-assisted prediction of preeclampsia: Development and external validation of a nationwide health insurance dataset of the BPJS Kesehatan in Indonesia,” *EBioMedicine*, vol. 54, 2020. Citado na página 19.
- [42] A. Baghban, M. Kahani, M. A. Nazari, M. H. Ahmadi, and W. M. Yan, “Sensitivity analysis and application of machine learning methods to predict the heat transfer performance of cnt/water nanofluid flows through coils,” *International Journal of Heat and Mass Transfer*, vol. 128, pp. 825–835, 2019. Citado na página 19.
- [43] K. K. Venkatesh, R. A. Strauss, C. A. Grotegut, R. P. Heine, N. C. Chescheir, J. S. Stringer, D. M. Stamilio, K. M. Menard, and J. E. Jelovsek, “Machine learning and statistical models to predict postpartum hemorrhage,” *Obstetrics and Gynecology*, vol. 135, pp. 935–944, 2020. Citado na página 19.
- [44] M. W. Moreira, J. J. Rodrigues, N. Kumar, J. Al-Muhtadi, and V. Korotaev, “Nature-inspired algorithm for training multilayer perceptron networks in e-health environments for high-risk pregnancy care,” *Journal of Medical Systems*, vol. 42, 2018. Citado na página 19.
- [45] E. Malacova, S. Tippaya, H. D. Bailey, K. Chai, B. M. Farrant, A. T. Gebremedhin, H. Leonard, M. L. Marinovich, N. Nassar, A. Phatak, C. Raynes-Greenow, A. K. Regan, A. W. Shand, C. J. Shepherd, R. Srinivasjois, G. A. Tessema, and G. Pereira, “Stillbirth risk prediction using machine learning for a large cohort of births from Western Australia, 1980-2015,” 2020. Citado 2 vezes nas páginas 19 e 20.
- [46] A. Akbulut, E. Ertugrul, and V. Topcu, “Fetal health status prediction based on maternal clinical history using machine learning techniques,” *Computer Methods and Programs in Biomedicine*, vol. 163, pp. 87–100, 2018. Citado 2 vezes nas páginas 20 e 38.

- [47] S. A. Rasmussen, J. C. Smulian, J. A. Lednický, T. S. Wen, and D. J. Jamieson, “Coronavirus disease 2019 (covid-19) and pregnancy: what obstetricians need to know,” *American Journal of Obstetrics and Gynecology*, vol. 222, pp. 415–426, 2020. Citado na página 20.
- [48] D. Liu, L. Li, D. Zheng, J. Wang, L. Yang, C. Zheng, D. Liu, L. Li, and X. Wu, “Pregnancy and perinatal outcomes,” *American Journal of Roentgenology*, pp. 1–6, 2020. Citado na página 20.
- [49] F. Donders, R. Lonnée-Hoffmann, A. Tsiakalos, W. Mendling, J. M. D. Oliveira, P. Judlin, F. Xue, and G. G. Donders, “Recommendations concerning covid-19 and pregnancy,” *Diagnostics*, vol. 10, pp. 1–23, 2020. Citado na página 20.
- [50] M. Knight, K. Bunch, N. Vousden, E. Morris, N. Simpson, C. Gale, P. Obrien, M. Quigley, P. Brocklehurst, and J. J. Kurinczuk, “Characteristics and outcomes of pregnant women admitted to hospital with confirmed sars-cov-2 infection in uk: National population based cohort study,” *The BMJ*, vol. 369, 2020. Citado na página 20.
- [51] C. Gao, S. Osmundson, D. R. V. Edwards, G. P. Jackson, B. A. Malin, and Y. Chen, “Deep learning predicts extreme preterm birth from electronic health records,” *Journal of Biomedical Informatics*, vol. 100, p. 103334, 2019. Citado na página 20.
- [52] Y. Baştanlar and M. Özuysal, “Introduction to machine learning,” *Methods in Molecular Biology*, vol. 1107, pp. 105–128, 2014. Citado na página 21.
- [53] T. Bayes, “LII: An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, f. r. s. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S.,” *Philosophical Transactions of the Royal Society of London*, vol. 53, pp. 370–418, 12 1763. Citado 2 vezes nas páginas 21 e 65.
- [54] D. Berrar, “Bayes’ theorem and naive bayes classifier,” *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, vol. 403, 2018. Citado na página 21.
- [55] K. M. Leung, “Naive bayesian classifier,” *Polytechnic University Department of Computer Science/Finance and Risk Engineering*, vol. 2007, pp. 123–156, 2007. Citado na página 23.
- [56] B. Peter and A. Bruce, *Estatística Prática para Cientistas de Dados- 50 Conceitos*, vol. 5. Alta Books. Citado na página 23.
- [57] K. P. Murphy *et al.*, “Naive bayes classifiers,” *University of British Columbia*, vol. 18, no. 60, pp. 1–8, 2006. Citado na página 24.

