

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM
COMPUTACIONAL

Rael Fonseca Andretto

Aprendizado de Máquina para Alocação de Equipagem Ferroviária

Juiz de Fora

2023

Rael Fonseca Andretto

Aprendizado de Máquina para Alocação de Equipagem Ferroviária

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Modelagem Computacional.

Orientador: Prof. D. Sc. Eduardo Pestana de Aguiar

Juiz de Fora

2023

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Andretto, Rael Fonseca Andretto.

Aprendizado de Máquina para Alocação de Equipagem Ferroviária /
Rael Fonseca Andretto. – 2023.

74 f. : il.

Orientador: Eduardo Pestana de Aguiar

Dissertação (Mestrado) – Universidade Federal de Juiz de Fora, Faculdade
de Engenharia. Programa de Pós-Graduação em Modelagem Computacional,
2023.

1. Ferrovia. 2. Aprendizado de Máquina. 3. Inteligência Computacional.
I. Aguiar, Eduardo Pestana de, orient. II. Título.

Rael Fonseca Andretto

Aprendizado de Máquina para Alocação de Equipagem Ferroviária

Dissertação
apresentada ao
Programa de Pós-
Graduação em
Modelagem
Computacional
da Universidade
Federal de Juiz de
Fora como requisito
parcial à obtenção do
título de Mestre em
Modelagem
Computacional. Área
de
concentração: Modelagem
Computacional.

Aprovada em 16 de março de 2023.

BANCA EXAMINADORA

Prof(a) Dr(a). Eduardo Pestana de Aguiar - Orientador
Universidade Federal de Juiz de Fora

Prof(a) Dr(a). Marley Maria Bernardes Rebuzzi Vellasco
Pontifícia Universidade Católica do Rio de Janeiro

Prof(a) Dr(a) . Leonardo Goliatt da Fonseca
Universidade Federal de Juiz de Fora

Juiz de Fora, 15/03/2023.



Documento assinado eletronicamente por **Eduardo Pestana de Aguiar, Professor(a)**, em 16/03/2023, às 10:19, conforme horário oficial de Brasília, com fundamento no § 3º

do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Marley Maria Bernardes Rebuzzi Vellasco, Usuário Externo**, em 16/03/2023, às 10:19, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Leonardo Goliatt da Fonseca, Professor(a)**, em 16/03/2023, às 10:19, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **1188494** e o código CRC **3D159B75**.

Dedico este trabalho a minha mãe Cristina, meu pai José Angelo, minha esposa Luana, minha tia Regina, minha avó Edmé e meus irmãos Rafael e Rafaela.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por me dar forças em todas as manhãs.

À minha mãe Cristina, meu pai Angelo e minha tia Regina por me apoiarem em cada segundo desta caminhada.

À minha esposa Luana, por todo o amor.

À minha avó Edmé e meus irmãos Rafael e Rafaela, pelo carinho.

Aos meus amigos da MRS, pelas conversas, conselhos e trocas de turnos, para que eu pudesse assistir as aulas.

Aos meus gestores da MRS, que viram a importância de trazer análise de dados para dentro da operação ferroviária.

Ao meu orientador Eduardo Aguiar, por ter acreditado em mim desde o início e confiado nesta pesquisa.

Agradeço à Universidade Federal de Juiz de Fora por todo o apoio.

RESUMO

Para se tornarem cada vez mais competitivas no mercado, as empresas ferroviárias buscam otimizar recursos para reduzir custos. Uma grande parcela das despesas ferroviárias está no pagamento da mão de obra, que, em adição ao salário, recebe por horas de prontidão, em transporte, além de pagamento de alimentação e hotéis. Apesar de existirem profissionais especializados, cuja função é realizar a alocação de equipe de forma mais econômica, muitas incertezas existem no momento da decisão, levando a muitos maquinistas a realizarem ciclos ociosos. Técnicas de aprendizado de máquina são ideais para realizar generalizações de novos cenários a partir de uma base de dados de treinamento e podem ser aplicadas para reduzir incertezas em diversos problemas. A literatura científica demonstra que apesar de haver muitos estudos envolvendo otimização de recursos humanos em sistemas de transportes, esta ainda está em desenvolvimento quando é necessária a utilização de técnicas de inteligência computacional, principalmente no meio ferroviário. Neste trabalho, aplicado na MRS Logística S.A., em uma primeira etapa, técnicas de análise exploratória de dados são usadas em conjunto com as regras geradas por uma árvore de decisão, para criar diretrizes a serem aplicadas no dia a dia da alocação de equipagem, a fim de evitar ciclos ociosos. Na segunda etapa, cinco algoritmos de aprendizado de máquina são utilizados para aprender a real necessidade de um maquinista no resto de seu ciclo, ou seja, se haverá ciclo útil ou ocioso, com o objetivo de reduzir a dúvida no momento de sua alocação, podendo até gerar sua dispensa do serviço. Dos algoritmos treinados, além dos mais clássicos como rede neural, máquina de vetor de suporte, árvore de decisão e floresta aleatória, foi utilizado o ALMMo, um modelo neuro-*fuzzy*, evolutivo e não paramétrico, que se atualiza a cada dado novo inserido. Os resultados da primeira etapa mostraram uma redução superior a 50% na ocorrência de ciclos ociosos na empresa. Na segunda etapa do estudo, os algoritmos mostraram uma acurácia acima de 86%, em média, o que atende a todos os níveis de serviço estabelecidos pela empresa. Por fim, tanto a árvore de decisão, a floresta aleatória e o ALMMo mostraram ser soluções adequadas para aplicações na empresa, devido aos seus desempenhos e características.

Palavras-chave: Ferrovia. Aprendizado de Máquina. Inteligência Artificial.

ABSTRACT

Railway companies seek to optimize resources and reduce costs, which leads to becoming increasingly competitive in the market. A large portion of railway expenses is in the workforce's payment, which, in addition to the salary, receives for hours of readiness, transport, food, and hotel expenses. Although there are specialized professionals whose function is to allocate the drivers more economically, many uncertainties exist at the time of the decision, leading many train drivers to carry out idle cycles. Machine learning techniques are ideal for generalizing new scenarios from a training database and can be applied to reduce uncertainties in many problems. The scientific literature shows that although there are many studies about the optimization of human resources in transport systems, this is still under development when it is necessary to use computational intelligence techniques, mainly in the railway environment. In this work, applied at MRS Logística S.A., a Brazilian railway operator, in a first step, exploratory data analysis techniques are used together with the rules generated by a decision tree to create and apply guidelines in crew allocation in the corporation. In the second stage, five machine learning algorithms learn the real need of a train driver for the rest of his cycle to reduce doubts at the time of his allocation, and may even lead to his dismissal from the service. Of the trained algorithms, in addition to the more classic ones such as neural network, support vector machine, decision tree, and random forest, ALMMo was used, a neuro-*fuzzy*, evolutionary and non-parametric model, which updates itself with the addition of new data. The results of the first stage showed a reduction of more than 50% in the occurrence of idle cycles in the company. In the second stage of the study, the algorithms showed an accuracy above 86%, on average, which meets all service levels established by the company. Finally, the decision tree, the random forest, and the ALMMo were considered suitable solutions for application in the company due to their performances and characteristics.

Keywords: Railway. Machine Learning. Artificial Intelligence.

LISTA DE ILUSTRAÇÕES

Figura 1 - Evolução da Produção Ferroviária (Milhões de TU).	12
Figura 2 - Redução de Acidentes Ferroviários.	12
Figura 3 - Revisão Bibliográfica.	17
Figura 4 - Trens Durante a Troca de Equipagem em Dias Tavares.	26
Figura 5 - Ciclo do Maquinista de FDT.	26
Figura 6 - Cenário com um Grande Espaçamento de Trens.	30
Figura 7 - Cenário com Poucos Trens.	30
Figura 8 - Cenário com Muitos Trens.	31
Figura 9 - Exemplo de Hiperplano Criado pelo SVM com Dados de Duas Variáveis.	36
Figura 10 - Perceptron.	38
Figura 11 - Exemplo de Rede Neural.	39
Figura 12 - Funções de Ativação.	40
Figura 13 - Regras <i>Fuzzy</i> do ALMMo.	41
Figura 14 - Projeto de Aprendizado de Máquina.	47
Figura 15 - Correlação Linear entre as Variáveis e o Alvo (2019).	50
Figura 16 - Número de Maquinistas x Espaçamento Anterior de Trens.	51
Figura 17 - Número de Maquinistas x Espaçamento Posterior de Trens.	51
Figura 18 - Árvore de Decisão que Gerou a “Regra de Duplagem”.	53
Figura 19 - Ciclos Ociosos vs Redução Esperada.	54
Figura 20 - Correlação Linear entre as Variáveis e o Alvo (2022).	57
Figura 21 - Ciclos Ociosos x Trens Posteriores.	58
Figura 22 - Gráfico de Caixa para N_Trains e Headway.	58
Figura 23 - Histograma das Variáveis Categóricas.	59
Figura 24 - Curva de Densidade para os Modelos.	63
Figura 25 - Ciclos Ociosos com Floresta Aleatória.	64
Figura 26 - Ciclos Ociosos com Árvore/ALMMo.	65

LISTA DE TABELAS

Tabela 1 – Exemplo de uma Programação de Equipe Um pra Um.	27
Tabela 2 – Exemplo de uma Programação com Atraso no Trem.	28
Tabela 3 – Exemplo de uma Programação de Equipe com Falta.	28
Tabela 4 – Exemplo de uma Programação de Equipe com Sobre.	28
Tabela 5 – Exemplo de uma Programação de Equipe com Sobre e Otimizada. . . .	29
Tabela 6 – Exemplo da Tabela do Teste de Wilcoxon	45
Tabela 7 – Exemplo da Base de Dados de 2019.	49
Tabela 8 – Resultados da Árvore de Decisão.	52
Tabela 9 – Exemplo da Base de Dados de 2022.	56
Tabela 10 – Hiperparâmetros Testados.	61
Tabela 11 – Hiperparâmetros Selecionados.	62
Tabela 12 – Desempenho dos Modelos (com Desvio Padrão).	62
Tabela 13 – Valores P do Teste de Wilcoxon para a Acurácia dos Modelos.	63
Tabela 14 – Valores P do Teste de Wilcoxon para a Revocação dos Modelos.	63