- [58] A. I. Schein and L. H. Ungar, *Active learning for logistic regression: An evaluation*, vol. 68. 2007. Citado 2 vezes nas páginas 27 e 54.
- [59] L. Zhao, Y. Chen, and D. W. Schaffner, “Comparison of logistic regression and linear regression in modeling percentage data,” *Applied and Environmental Microbiology*, vol. 67, pp. 2129–2135, 2001. Citado na página 27.
- [60] S. I. Lee, H. Lee, P. Abbeel, and A. Y. Ng, “Efficient l1 regularized logistic regression,” *Proceedings of the National Conference on Artificial Intelligence*, vol. 1, pp. 401–408, 2006. Citado na página 30.
- [61] T. Chen, “Scalable and intelligent learning systems,” 2019. Citado 2 vezes nas páginas 30 e 54.
- [62] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 27. Springer US, 2006. Citado 2 vezes nas páginas 32 e 54.
- [63] D. L. Wilson, “Asymptotic properties of nearest neighbor rules using edited data,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 2, pp. 408–421, 1972. Citado 3 vezes nas páginas 32, 87 e 102.
- [64] A. M. Oprescu, G. Miro-Amarante, L. Garcia-Diaz, L. M. Beltran, V. E. Rey, and M. Romero-Ternero, “Artificial intelligence in pregnancy: A scoping review,” *IEEE Access*, vol. 8, pp. 181450–181484, 2020. Citado na página 35.
- [65] A. Braga, A. Carvalho, and T. Ludermir, “Redes neurais artificiais,” *Sistemas Inteligentes: Fundamentos e Aplicações*, 2003. Citado na página 35.
- [66] F. J. V. Zuben, “Redes neurais não-recorrentes 1,” 2020. Citado na página 35.
- [67] S. Haykin, *Neural networks and learning Machines*, vol. 127. McMaster University, 1999. Citado 5 vezes nas páginas 35, 39, 40, 45 e 54.
- [68] Y. H. Shao, W. J. Chen, and N. Y. Deng, “Nonparallel hyperplane support vector machine for binary classification problems,” *Information Sciences*, vol. 263, pp. 22–35, 2014. Citado na página 38.
- [69] A. C. Lorena and A. C. P. L. F. D. Carvalho, “Uma introdução às support vector machines,” 2007. Citado 2 vezes nas páginas 38 e 39.
- [70] A. R. Gonçalves, *Fundamentos e Aplicações de Técnicas de Aprendizado de Máquina*. 2008. Citado 2 vezes nas páginas 39 e 43.
- [71] D. Studi, D. I. Firenze, and A. Passerini, “Kernel methods , multiclass classification and applications to computational molecular biology.” Citado na página 41.

- [72] S. V. Machines, B. Scholkopf, and A. J. Smola, *Learning with Kernels*. 2001. Citado na página 41.
- [73] K. R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, “An introduction to kernel-based learning algorithms,” *IEEE Transactions on Neural Networks*, vol. 12, pp. 181–201, 2001. Citado na página 41.
- [74] Y. Lin, Y. Lee, and G. Wahba, “Support vector machines for classification in nonstandard situations,” *Machine Learning*, vol. 46, pp. 191–202, 2002. Citado na página 42.
- [75] Y. Lee, Y. Lin, and G. Wahba, “Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data,” *Journal of the American Statistical Association*, vol. 99, pp. 67–81, 3 2004. Citado na página 42.
- [76] F. J. V. Zuben, *Sistemas Baseados em Regras e Árvores de Decisão*. 1 ed., 2011. Citado na página 43.
- [77] B. N. B. Lima and L. M. Cioletti, “Entropia: introdução à teoria matemática da (des)informação,” 2004. Citado na página 44.
- [78] N. R. Pal and S. K. Pal, “Entropy: A new definition and its applications,” vol. 21, pp. 1260–1270, 1991. Citado na página 44.
- [79] L. Breiman, “Random forests leo,” *International Journal of Advanced Computer Science and Applications*, vol. 7, pp. 1–33, 2001. Citado na página 45.
- [80] T. Hastie, R. Tibshirani, and J. Friedman, “Springer series in statistics,” *The Elements of Statistical Learning*, vol. 27, p. 83–85, 2009. Citado na página 45.
- [81] M. N. Bernstein, “Random forests,” *Machine Learning*, vol. 45, pp. 5–32, 2011. Citado na página 46.
- [82] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, “Random forests for land cover classification,” *Pattern Recognition Letters*, vol. 27, pp. 294–300, 2006. Citado na página 46.
- [83] L. Breiman, “Classification algorithms and regression trees,” *Classification and Regression Trees*, pp. 246–280, 1984. Citado na página 47.
- [84] G. James, D. Witten, T. Hastie, and R. Tibshirani, *Springer Texts in Statistics An Introduction to Statistical Learning - with Applications in R*. 2013. Citado na página 47.