SUMÁRIO

1	INTRODUÇÃO	11
1.1	OBJETIVOS	15
1.1.1	Objetivo Geral	15
1.1.2	Objetivos Específicos	15
1.2	CONTRIBUIÇÕES	15
1.3	ORGANIZAÇÃO DO TRABALHO	16
2	REVISÃO BIBLIOGRÁFICA	17
2.1	ALOCAÇÃO DE TRIPULAÇÃO EM SISTEMAS DE TRANSPORTES	17
2.2	APRENDIZADO DE MÁQUINA NA ALOCAÇÃO DE TRIPULAÇÃO EM SISTEMAS DE TRANSPORTES	18
2.3	OTIMIZAÇÃO DA ALOCAÇÃO DE TRIPULAÇÃO NA MRS LOGÍSTICA S.A.	19
2.4	CONCLUSÃO DA REVISÃO BIBLIOGRÁFICA	20
3	FORMULAÇÃO DO PROBLEMA	22
3.1	PRINCIPAIS CONCEITOS DE ALOCAÇÃO DE EQUIPAGEM FERROVIÁ- RIA	22
3.2	REGULAMENTAÇÃO DO TRABALHO DA EQUIPAGEM FERROVIÁ- RIA	23
3.3	REDUÇÃO DE CICLOS OCIOSOS NA MRS LOGÍSTICA S.A.	25
4	REFERENCIAL TEÓRICO	32
4.1	ANÁLISE EXPLORATÓRIA DE DADOS	32
4.2	PREPARAÇÃO DOS DADOS	33
4.3	AJUSTE DE HIPERPARÂMETROS, REGULARIZAÇÃO E VALIDAÇÃO CRUZADA	33
4.4	MODELOS DE APRENDIZADO DE MÁQUINA	34
4.4.1	Árvore de Decisão	34
4.4.2	Floresta Aleatória	35
4.4.3	Máquina de Vetor de Suporte	36
4.4.4	Rede Neural	38
4.4.5	ALMMo	40
4.5	MÉTRICAS DE CLASSIFICAÇÃO BINÁRIA	43
4.6	TESTE ESTATÍSTICO	44
5	EXPERIMENTOS COMPUTACIONAIS	47
5.1	O EXPERIMENTO DE 2019	48
5.1.1	Estudo do Problema	48
5.1.2	Base de Dados	48
5.1.3	Análise Exploratória de Dados	49

5.1.4	Preparação dos Dados	52
5.1.5	Ajuste de Hiperparâmetros	52
5.1.6	Resultados	52
5.1.7	Aplicação na Empresa	53
5.2	O EXPERIMENTO DE 2022	54
5.2.1	Estudo do Problema	54
5.2.2	Base de Dados	55
5.2.3	Análise Exploratória de Dados	56
5.2.4	Preparação dos Dados	59
5.2.5	Ajuste de Hiperparâmetros	60
5.2.6	Resultados	61
5.2.7	Aplicação na Empresa	64
6	CONCLUSÃO	67
6.1	TRABALHOS FUTUROS	68
	REFERÊNCIAS	69

1 INTRODUÇÃO

A história do transporte ferroviário iniciou no século XVI, para atender às necessidades das minas de minério de ferro e carvão, com o transporte de vagões sobre trilhos de madeira, inicialmente puxados por homens ou carroças [44]. Com o passar dos anos, a ferrovia foi sendo aprimorada com a substituição do trilho de madeira por ferro, a inclusão de bordas nas rodas dos veículos, a criação da máquina a vapor de James Watt e o primeiro veículo capaz de se movimentar com a pressão do vapor, inventado por Richard Trevithick, em 1804 [59]. Aperfeiçoando o modelo de Trevithick, George Stephenson construiu a locomotiva *Locomotion* que circulou, em 1825, na *Stockton & Darlington Railway*, a primeira ferrovia pública do mundo, também construída por Stephenson [44].

Somente em 1854 a ferrovia chegaria ao Brasil, com a inauguração da Estrada de Ferro Mauá, que seguia por 14,5 km entre a estação de Guia de Pacobaíba (Magé - RJ) e Raiz da Serra (atual Vila Inhomirim, Petrópolis - RJ), construída por Irineu Evangelista de Souza, o Barão de Mauá. Em 1858, foram inauguradas a estrada de ferro Recife a São Francisco e a Estrada de Ferro Dom Pedro II (esta posteriormente renomeada Central do Brasil), sendo seguidas por diversas outras iniciativas por todo país [60].

O sucesso financeiro das empresas ferroviárias no Brasil foi comprometido por diversos fatores, como políticas que favoreceram ao automóvel, a sazonalidade de produção de certas mercadorias, a falta de produtos de retorno e o alto custo de implantação da via e de compra de veículos. Além disso, devido a terem sido construídas de forma independente por entidades diversas, cada ferrovia possuía uma bitola (espaçamento entre trilhos) diferente, o que impedia o intercâmbio de trens, dificultando a operação por longas extensões, que é um dos viabilizadores econômicos do modal. Em 1957, foi fundada a RFFSA (Rede Ferroviária Federal S.A.), um esforço do governo federal em absorver as ferrovias deficitárias e mantê-las funcionando. Porém, o cenário econômico das ferrovias brasileiras não mudou sob a gestão da Rede, continuando ainda os deficits, com exceção da superintendência da Central do Brasil [49].

Iniciada em 1975, a Ferrovia do Aço foi um projeto desafiador que pretendia ligar as cidades de Belo Horizonte, Volta Redonda e São Paulo, em linha dupla, com rampas máximas de 1% e raio mínimo de 900 metros, sendo um percurso montanhoso que necessitava da construção de diversos túneis e pontes. Com a situação financeira do país agravada pela crise internacional do petróleo, o projeto precisou ser revisto, sendo concluído somente o trecho de Jeceaba a Saudade em 1989, em linha singela, após parceria com as empresas mineradoras da região. Em 1993, a Ferrovia do Aço foi estendida até Itabirito [67]. Com o projeto feito somente em linha singela, a circulação dos trens passou a ser realizada em forma de “carrossel”, no qual os trens faziam a circulação vazios pela Linha do Centro (Central do Brasil) e desciam carregados pela Ferrovia do Aço.

Visando o aumento da produção e redução de acidentes, em março de 1992, a RFFSA foi incluída no Plano Nacional de Desestatização, que previa a concessão do transporte de carga para o setor privado, com todas as malhas incluídas no plano leiloadas até 1998 [49]. Neste modelo de operação, empresas privadas usufruem da malha e ativos ferroviários concedidos pelo governo, em contrapartida estas devem atender a metas de produção, segurança e investimentos, além de pagarem pela exploração do transporte. De acordo com a ANTF (Associação Nacional dos Transportadores Ferroviários) [7], um novo patamar no modal ferroviário foi alcançado com a privatização, com o aumento do volume transportado pelas ferrovias brasileiras (Figura 1) e a redução do índice de acidentes ferroviários [13] (Figura 2).



Figura 1 - Evolução da Produção Ferroviária (Milhões de TU).
Fonte: ANTF (2022) [7].

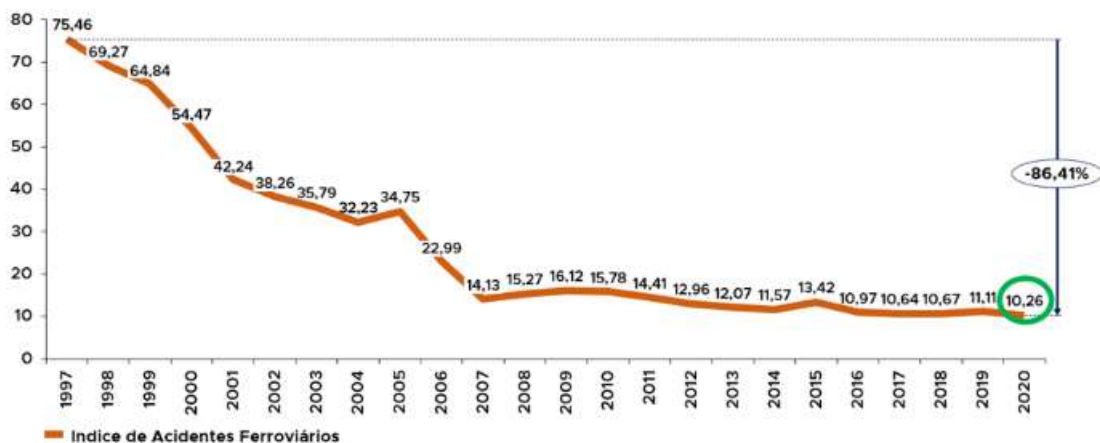


Figura 2 - Redução de Acidentes Ferroviários.
Fonte: ANTF (2022) [7].

As expectativas para o setor são de intenso crescimento: em março de 2022, o transporte de cargas por ferrovias apresentou um crescimento de 30% em relação ao mesmo período do ano anterior [10]. Ainda, são previstos mais de R\$250 bilhões de investimentos com o novo modelo de autorizações [12], que permite que empresas privadas construam suas próprias ferrovias, e mais investimentos vindos das operadoras com o adiantamento da renovação de suas concessões [14]. Atualmente o modal ferroviário é responsável por 14,95% do transporte de cargas no Brasil [19], porém é estimado que esta participação possa aumentar para 35% nos próximos anos [59].

Dado o histórico de anos de priorização da rodovia no país e possuindo somente uma pequena parcela da matriz de transporte nacional, as empresas ferroviárias buscam reduzir seus custos e otimizar processos para se tornarem cada vez mais competitivas no mercado e captar novos clientes. Dentro dos custos operacionais, destaca-se o custo com mão de obra qualificada, que não só é escassa, demandando anos de formação, como também sujeita a pagamento por horas improdutivas, isto é, horas de aguardo pelo início da jornada em trem, horas extras, em transportes, além dos custos com alimentação e hotel [53].

Além disso, um planejamento inadequado pode levar a não haver maquinistas o suficiente para atender à demanda. A falta de maquinistas é um problema que afeta as operadoras ferroviárias pelo mundo todo. De acordo com a BBC [47], aproximadamente um quinto dos atrasos de trens na Inglaterra entre 2013 e 2015 foram causados por falta de maquinistas. Em 2019, foi noticiado no *The Times* que mais de 35 mil trens foram cancelados nos últimos seis anos por falta de tripulação [20]. Com a pandemia de covid-19, a situação das ferrovias foi agravada, além dos constantes afastamentos por sintomas da doença, o número de contratações de maquinistas foi abaixo do esperado devido às escolas de treinamento para novos maquinistas terem sido canceladas [29].

No transporte de passageiros, principalmente na Inglaterra, em que o metrô existe há mais de um século, os horários de pico e sazonalidades são bem conhecidos, facilitando o dimensionamento de recursos. Ainda assim, conforme as notícias citadas, a alocação de maquinistas é um desafio. O problema se torna ainda mais complexo no setor de cargas, em que a demanda pode variar a qualquer momento, dependendo de estoque de clientes e até mesmo de variação na cotação do dólar.

Na MRS Logística S.A., empresa ferroviária que possui a concessão de malhas dos estados de Minas Gerais, Rio de Janeiro e São Paulo, o gasto com salários de funcionários chega a 155 milhões de reais por trimestre [42], aproximadamente R\$51,7 milhões mensais, sendo grande parte dos funcionários diretamente ligados com a operação. Com a renovação da concessão da empresa, que estenderá o contrato em mais 30 anos, e a implantação pelo governo federal do novo modelo de autorização de ferrovias, o desafio de alocação de maquinistas será ainda maior, pois haverá um aumento de demanda em curto prazo que

será maior que o da mão de obra, dado que o tempo necessário para formação pode levar anos.

Neste contexto, a redução de ciclos ociosos de maquinistas gera diretamente economia financeira para a empresa, pois, além de reduzir os custos relacionados ao ciclo, também pode evitar a falta de tripulação em trem, já que, se dispensado, o maquinista estará disponível novamente em sua sede antecipadamente, podendo evitar um futuro cenário de falta de tripulação no local. Ao ser liberado, o maquinista recebe pelas oito horas de trabalho (sem considerar horas extras e improdutivas) e deve estar disponível novamente em sua sede após doze horas de sua dispensa. Com isso, a empresa economiza no pagamento de horas improdutivas, extras, em transporte e hotéis, já o maquinista recebe pela hora trabalhada normal, além de ter um imediato descanso, ao invés de realizar um ciclo sem necessidade. Dado o cenário de falta de tripulação para trem já descrito, a liberação do maquinista retirará este de um cenário em que está ocioso para um cenário em que ele será necessário.

Na literatura, a alocação de equipagem é feita através de modelos de otimização que buscam encontrar a combinação de maquinistas e trens que geram o cenário mais econômico. Porém, para que esta solução funcione, é necessário que os horários de trens sejam conhecidos, caso contrário, pode-se gerar um desperdício de recursos. Em locais em que a circulação de trens não tem grade horária fixa, como no trecho ferroviário compreendido entre as cidades de Juiz de Fora - MG e Jeceaba - MG, alguns maquinistas realizam ciclos ociosos devido ao pouco conhecimento de cenários futuros. Para contornar esta situação, a previsão do real uso do maquinista em seu ciclo será realizada através de modelos de inteligência computacional.

O estudo de técnicas de aprendizado de máquina iniciou-se no século passado. Um dos primeiros estudos que visou modelar um neurônio matematicamente foi realizado em 1943 por McCulloch e Pitts [40]. Em 1958, Rosenblatt [56] definiu o modelo *perceptron*, que seria a base para a criação das futuras redes neurais. Em 1986, com a generalização da regra delta, também conhecida como *backpropagation* [57], através da retropropagação do erro da função de custo para os neurônios, os pesos da rede puderam ser atualizados de acordo com os dados. Logo surgiram outras variações de redes neurais como a convolucional [37], a rede neural recorrente [30] e as redes neurais generativas adversariais [26].

Além das redes neurais, outros modelos capazes de aprender com os dados foram desenvolvidos, como a máquina de vetor suporte (*support vector machine*, ou SVM) [9], capaz de gerar o hiperplano que maximiza a diferença entre as classes de dados. Os modelos de árvore de decisão [15] ganharam bastante destaque por serem capazes de explicitar claramente suas regras de classificação. Já ao criar um agrupamento de várias árvores de decisão, as florestas aleatórias [16] mostraram-se ser um modelo de alta performance ao prever a classe mais votada pelos submodelos.

Recentemente, modelos evolutivos e não paramétricos, como o ALMMo (*Autonomous Learning Multi-Model System*) [5], têm surgido e ganhado bastante espaço. A principal característica deste sistema *fuzzy* evolutivo, além de sua capacidade de predição, é ser um modelo que aprende sem nenhuma suposição prévia sobre os dados, ou seja, sem que haja nenhum hiperparâmetro para ser ajustado. O ALMMo também possui importantes atributos comuns a outros modelos como não necessitar que os dados sejam distribuídos normalmente e nem escalados.

Apesar de muitos destes modelos terem surgido no século passado, somente com o aumento da capacidade de processamento dos computadores é que soluções baseadas em aprendizado de máquina ganharam ampla aplicação. Atualmente, inteligência computacional é utilizada com sucesso em diversas áreas como medicina [62], direito [63], jogos [23], carros autônomos [52], criação de artes [61] e também nas áreas de transportes [45]. Na ferrovia, o uso de aprendizado de máquina é muito concentrado nos campos de manutenção e gerenciamento de tráfego, sendo o uso em alocação de tripulação um campo ainda em desenvolvimento, com muita oportunidade de novos estudos [66].

1.1 OBJETIVOS

1.1.1 Objetivo Geral

O principal objetivo deste trabalho é desenvolver uma solução, baseada em técnicas de inteligência computacional, para reduzir o número de maquinistas alocados em um trecho ferroviário em forma de carrossel da MRS Logística S.A. No estudo, serão criadas diretrizes para a alocação da tripulação excedente e será avaliada a performance de algoritmos de aprendizado de máquina para aplicação na empresa.

1.1.2 Objetivos Específicos

1. Desenvolver diretrizes baseadas em estudo dos dados para a sobre-alocação de um maquinista;
2. Avaliar a performance de algoritmos de aprendizado de máquina para a previsão da necessidade da alocação do maquinista.

1.2 CONTRIBUIÇÕES

As principais contribuições deste trabalho são:

- Ampliação da literatura existente através de um estudo inédito que utiliza modelos de aprendizado de máquina para prever e reduzir a realização de ciclos ociosos de maquinistas;

- Avaliação de Rede Neural, SVM, Árvore de Decisão, Floresta Aleatória e ALMMo para solução do problema e posterior recomendação de utilização na empresa;
- Aplicação em ambiente corporativo de uma solução baseada em árvore de decisão que obteve resultados satisfatórios.

As principais conclusões deste trabalho são:

- As regras definidas através da árvore de decisão reduziram em mais da metade a ocorrência de ciclos ociosos na empresa avaliada;
- O uso de modelos de aprendizado de máquina seria ainda mais eficiente na economia de maquinistas que o método de regras;
- O modelo de floresta aleatória possui melhor acurácia para o caso avaliado, porém algoritmos com regras mais explícitas obtêm resultados semelhantes e podem ser melhor aceitos em ambiente corporativo.

1.3 ORGANIZAÇÃO DO TRABALHO

Esta dissertação está dividida em mais cinco capítulos. Uma revisão da bibliografia de assuntos correlatos à solução do problema é apresentada no Capítulo 2. No Capítulo 3 é formulado o problema a ser resolvido através de técnicas de inteligência computacional. A base teórica sobre aprendizado de máquina é desenvolvida no Capítulo 4. No Capítulo 5 os resultados são apresentados e discutidos, ressaltando as aplicações e próximos passos. Por último, é concluído o estudo e são apresentadas sugestões de futuros trabalhos.

2 REVISÃO BIBLIOGRÁFICA

A revisão bibliográfica foi dividida em três seções, de forma a melhor expor os assuntos que compõem este estudo. Primeramente, será feita uma análise da literatura de alocação de tripulação em sistemas de transportes, destacando os diferenciais da aplicação em ferrovias. Em seguida, os estudos que aplicaram aprendizado de máquina em alocação de tripulação serão abordados. Por último, serão analisadas as pesquisas realizadas na MRS Logística S.A. com respeito à economia nos custos de equipagem, de forma a localizar este trabalho. A Figura 3 demonstra os tópicos que compoem esta pesquisa.

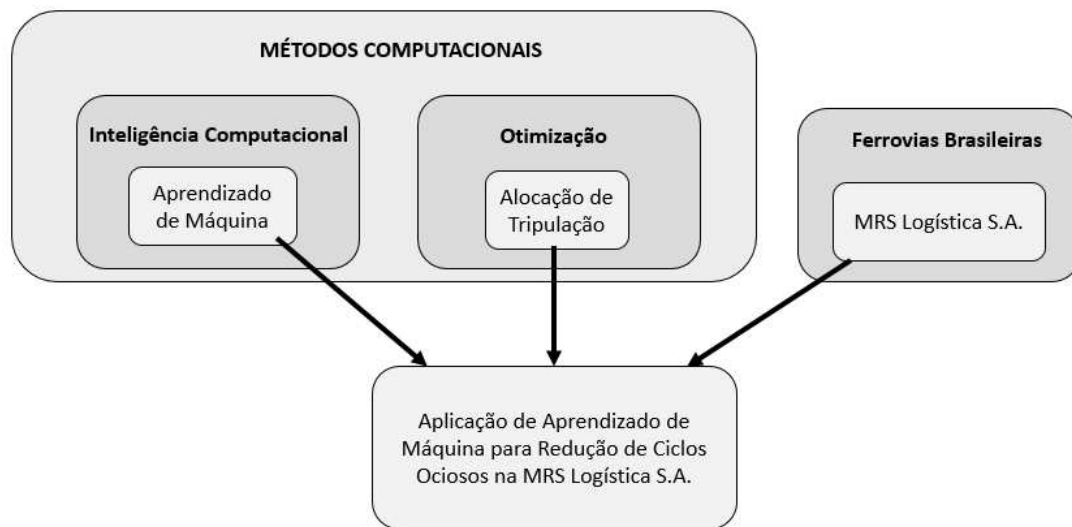


Figura 3 - Revisão Bibliográfica.
Fonte: O autor (2023).