- [85] M. Bohanec and I. Bratko, “Trading accuracy for simplicity in decision trees,” *Machine Learning*, vol. 15, pp. 223–250, 6 1994. Citado na página 47.
- [86] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Annals of Statistics*, vol. 29, pp. 1189–1232, 2001. Citado 2 vezes nas páginas 53 e 54.
- [87] O. Sagi and L. Rokach, “Approximating XGBoost with an Interpretable Decision Tree,” *Information Sciences*, vol. 572, pp. 522–542, 9 2021. Citado 2 vezes nas páginas 53 e 66.
- [88] T. G. Dietterich, “Ensemble methods in machine learning,” *International Workshop on multiple classifier systems*, pp. 1–15, 2000. Citado na página 54.
- [89] Maiti and Bidinger, *Pattern Classification Using Ensemble Methods*, vol. 53. 1981. Citado na página 54.
- [90] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, “How many trees in a random forest?,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7376 LNAI, pp. 154–168, 2012. Citado na página 54.
- [91] L. K. Hansen and P. Salamon, “Neural network ensembles,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 993–1001, 1990. Citado na página 54.
- [92] A. Fausto and R. Araújo, “Máquinas de comitê.” <https://www.cin.ufpe.br/~aluizioa/RN/RN-09-maquinas-de-comite.pdf>, 2019. Citado na página 55.
- [93] A. Barbez, F. Khomh, and Y. G. Guéhéneuc, “A machine-learning based ensemble method for anti-patterns detection,” *Journal of Systems and Software*, vol. 161, 2020. Citado na página 55.
- [94] G. Brown, “Ensemble learning,” *Encyclopedia of Machine Learning*, p. 312 – 320, 2010. Citado na página 55.
- [95] R. Polikar, “Ensemble based systems in decision making,” *IEEE Circuits and Systems Magazine*, vol. 6, pp. 21–44, 2006. Citado na página 55.
- [96] S. Karlos, G. Kostopoulos, and S. Kotsiantis, “A soft-voting ensemble based co-training scheme using static selection for binary classification problems,” *Algorithms*, vol. 13, 2020. Citado na página 55.

- [97] R. Simanaviciene and L. Ustinovichius, “Sensitivity analysis for multiple criteria decision making methods: Topsis and Saw,” vol. 2, pp. 7743–7744, 2010. Citado na página 56.
- [98] I. Kaliszewski and D. Podkopaev, “Simple additive weighting - a metamodel for multiple criteria decision analysis methods,” *Expert Systems with Applications*, vol. 54, pp. 155–161, 7 2016. Citado na página 57.
- [99] C.-L. Hwang, Y.-J. Lai, and T.-Y. Liu, “A new approach for multiple objective decision making,” 1993. Citado 2 vezes nas páginas 57 e 58.
- [100] M. N. Yahya, H. Gökçekuş, D. U. Ozsahin, and B. Uzun, “Evaluation of wastewater treatment technologies using Topsis,” *Desalination and Water Treatment*, vol. 177, pp. 416–422, 2 2020. Citado na página 57.
- [101] A. F. Tehrani, W. Cheng, K. Dembczyński, and E. Hüllermeier, “Learning monotone nonlinear models using the Choquet Integral,” 2012. Citado na página 59.
- [102] G. D. Pelegrina, L. T. Duarte, M. Grabisch, and J. M. T. Romano, “The multilinear model in multicriteria decision making: The case of 2-additive capacities and contributions to parameter identification,” *European Journal of Operational Research*, vol. 282, pp. 945–956, 2020. Citado 4 vezes nas páginas 60, 63, 70 e 109.
- [103] L. Anzilli and S. Giove, “Multi-criteria and medical diagnosis for application to health insurance systems: a general approach through non-additive measures,” *Decisions in Economics and Finance*, vol. 43, pp. 559–582, 2020. Citado na página 60.
- [104] M. Grabisch and M. Roubens, “Application of the choquet integral in multicriteria decision making,” *Fuzzy Measures and Integrals*, pp. 348 – 374, 2000. Citado 2 vezes nas páginas 60 e 109.
- [105] M. Grabisch and M. Roubens, “Application of the Choquet integral in multicriteria decision making Multi-attribute decision making by consumers associations View project Application of the Choquet Integral in Multicriteria Decision Making,” 2008. Citado na página 60.
- [106] L. T. Duarte and C. Torezzan, *Introduction to the Choquet integral and its application to MCDA Introduction*. Unicamp, 2020. Citado na página 62.
- [107] K. E. Heikes, D. M. Eddy, B. Arondekar, and L. Schlessinger, “Diabetes risk calculator,” *Diabetes Care*, vol. 31, pp. 1040–1045, 5 2008. Citado na página 64.
- [108] U. Mansmann, A. Rieger, B. Strahwald, and A. Crispin, “Risk calculators—methods, development, implementation, and validation,” *International Journal of Colorectal Disease*, vol. 31, pp. 1111–1116, 6 2016. Citado na página 64.