2.1 ALOCAÇÃO DE TRIPULAÇÃO EM SISTEMAS DE TRANSPORTES

O problema de agendamento e alocação de recursos humanos em sistemas de transportes, *Crew Scheduling Problem* (CSP), é algo em constante pesquisa no meio científico. Porém, dada a natureza do problema, essas aplicações estão muitas vezes restritas ao uso de técnicas de otimização, como programação linear ou mesmo metaheurísticas. Silva e Reis (2014) [64] resolvem o CSP para uma aplicação em uma companhia de ônibus da cidade de Belo Horizonte - MG, em que comparam diferentes metaheurísticas como pesquisa local iterada, busca de vizinhança variável e busca de vizinhança em grande escala para chegar a soluções com melhor resultado que as atuais utilizadas pela empresa. Em Kasirzadeh, Saddoune e Soumis (2017) [34], uma revisão de literatura de aplicações do CSP no transporte aéreo é feita e destaca-se que as soluções estão restritas a modelar o CSP como um problema de cobertura de conjuntos ou partição de conjuntos e solucioná-lo com técnicas como relaxação Lagrangiana ou gerador de colunas. O autor ainda mostra que apesar de possuir pequenas variações dependendo do local ou da companhia aérea em estudo, algumas restrições do problema são comuns como: horas de descanso entre

turnos, quantidade de tarefas permitidas dentro de um mesmo turno e quantidade de turnos permitidos no mês. Uma modelagem do CSP em programação inteira é feita por Leggate *et al.* (2018) [38] para alocar tripulações de navios de abastecimento de plataformas. Destacam-se neste trabalho os grandes tempos de planejamento necessários, pois as equipes realizam jornadas de trabalhos embarcadas que podem durar quatro semanas.

Assim como os meios de transportes citados, no modal ferroviário também há diferentes pesquisas e aplicações do CSP, com variações de restrições em relação ao tipo de ferrovia, regulamentações locais ou regras da empresa em estudo. Em Caprara *et al.* (2000) [17] é desenvolvido um dos primeiros estudos de CSP para ferrovias, *Railway Crew Scheduling Problem* (RCSP). Entre as restrições do problema estão: o maquinista deve retornar ao seu local de trabalho após o término de um serviço e existência de tempo mínimo de descanso entre tarefas e serviços. Os autores formulam o problema de diversas formas, como programação linear inteira, partição de conjuntos, cobertura de conjuntos e até uma heurística para RCSP em grande escala. Um solução com gerador de colunas para resolver problemas de longo prazo em RCSP é apresentada por Albers (2009) [2]. Buscando uma melhor eficiência computacional, um algoritmo de decomposição “divisão e preço” foi implementada por Silke (2019) [33] para resolver de forma mais rápida o mesmo problema de Albers (2009) [2].

2.2 APRENDIZADO DE MÁQUINA NA ALOCAÇÃO DE TRIPULAÇÃO EM SISTEMAS DE TRANSPORTES

Técnicas de aprendizado de máquina são bastante usadas para solucionar problemas ferroviários. Mas, como descrito por Tang *et al.* (2022) [66], as pesquisas estão mais concentradas nas áreas de manutenção e inspeção. Já a área de planejamento, onde encaixa-se a alocação de equipe, realizou estudos mais focados em gestão da circulação de trens e locomotivas. Após extensa pesquisa, a única publicação com aplicação de aprendizado de máquina em RCSP encontrada foi a desenvolvida por Gattermann-Itschert, Poreschack e Thonemann (2022) [24]. Neste trabalho, há uma restrição em que o agendamento de equipes deve ser feito com uma distribuição equilibrada de serviços considerados ruins pelos maquinistas. Para resolver essa restrição, os autores utilizaram uma floresta aleatória para prever se o serviço seria considerado bom ou ruim e posteriormente aplicá-lo dentro do algoritmo de otimização. Ainda, foi utilizado um método para computar a importância de cada variável e concluiu-se que o fator mais representativo para ser considerado um serviço bom ou ruim era a duração da atividade.

Já no transporte aéreo, o uso de técnicas de inteligência computacional aliadas ao CSP é mais comum. Em Gopalakrishnan e Johnson (2005) [28], os autores desenvolveram uma rede neural *feedforward* para prever a probabilidade de um serviço ser alocado a

uma tripulação, de acordo com suas características, e reduzir o custo computacional da otimização do CSP, porém os resultados dessa aplicação não foram conclusivos, de acordo com os autores. Já em Yaakoubi, Soumis e Lacoste-Julien (2020) [72], uma rede neural convolucional foi desenvolvida para prever qual avião seria o próximo que o piloto seria designado, de forma a reduzir a complexidade do algoritmo de otimização. Neste trabalho, os autores obtiveram sucesso e reduziram o custo computacional quando comparado à solução somente com otimização. É possível perceber que a aplicação de inteligência computacional em problemas de agendamento/alocação de recursos humanos em sistemas de transportes é um tema ainda pouco explorado e com muitas oportunidades de desenvolvimento. As aplicações existentes necessitam de uma grade horário fixa dos trens e usam o aprendizado de máquina para reduzir o custo computacional da otimização, ou para contornar alguma restrição do problema, enquanto o presente trabalho busca prever a necessidade de maquinistas em cenários futuros sem um horário fixo dos trens.

2.3 OTIMIZAÇÃO DA ALOCAÇÃO DE TRIPULAÇÃO NA MRS LOGÍSTICA S.A.

Métodos e soluções para otimização da alocação de equipagens ferroviárias estão em constante pesquisa pela MRS Logística S.A. Em Santos (2007) [58], para que os maquinistas realizem o mínimo de horas extras e horas ociosas possíveis, o autor propõe um modelo para definição dos locais ideais para serem sedes. Baseado nas restrições de tempos de trabalho dispostos na Consolidação das Leis do Trabalho, CLT [11], e de posse dos tempos de circulação na malha da MRS, o modelo cria uma árvore de soluções de escolhas de postos e a solução ideal é o ramo que reduziu o maior número de horas extras ou ociosas.

Fernandes (2012) [21] defendeu a criação do cargo de programador de equipagens, com o objetivo de ter um colaborador dedicado à otimização das alocações, visando o atendimento a todos os trens, mas também a reduzir o número de horas improdutivas dos maquinistas. Nesta monografia, o autor compara o método de alocação antigo “descentralizado”, realizado por cada operador de escala em cada sede, com o modelo proposto “centralizado”, em que haveria um colaborador lotado no centro de controle que absorveria a função de alocação das equipagens. Conforme explicado pelo autor, o operador de escalas era responsável por alocar as equipes em trens, planejar a logística dos carros da empresa, realizar teste de etilômetro nos maquinistas, garantir que estes recebessem a alimentação e fazer a digitação de todas as suas atividades. Logo, sobrava pouco tempo para que o operador de escalas dedicasse a verificar o cenário de trens e realizasse a alocação mais econômica, geralmente seguindo somente o FIFO (*First-in First-out*). Com a criação da função do programador de equipagens, este novo cargo teria uma visão macro da circulação dos trens em diversas sedes e poderia fazer a alocação buscando o ganho ótimo do sistema, enquanto o operador de escalas ficaria focado nas outras tarefas de

execução. Para concluir, o autor realiza simulações de cenários de alocações, mostrando a redução em horas improdutivas que um colaborador dedicado e com maior visão sistêmica traria para a empresa. A criação do cargo de programador de equipagens foi uma das primeiras grandes iniciativas da MRS Logística S.A. para a otimização das alocações.

Outra forma de reduzir os custos relacionados à equipagem é realizando uma logística de carro eficiente, reduzindo a quilometragem rodada por maquinista e consequentemente o custo das horas em trânsito. Usando análise envoltória de dados em cenários futuros gerados a partir de regressão linear, Silva (2013) [65] mostrou que uma melhor distribuição dos carros que fazem a logística dos maquinistas em um trecho da malha da MRS Logística S.A., no estado do Rio de Janeiro, seria necessária para diminuir o número de horas extras e horas improdutivas pagas. Os modelos de regressão linear usaram o volume transportado pela empresa e a quantidade de maquinistas na localidade como dados de entrada para prever a quilometragem percorrida em carro no ano seguinte e a quantidade de carros necessária. Além disso, foram realizadas regressões para estimar também as horas extras pagas para os maquinistas, chegando a valores de alto custo.

O método simplex de programação linear foi utilizado por Rodrigues (2015) [53] para gerar a alocação de equipagem mais econômica no trecho de Juiz de Fora a Lafaiete. Os indicadores a serem minimizados foram: redução de horas de prontidão, redução de THP (Trem Hora Parado), redução de jornadas maiores que oito horas e redução de jornadas que gerassem reprogramações. Para aplicar a metodologia, o autor utilizou o solver do *Microsoft Excel* e mostrou que a solução do otimizador é melhor que a utilização simples do FIFO.

Trinchinato (2017) [68] mostrou que a aplicação do modelo centralizado, defendido por Fernandes (2012) [21], de alocação nas sedes de Itu e Jundiaí, da MRS Logística S.A., reduziu 2800 horas improdutivas por mês, comparando os anos de 2014 e 2015. Além disso, o autor mostra que a criação de uma sede em Campinas reduziu em sete mil reais o custo mensal com horas de passagem.

A redução de horas improdutivas de maquinistas na MRS Logística S.A. é tratada por Fuzeto, Silveira e Castro (2017) [22] através de uma proposta de flexibilização das habilitações. Os autores mostram que, no posto de Embu-Guaçu, a quantidade de horas improdutivas é positivamente correlacionada com a quantidade de maquinistas trocados por não terem habilitação para seguirem viagem. Analisando o oitavo decil dos tempos de circulação e de horas improdutivas, é proposto um aumento de habilitação dos maquinistas para mais 34 quilômetros.

2.4 CONCLUSÃO DA REVISÃO BIBLIOGRÁFICA

São poucos os estudos que aplicaram aprendizado de máquina em alocação de tripulação em transportes publicados até a presente data, sendo somente um deles específico

em ferrovia. Além destes estudos terem sido modelados para os casos específicos em que foram aplicados, seus objetivos se diferem do presente trabalho, em que será aplicado o aprendizado de máquina para conseguir uma alocação mais econômica em cenários de imprevisibilidade de trens. Já os trabalhos publicados buscaram reduzir complexidade de otimização, ou resolver restrições. Desta forma, pode-se concluir que o presente estudo amplia a literatura existente ao aplicar aprendizado de máquina de forma a fornecer diretrizes para sobre-alocar ou dispensar um maquinista na programação de equipagem diária. Em relação às pesquisas realizadas na MRS Logística S.A., este estudo pode ser considerado uma sequência do trabalho de Rodrigues (2015) [53], que modelou a otimização do trecho de Juiz de Fora à Lafaiete, porém sem especificar quando dispensar ou sobre-alocar os maquinistas que estão sobrando na alocação.

3 FORMULAÇÃO DO PROBLEMA

3.1 PRINCIPAIS CONCEITOS DE ALOCAÇÃO DE EQUIPAGEM FERROVIÁRIA

Para conduzir as locomotivas dos trens é necessário ter um maquinista devidamente habilitado. A situação em que o profissional realiza a circulação do trem sozinho, sem a presença de um auxiliar, é chamada de monocondução [55] e já é realidade nas maiores empresas de carga do país.

O planejamento da equipagem ferroviária é o processo que garante que todos os trens tenham um maquinista para sua condução, da forma mais econômica e satisfazendo às restrições de cada empresa [1]. Como dito por Fuzeto, Silveira e Castro (2017) [22], o planejamento de equipes é feito em vários estágios: orçamento (anual), plano mensal, plano semanal e plano diário. Em outra classificação, o orçamento pode ser considerado como o planejamento estratégico, os planos mensais e semanais como planejamento operacional e o plano diário como um plano de ultra-curto prazo [1].

De acordo com Albers (2009) [2] e Fuzeto, Silveira e Castro (2017) [22], as principais restrições para um problema de planejamento de equipagem ferroviária são: evitar que os trens fiquem sem maquinistas (alocação de pelo menos um maquinista habilitado), reduzir o THP (Trem Hora Parado), redução do número de horas extras, redução do pagamento de hotéis e custos de transporte de carro (quando não há trem para o maquinista ir do ponto A ao B e é necessário usar um carro privado).

Em [1], os autores descrevem os principais conceitos de alocação de equipagem:

- Tarefa: Um serviço único, como, por exemplo, levar um trem do pátio A ao pátio B.
- Serviço: Um conjunto de tarefas a serem feitas em um ciclo. Exemplo: Conduzir um trem do pátio A ao pátio B, tirar o tempo de descanso regulamentado por lei e retornar de trem ao pátio de origem.
- Destacamento: Local onde o serviço inicia e/ou termina. Para redução de custos é ideal que a cidade em que esteja o destacamento do maquinista, também seja sua cidade de residência.
- Alocação: Designar uma tarefa a um maquinista.
- Tarefa Sobre-Coberta (do original “*over-covered task*”): Alocar mais de um maquinista para o mesma tarefa sem que haja necessidade. Apesar de inicialmente parecer dispendiosa, a tarefa “sobre-coberta” pode ser necessária visando uma redução do custo global, seja retornando dois maquinistas em um trem para sua sede origem (por falta de outro trem para seu retorno) ou enviando dois maquinistas em um só trem para um pátio, com o objetivo de que neste pátio cada um dos maquinistas

assumam um trem diferente. No jargão ferroviário, “sobre-cobrir”, ou “sobre-alocar” um trem é também chamado de “duplar”. O termo “duplar” será utilizado ao longo deste trabalho, como um sinônimo de sobre-cobrir.

As tarefas que podem ser atribuídas a um maquinista são descritas por Trinchinato (2017) [68]:

- Manobra: Tarefa executada no destacamento de origem, sem deslocamento para outra sede.
- Auxílio: Tarefa semelhante a de manobra, porém específica para acoplar locomotivas na cauda de trens, com o objetivo de fornecer maior tração para vencer rampas acentuadas.
- Jornada de trabalho sem reaproveitamento: Viagem curta, na qual, após a chegada no destino, o maquinista retorna de carro para seu destacamento no final da jornada e segue para descanso em sede (22 horas). Essa tarefa também é conhecida como escala de retorno;
- Jornada de trabalho com reaproveitamento: viagem longa, com descanso fora da sede (10 horas) e retorno à sede em viagem de trem. Esse item é classificado como um conjunto de tarefas (trem de ida, descanso e trem de volta), ou seja, um serviço e também pode ser chamado de escala de prontidão.

3.2 REGULAMENTAÇÃO DO TRABALHO DA EQUIPAGEM FERROVIÁRIA

De forma a embasar o estudo do modelo para equipagem ferroviária, é importante entender as leis que regem esta função, ou seja, a Consolidação das Leis do Trabalho [11]. A definição do que é considerado serviço ferroviário está no artigo 236, sendo a equipagem dos trens categorizada na classe C, conforme o artigo 237.

Art. 236 - No serviço ferroviário - considerado este o de transporte em estradas de ferro abertas ao tráfego público, compreendendo a administração, construção, conservação e remoção das vias férreas e seus edifícios, obras-de-arte, material rodante, instalações complementares e acessórias, bem como o serviço de tráfego, de telegrafia, telefonia e funcionamento de todas as instalações ferroviárias - aplicam-se os preceitos especiais constantes desta Seção (BRASIL, 1943).

Art. 237 - O pessoal a que se refere o artigo antecedente fica dividido nas seguintes categorias: a) funcionários de alta administração, chefes e ajudantes de departamentos e seções, engenheiros residentes, chefes de depósitos, inspetores e demais empregados que exercem funções administrativas ou fiscalizadoras; b) pessoal que trabalhe em lugares ou trechos determinados e cujas tarefas requeiram atenção constante; pessoal de escritório, turmas de conservação e construção da via permanente, oficinas e estações principais, inclusive os respectivos telegrafistas; pessoal

de tração, lastro e revistadores; c) das equipagens de trens em geral; d) pessoal cujo serviço é de natureza intermitente ou de pouca intensidade, embora com permanência prolongada nos locais de trabalho; vigias e pessoal das estações do interior, inclusive os respectivos telegrafistas (BRASIL, 1943).

Os artigos 238 e 239 descrevem que a jornada máxima de trabalho do maquinista deve ser de 12 horas, havendo um descanso entre turnos de pelo menos 10 horas. Tempos de traslado antes do início e após o término da viagem não são contabilizados como horários efetivos de trabalho.

Art. 238. Será computado como de trabalho efetivo todo o tempo, em que o empregado estiver à disposição da estrada. § 1º Nos serviços efetuados pelo pessoal da categoria c, não será considerado como de trabalho efetivo o tempo gasto em viagens do local ou para o local de terminação e início dos mesmos serviços

§ 4º Para o pessoal da equipagem de trens, só será considerado esse trabalho efetivo, depois de chegado ao destino, o tempo em que o ferroviário estiver ocupado ou retido à disposição da Estrada. Quando, entre dois períodos de trabalho, não mediar intervalo superior a uma hora, será essa intervalo computado como de trabalho efetivo (BRASIL, 1943).

Art. 239 - Para o pessoal da categoria "c", a prorrogação do trabalho independe de acordo ou contrato coletivo, não podendo, entretanto, exceder de 12 (doze) horas, pelo que as empresas organizarão, sempre que possível, os serviços de equipagens de trens com destacamentos nos trechos das linhas de modo a ser observada a duração normal de oito horas de trabalho.

§ 1º - Para o pessoal sujeito ao regime do presente artigo, depois de cada jornada de trabalho haverá um repouso de 10 (dez) horas contínuas, no mínimo, observando-se, outrossim, o descanso semanal (BRASIL, 1943).

O artigo 241 descreve as quantias recebidas como hora-extra após as oito horas de trabalho efetivo. Desta forma, mesmo com uma jornada limitada a doze horas, a lei mostra que a equipe ferroviária deverá receber adicionalmente a partir da oitava hora trabalhada. Sendo assim, visando à diminuição de custos, as empresas tentarão manter a jornada dos trabalhadores em, no máximo, oito horas.

Art. 241 - As horas excedentes das do horário normal de oito horas serão pagas como serviço extraordinário na seguinte base: as duas primeiras com o acréscimo de 25% (vinte e cinco por cento) sobre o salário-hora normal; as duas subseqüentes com um adicional de 50% (cinquenta por cento) e as restantes com um adicional de 75% (setenta e cinco por cento).
Parágrafo único - Para o pessoal da categoria "c", a primeira hora será majorada de 25% (vinte e cinco por cento), a segunda hora será paga com o acréscimo de 50% (cinquenta por cento) e as duas subseqüentes com o de 60% (sessenta por cento), salvo caso de negligência comprovada (BRASIL, 1943).

No artigo 244 é descrito o serviço de prontidão, ou seja, de aguardo de serviço, e reforça que este trabalho também é remunerado. Assim como as horas extras, durante a alocação de equipe, as horas em espera (de prontidão) deverão ser reduzidas ao máximo possível, visando redução de custos.