- [109] C. J. Gill, L. Sabin, and C. H. Schmid, “Why clinicians are natural bayesians,” 5 2005. Citado na página 65.
- [110] A. M. Lipsky and R. J. Lewis, “Placing the bayesian network approach to patient diagnosis in perspective,” 2005. Citado na página 65.
- [111] L. Mvo, S. Vm, and A. Tl, “A análise de diagnósticos de enfermagem sob uma perspectiva bayesiana,” 2012. Citado na página 65.
- [112] J. Harbison, B. Mphil, R. Scm, and R. Rnt, “Clinical decision making in nursing: theoretical perspectives and their relevance to practice.” Citado na página 65.
- [113] A. M. N. de Faria Stamm, “Raciocínio clínico no diagnóstico médico,” 2007. Citado na página 65.
- [114] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, “Applications of machine learning to machine fault diagnosis: A review and roadmap,” *Mechanical Systems and Signal Processing*, vol. 138, p. 106587, 4 2020. Citado na página 65.
- [115] V. Tresp, “Committee machines,” vol. 1, pp. 1–12, 2001. Citado na página 66.
- [116] B. Cai, L. Huang, and M. Xie, “Bayesian networks in fault diagnosis,” *IEEE Transactions on Industrial Informatics*, vol. 13, pp. 2227–2240, 10 2017. Citado na página 66.
- [117] G. D. Pelegina, M. Grabisch, L. T. Duarte, and J. M. T. Romano, “Multilinear model: New issues in capacity identification.” Citado na página 70.
- [118] L. Vandenberghe, “The cvxopt linear and quadratic cone program solvers,” *Online: <http://cvxopt.org/documentation/coneprog.pdf>*, 2010. Citado na página 71.
- [119] M. Andersen, J. Dahl, Z. Liu, L. Vandenberghe, S. Sra, S. Nowozin, and S. Wright, “Interior-point methods for large-scale cone programming,” *Optimization for machine learning*, vol. 5583, 2011. Citado na página 71.
- [120] K. G. 1982-Fernandes, “Cor da pele/raça como fator associado à morbidade e mortalidade materna e perinatal: Skin color/race as a factor associated with maternal and perinatal morbidity and mortality,” 2019. Citado na página 92.
- [121] F. T. Fernandes and A. D. P. C. Filho, “Perspectivas do uso de mineração de dados e aprendizado de máquina em saúde e segurança no trabalho,” *Revista Brasileira de Saúde Ocupacional*, vol. 44, pp. 1–12, 2019. Citado na página 93.
- [122] Q. Zhao, J. Chen, F. Li, A. Li, and Q. Li, “An integrated model for evaluation of maternal health care in china,” *Plos One*, vol. 16, p. e0245300, 2021. Citado na página 93.

-
- [123] R. Tibshirani, “Regression Shrinkage and Selection via the Lasso,” *J. R. Statist. Soc. B*, vol. 58, pp. 267–288, 1996. Citado na página 109.
- [124] M. Grabisch, “The application of fuzzy integrals in multicriteria decision making,” *European Journal of Operational Research*, vol. 89, pp. 445–456, 3 1996. Citado na página 109.
- [125] L. S. Shapley, *A Value for n -Person Games*, vol. Volume II. Princeton University Press, 12 1953. Citado na página 109.