Art. 244. As estradas de ferro poderão ter empregados extranumerários, de sobre-aviso e de prontidão, para executarem serviços imprevistos ou para substituições de outros empregados que faltem à escala organizada. § 3º Considera-se de "prontidão" o empregado que ficar nas dependências da estrada, aguardando ordens. A escala de prontidão será, no máximo, de doze horas. As horas de prontidão serão, para todos os efeitos, contadas à razão de 2/3 (dois terços) do salário-hora normal (BRASIL, 1943).

Com essa revisão das leis que regem o trabalho dos maquinistas, em resumo, é possível observar que uma escala otimizada de maquinistas terá como objetivo, além de atender à produção da empresa, também reduzir os gastos associados à jornada dos trabalhadores. O tempo em prontidão e em hora-extra deverá ser o menor possível, sem prejudicar a circulação dos trens, buscando maior economia, segurança e levando à redução de custos e aumento de competitividade.

3.3 REDUÇÃO DE CICLOS OCIOSOS NA MRS LOGÍSTICA S.A.

Na MRS Logística S.A., como descrito por Fuzeto, Silveira e Castro (2017) [22] e Trinchinato (2017) [68], há planejamentos de longo e curto prazo. Porém a alocação do maquinista para um trem específico somente é feita na programação diária, de ultra-curto prazo [1]. Esta alocação é feita em média a cada quatro horas, com visão das próximas doze horas. Assim as demandas mais próximas são atualizadas e as mais distantes entram ao final da programação. A alocação é feita neste ultra-curto prazo devido aos cenários serem dinâmicos, podendo mudar por causa de ocorrências no trecho ou mesmo mudanças de pedido de clientes. Com isso, as previsões para mais de doze horas perdem o sentido, dado que o cenário real será completamente diferente, pois pode haver muitas mudanças na circulação de trens em curto prazo. Vale ressaltar que os trens de minério não possuem grade horária fixa e circulam em regime direto de carga e descarga, dependendo somente de demanda dos clientes.

Esta pesquisa tem como foco a implementação de inteligência computacional na alocação de equipagem ferroviária com a visão de ultra-curto prazo. A realização desta alocação é realizada pelo cargo do programador de equipagens, sendo os dados de entrada: previsão de trens nas próximas 12 horas e maquinistas que se apresentam nesta faixa de horário, bem como suas habilitações e tipos de escala. Mais especificamente, a aplicação será realizada na alocação dos maquinistas em regime de prontidão no carrossel do minério, no pátio ferroviário de Dias Tavares em Juiz de Fora - MG, Figura 4, e no pátio ferroviário

P1-07 (São Brás do Suaçui - MG). Dias Tavares e P1-07 também são chamados pelas suas códigos: FDT e FJC, respectivamente.



Figura 4 - Trens Durante a Troca de Equipagem em Dias Tavares.
Fonte: O autor (2023).

Para o regime de prontidão, conforme Rodrigues (2015) [53], o maquinista de Juiz de Fora só tem um tipo de serviço possível: equipar um trem em Juiz de Fora, conduzindo-o até Conselheiro Lafaiete, seguindo para o descanso em hotel com duração de 10 horas, regulamentado por lei. Após este tempo, irá de carro para o P1-07, onde conduzirá um trem até a cidade de Bom Jardim de Minas, retornando de carro para seu destacamento em Juiz de Fora, saindo então para a folga. Neste trecho, não há restrições de habilitações, ou seja, todos os maquinistas podem equipar todos os trens. Considerando tempos improdutivos, em transporte, em trem e em descanso, pode se dizer que aproximadamente o maquinista estará disponível no P1-07 no dia seguinte a sua apresentação em Juiz de Fora. O ciclo do maquinista de Juiz de Fora é ilustrado na Figura 5.

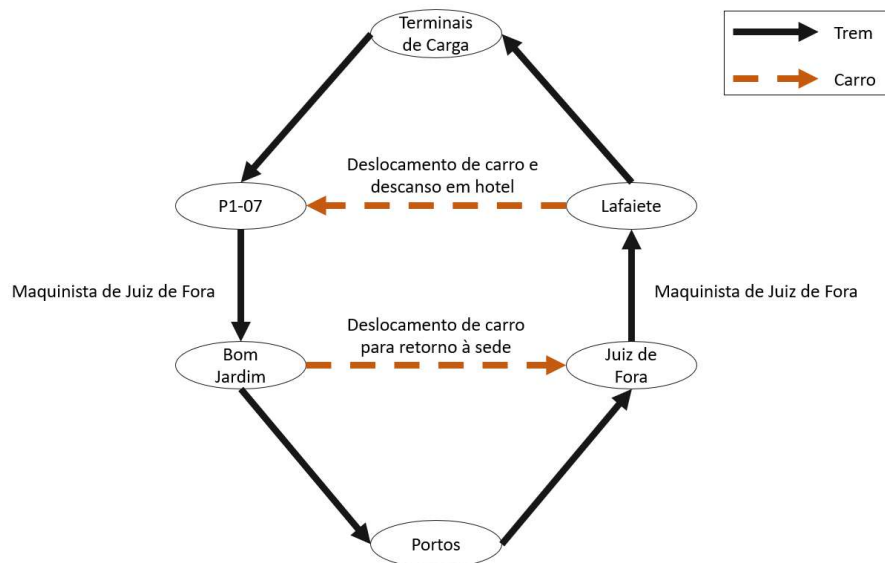


Figura 5 - Ciclo do Maquinista de FDT.
Fonte: O autor (2023).

Os outros trechos da Figura 5 acontecem de forma semelhante. Um maquinista equipa em Lafaiete, conduz até o terminal de carga, entra para o descanso e fica disponível no dia seguinte para retornar. Semelhantemente é feito em Bom Jardim para os portos. Porém, devido a cada trecho ter sua especificidade, este estudo focará somente no ciclo dos maquinistas de Juiz de Fora.

As alocações são realizadas seguindo a ordem FIFO (*First-in First-out*). A regra geral para alocação de um maquinista em um serviço são:

- Um maquinista só deve ser alocado a uma tarefa que ele tenha condições de cumprir dentro da jornada de trabalho regulamentada por lei.
- O atraso de trem por falta de maquinista deve ser evitado.
- O tempo de improdutivo (tempo ocioso) do maquinista deve ser minimizado, desde que não gere grandes impactos na circulação de trens, pois, neste tempo, o colaborador não pode realizar nenhuma atividade, sendo um tempo remunerado somente para que ele espere o trem.

Em um cenário com cinco maquinistas disponíveis e cinco trens previstos, a alocação de equipagem se daria conforme a Tabela 1: um maquinista por trem, sem gerar atrasos.

Tabela 1 – Exemplo de uma Programação de Equipe Um pra Um.

Maquinista	Horário do Maquinista	Trem	Saída do Trem
Maquinista 1	8:00	Trem A	9:00
Maquinista 2	10:00	Trem B	12:00
Maquinista 3	14:00	Trem C	15:00
Maquinista 4	17:00	Trem D	18:00
Maquinista 5	19:00	Trem E	19:30

Fonte: O autor (2023).

O cenário da Tabela 1 é o ideal, todos os trens atendidos, sem gerar atrasos. Já na Tabela 2, apesar de todos os trens estarem atendidos, há um atraso previsto no trem D, devido ao seu maquinista só chegar após 30 minutos. Como todos os outros maquinistas estão alocados em trem, não há nenhum outro disponível para que o atraso seja evitado. Desta forma, uma opção seria o programador de equipagens tentar negociar a antecipação do maquinista 4 em meia hora. Caso não consiga o aceite, o trem irá aguardar 30 minutos até o maquinista se apresentar.

Um cenário ainda pior é o da Tabela 3. Neste há um trem a mais do que maquinistas disponíveis. Este é um cenário indesejado, em que haverá necessidade de trazer algum maquinista de outro destacamento ou mesmo de antecipar em muitas horas alguma apresentação do dia seguinte para que o trem não fique parado. Para trazer o recurso de outro destacamento, deve haver sobra de maquinistas neste outro local e este deve

ser habilitado no trecho em que será levado. Em relação à antecipação, além de aceite do próprio maquinista, ele não poderá ser antecipado se isso acarretar em um tempo de descanso menor que o regulamentado.

Tabela 2 – Exemplo de uma Programação com Atraso no Trem.

Maquinista	Horário do Maquinista	Trem	Saída do Trem
Maquinista 1	8:00	Trem A	9:00
Maquinista 2	10:00	Trem B	12:00
Maquinista 3	14:00	Trem C	15:00
Maquinista 4	18:30	Trem D	18:00
Maquinista 5	19:00	Trem E	19:30

Fonte: O autor (2023).

Tabela 3 – Exemplo de uma Programação de Equipe com Falta.

Maquinista	Horário do Maquinista	Trem	Saída do Trem
Maquinista 1	8:00	Trem A	9:00
Maquinista 2	10:00	Trem B	12:00
Maquinista 3	16:00	Trem C	18:00
Maquinista 4	17:00	Trem D	19:30
—	—	Trem E	20:00

Fonte: O autor (2023).

Ainda, pode haver um cenário em que haja sobra de maquinistas. Em um FIFO simples teríamos a situação da Tabela 4. Neste cenário de FIFO, é possível observar que os maquinistas estão aguardando muito tempo os trens e o último colaborador está sobrando. Considerando a redução do horário de espera dos maquinistas e sem gerar THP (Trem Hora Parado), a alocação ideal para o cenário de sobra mostrado pode ser vista na Tabela 5.

Tabela 4 – Exemplo de uma Programação de Equipe com Sobra.

Maquinista	Horário do Maquinista	Trem	Saída do Trem
Maquinista 1	8:00	Trem A	9:00
Maquinista 2	10:00	Trem B	12:00
Maquinista 3	16:00	Trem C	18:00
Maquinista 4	17:00	Trem D	19:30
Maquinista 5	19:00	—	—

Fonte: O autor (2023).

Desta forma, todos os trens foram atendidos e cada maquinista está esperando por menos tempo o trem, gerando uma escala mais econômica (redução de pagamento de horas improdutivas). O maquinista que está sobrando pode ser liberado de seu serviço ou pode ir “sobre-alocado” no próximo trem. No jargão ferroviário, “sobre-alocar” é também chamado de “duplar” e significa colocar dois maquinistas em um mesmo trem. A vantagem da liberação é a economia direta de não haver o gasto com horas improdutivas, traslado de

carro e hotel para este maquinista, porém, como a escala funciona como um ciclo, se este maquinista não sobe de trem de Juiz de Fora (FDT) para Lafaiete, ele conseqüentemente não se apresentará no P1-07 (FJC) para equipar trens. Se um maquinista a menos fizer o ciclo, pode ser que algum trem fique sem quem o conduza em FJC após a hora estimada de sua chegada. Por isso, mesmo sem necessidade no primeiro trem, geralmente o maquinista não é dispensado, subindo duplado para estar disponível em FJC no dia seguinte. Porém, pode ocorrer que este maquinista suba duplado na ida e também não seja necessário na volta, retornando também duplado, ou seja, realizando um ciclo inteiro completamente ocioso.

Tabela 5 – Exemplo de uma Programação de Equipe com Sobra e Otimizada.

Maquinista	Horário do Maquinista	Trem	Saída do Trem
Maquinista 1	8:00	Trem A	9:00
Maquinista 2	10:00	Trem B	12:00
Maquinista 3	16:00	—	—
Maquinista 4	17:00	Trem C	18:00
Maquinista 5	19:00	Trem D	19:30

Fonte: O autor (2023).

Considerando que o maquinista só fica disponível em FJC no dia seguinte à sua apresentação em FDT, fica claro que quando este maquinista é alocado para o primeiro trem em Juiz de Fora, dado que o horizonte de programação é de somente doze horas, não é possível saber neste primeiro momento qual será sua real necessidade em FJC. Não se sabe qual trem ou mesmo se haverá trem para que este faça o percurso de retorno do P1-07 para Bom Jardim de Minas. Desta forma, adotou-se como uma regra na programação de equipagem que a sobra deveria subir sempre duplada, pois, pela falta desta informação, era melhor haver sobra de maquinista em FJC, do que falta. Porém, caso houvesse uma forma de identificar quando duplar ou quando dispensar o maquinista, isso geraria não só economia imediata para a empresa, mas também um ciclo menor do maquinista, pois este estaria disponível novamente na sede origem mais rapidamente, dado que não realizou o ciclo longo de viagem. Com isso, além de não gastar recursos sem necessidade, a apresentação deste maquinista em sede antecipadamente ainda poderia evitar uma possível falta de equipes.

Todo este processo descrito para a alocação e diminuição de horas improdutivas é feito pelo programador de equipagens de forma manual. Um modelo para otimização e automatização do processo de alocação de equipes no trecho de Juiz de Fora a Lafaiete foi desenvolvido por Rodrigues (2015) [53], porém ainda não implementado na empresa. Ainda, este modelo, em cenários de sobra de maquinistas, somente mostra que o maquinista não precisa ser utilizado em trem na ida, não dizendo sobre sua necessidade no trecho de retorno, ou seja, ele não diz o que fazer com a sobra, se deve ser dispensada ou duplada.

A chegada de maquinistas em FJC está diretamente ligada à circulação de trens pela Linha do Centro (trecho de Dias Tavares, Juiz de Fora - MG). Para ilustrar esta premissa, será utilizada a visão do gráfico horário de trens, em que no eixo y estão os pátios percorridos pelo trem e no eixo x está o tempo. Na Figura 6, há um espaçamento de chegada de trens entre cinco e dez horas da manhã em Juiz de Fora, o que significa que ficarão cinco horas sem que maquinistas equipem trens para Lafaiete, causando no dia seguinte um espaçamento de cinco horas sem haver apresentações de maquinistas em FJC. Com isso, havendo a disponibilidade de maquinistas, seria importante colocar mais de um maquinista no trem que antecede o espaçamento, pois quando estes maquinistas em excedente chegarem em FJC, supririam o “buraco” do gráfico.



Figura 6 - Cenário com um Grande Espaçamento de Trens.
Fonte: O autor (2023).

Já na Figura 7, não há um grande espaçamento de trens, porém o espaçamento médio entre eles é maior que na Figura 6. Neste cenário, não está claro se duplicar algum trem traria benefício ou somente custos, ou seja, não é possível saber se há chances de faltar maquinistas em FJC no futuro, se nenhum destes trens for duplicado.

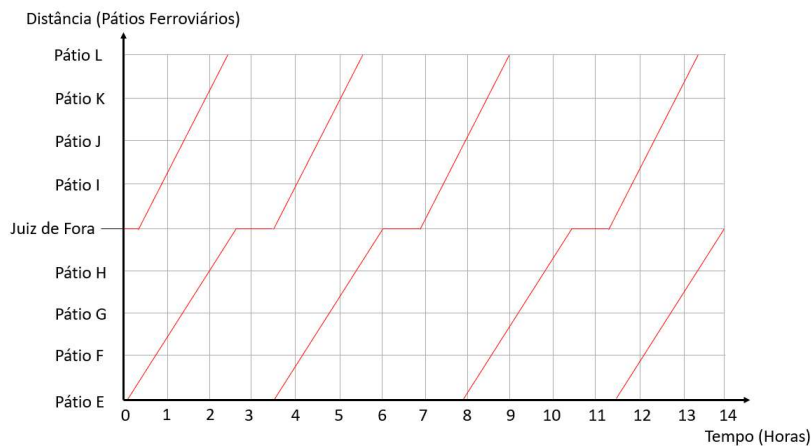


Figura 7 - Cenário com Poucos Trens.
Fonte: O autor (2023).

Por último, na Figura 8, há uma sequência de muitos trens em circulação em intervalos muito pequenos de tempo. Neste cenário, dificilmente haveria necessidade de se colocar um maquinista a mais em trem, já que futuramente FJC estaria recebendo muitos maquinistas em um curto período de tempo. Com isso, se houver sobra, a chance deste maquinista realizar um ciclo ocioso é grande, sendo o ideal dispensá-lo.

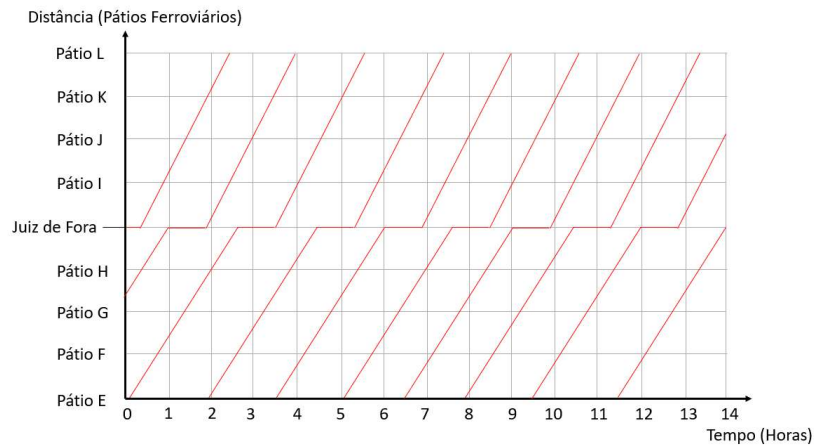


Figura 8 - Cenário com Muitos Trens.

Fonte: O autor (2023).

Enquanto o cenário de FDT pode sofrer grandes variabilidades, conforme as figuras apresentadas, o de FJC tende a ser mais próximo da Figura 8, pois recebe trens de diversas localidades, e uma ocorrência em uma linha não impede que cheguem trens vindo de outros trechos, ou seja, em FJC há um fluxo mais constante de trens. Como observado, há diversos cenários de circulação em FDT e não é claro quando duplar um trem é realmente necessário.

Pode-se definir o problema a ser estudado como: busca da redução de ciclos ociosos de maquinistas, no trecho compreendido por Juiz de Fora e P1-07 da MRS Logística S.A., através da previsão da real necessidade de sobre-alocar um trem, após verificada a sobra de tripulação na programação de equipagens de ultra-curto prazo.

4 REFERENCIAL TEÓRICO

Neste capítulo, são descritos os conceitos de inteligência computacional e análise de dados utilizados no desenvolvimento do trabalho.

Segundo Goodfellow, Bengio e Courville (2016), aprendizado de máquina é a capacidade de sistemas de adquirirem conhecimento extraindo padrões de dados [27]. Já Géron (2019) define o aprendizado de máquina como a ciência e a arte da programação de computadores para que eles possam aprender com os dados [25]. Ainda, Mitchell (1997) afirma que “Diz-se que um programa computacional aprende pela experiência E em relação a algum tipo de tarefa T e a alguma medida de performance P, se sua performance nas tarefas T, medidas por P, melhoram com a experiência E” [41].

No aprendizado de máquina supervisionado, os dados fornecidos ao algoritmo possuem as soluções desejadas, chamadas de rótulos [25]. Desta forma, o modelo aprende ao comparar seus resultados com os reais. A classificação é um tipo de aprendizado de máquina supervisionado em que os dados de saída são discretos e sem ordem [36], ou seja, dados categóricos.

4.1 ANÁLISE EXPLORATÓRIA DE DADOS

Conforme Komorowski *et al.* (2016) [35], um dos objetivos da análise exploratória de dados é auxiliar o analista no reconhecimento de padrões dos dados. Nesta etapa do processo, é comum a construção de gráficos de diversos tipos, buscando entender como as variáveis se relacionam entre si e como se relacionam com a variável alvo. Além do padrão visual, outra forma de achar relações é através do cálculo da covariância e correlação entre as variáveis.

A covariância [35] é a medida de interdependência linear entre duas variáveis e é calculada conforme a Equação 4.1.

$$Cov(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}, \quad (4.1)$$

em que x e y são as variáveis, n o número de dados da amostra, \bar{x} a média da variável x e \bar{y} a média da variável y .

Se a covariância possuir valor alto e positivo, significa que quando uma variável aumenta, a outra também aumentará. Já um valor alto e negativo significa que quando uma variável aumenta, a outra diminuirá. Por último, valores próximos de zero indicam que as variáveis não possuem relação linear entre si.

A correlação [35] é calculada pela Equação 4.2 e é similar à covariância, porém de maior interpretabilidade dos resultados, já que limita os possíveis valores entre -1 e +1.

$$\text{Corr}(x, y) = \frac{\text{Cov}(x, y)}{S_x S_y}, \quad (4.2)$$

em que S_x e S_y são os desvios padrões das variáveis x e y , respectivamente.

4.2 PREPARAÇÃO DOS DADOS

Para que os modelos de aprendizado entendam as variáveis categóricas, é necessário codificá-las em números. O chamado *Ordinal Encoder* [50] atribui um número a cada categoria. Esse tipo de codificação não cria novas colunas nos dados, porém acaba criando uma relação sequencial entre as variáveis, que pode não existir. Outra forma de codificar é através do *One Hot Encoder* [50], que cria uma coluna para cada classe existente da variável categórica e esta será assinalada como 1 se a amostra pertencer àquela classe ou 0, caso contrário. Outras formas de codificação são descritas por Potdar, Pardawala e Pai (2017) [50].

Alguns modelos de aprendizado de máquina não alcançam bons resultados de previsão se suas variáveis tiverem ordem de grandeza muito diferentes, como o SVM e a rede neural. Para resolver este problema, é usual escalar as variáveis para dentro de uma mesma faixa de valores. Uma das formas de fazer isso é o *MinMax Scaling* [69], em que cada variável é transformada de forma a só ter valores entre 0 e 1. A Equação 4.3 descreve este método.

$$x'_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}, \quad (4.3)$$

em que x_i é uma amostra da variável x , x'_i é o valor convertido de uma amostra da variável x , $\min(x)$ e $\max(x)$ são funções que retornam o mínimo e máximo valor, respectivamente, encontrado para a variável x .

4.3 AJUSTE DE HIPERPARÂMETROS, REGULARIZAÇÃO E VALIDAÇÃO CRUZADA

Os hiperparâmetros são parâmetros dos modelos que são definidos antes do momento do aprendizado, como, por exemplo, a arquitetura da rede neural, o tamanho da árvore de decisão, quantas árvores estarão compondo o modelo de floresta aleatória ou o valor do parâmetro de regularização. Não há um valor ideal de hiperparâmetro que desempenhe bem em todas as situações. Desta forma, o ideal é usar um método que encontre os hiperparâmetros que funcionem melhor para os dados a serem utilizados. Um destes métodos é o *Grid Search* [73], que consiste em testar todas as combinações possíveis dos hiperparâmetros escolhidos pelo usuário. O fato de este método testar todas as combinações faz com que esta etapa possa se tornar lenta, porém ainda é muito utilizado, como visto em [8, 32, 51]. Outras alternativas para ajuste de hiperparâmetros são métodos como o *Random Search* e otimização Bayesiana [73].

Para medir o poder de generalização do modelo, é necessário dividir os dados em um grupo de treinamento e outro de teste, de forma que o algoritmo seja treinado no primeiro grupo e testado no outro, com dados não usados no treinamento. Outra forma é usar a validação cruzada *K-fold*, em que os dados são divididos em K partições, o treino é feito em $K-1$ grupos e testado no restante. Após computar o resultado, muda-se o grupo de retenção e repete-se o processo até terem sido usadas todas as K combinações. Ainda, há a versão da validação cruzada *K-fold* estratificada, em que a divisão dos K grupos é feita de forma a se manter em cada um a mesma proporção de classes presente na base completa [25].

O modelo de aprendizado de máquina pode ter um alto poder preditivo nos dados de treinamento e não conseguir um resultado satisfatório em dados novos. Este fenômeno é conhecido como sobreajuste dos dados. Uma forma de evitar o sobreajuste é aplicando a regularização, ou seja, adicionando termos na forma de hiperparâmetros, que penalizam o modelo se ele se tornar muito complexo. O oposto do sobreajuste é o subajuste, em que o modelo não desempenha bem nem no treinamento nem no teste e as formas mais comuns de correção deste efeito é aumentando a complexidade do modelo ou adicionando mais dados [25].

4.4 MODELOS DE APRENDIZADO DE MÁQUINA

4.4.1 Árvore de Decisão

A árvore de decisão é um modelo de aprendizado que consiste na subdivisão do espaço de instâncias de acordo com o valor de uma variável que represente a melhor redução na medida de impureza [54]. Uma das grandes vantagens no uso da árvore de decisão é que este não é um algoritmo “caixa-preta”, ou seja, é de fácil rastreabilidade e possível de entender os caminhos que levaram o algoritmo a classificar uma instância como pertencendo a determinada classe. A árvore é formada por nós e folhas. Os nós são as subdivisões internas da árvore, representando uma tomada de decisão a partir de um valor de um atributo categórico ou de uma faixa de valores para atributos numéricos. As folhas são os nós externos da árvore e a cada folha deve ser atribuída uma classe, de acordo com a classe majoritária de suas instâncias.

O algoritmo de árvore de decisão divide o espaço de instâncias como o resultado de uma função dos atributos de entrada. Esta função é selecionada de acordo com o critério de divisão. Após a seleção da divisão apropriada, cada nó será subdividido até que não haja ganho significativo no critério de divisão. A subdivisão pode ser encerrada após o algoritmo chegar a algum ponto de parada pré-definido como, por exemplo, número máximo de níveis na árvore, número mínimo de instâncias em um nó ou número máximo de nós. Esses critérios de parada ajudam o modelo a não se tornar muito complexo, de difícil entendimento, e também evita o sobreajuste dos dados. Alguns modelos como o

CART [15] realizam a pós-poda, ou seja, após a construção de uma árvore maior, há uma etapa de poda de galhos com o objetivo de diminuir a complexidade sem perder significativamente na classificação.

Existem diversos métodos para a repartição de uma árvore, como descrito por Rokach e Maimon (2005) [54], porém os mais comuns são Entropia (Equação 4.4) e Índice de Gini (Equação 4.5).

Entropia:

$$Entropia = -\sum_{i=1}^n P_i \log_2(P_i). \quad (4.4)$$

Índice Gini:

$$Gini = 1 - \sum_{i=1}^n P_i^2, \quad (4.5)$$

em que P_i é a probabilidade de uma amostra ser da classe i no nó avaliado.

O procedimento para construção de árvores de decisão pode ser resumido nos seguintes passos, conforme Loh (2011) [39]:

- Seja X o conjunto de variáveis dos dados, e S o subconjunto de valores assumidos por X ;
- Iniciando pelo nó raiz, para cada variável em X , busque um subgrupo S que minimize a soma da medida de impureza adotada;
- Escolha a repartição que obteve o valor mínimo de impureza e divida o nó raiz em dois nós;
- Se algum critério de parada tiver sido encontrado, pare, senão repita o método de repartição até que um critério de parada seja satisfeito.

4.4.2 Floresta Aleatória

De acordo com Breiman (2001) [16], a floresta aleatória é um classificador que consiste em uma coleção de k classificadores de árvores de decisão $h(x, \theta_k)$, em que θ_k são vetores aleatórios que controlam os parâmetros das árvores, e cada árvore contribui para a floresta através de um voto para a classe predita através do input x .

Com o objetivo de construir a floresta com modelos diversos, cada árvore é treinada em um subconjunto aleatório dos dados de treinamento. Quando os dados são amostrados com substituição é chamado de *bagging* (abreviação de *bootstrap aggregating*); já quando não há reamostragem para a mesma árvore é chamado de *pasting*. Por inserir uma diversidade maior nos modelos, ao usar o *bagging* os preditores criados ficam sendo menos correlacionados, gerando um agrupamento menos sujeito ao sobreajuste. Outra aleatoriedade inserida nas árvores é a escolha de quais variáveis serão usadas. Com isso, são gerados diversos modelos mais fracos (treinados em subconjuntos de dados e variáveis),

porém bastante diversos e que, quando combinados, suas predições através de votação geram um agrupamento poderoso [25].

4.4.3 Máquina de Vetor de Suporte

A máquina de vetor de suporte, ou SVM (da sigla em inglês), é um modelo de aprendizado de máquina que busca o hiperplano que aumenta a margem de separação entre as classes [9]. Na Figura 9 é possível observar que apesar de ser possível criar uma infinidade de hiperplanos possíveis para separar os dados verdes dos azuis, somente o hiperplano H está posicionado na distância máxima possível (d) dos pontos das diferentes classes, além de classificá-los corretamente.

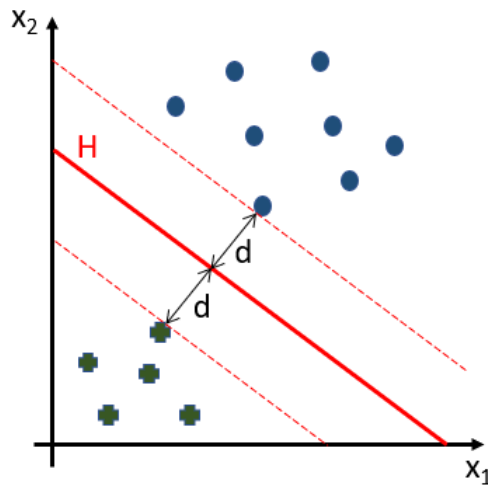


Figura 9 - Exemplo de Hiperplano Criado pelo SVM com Dados de Duas Variáveis.
Fonte: O autor (2023).

Conforme Kotsiantis (2007) [36], uma intuição simplificada do SVM mostra que este é um problema de otimização, em que a busca por minimizar a norma do vetor perpendicular ao hiperplano, atendendo às restrições necessárias, resultará no hiperplano com máxima margem entre as classes. A Equação 4.6, descreve o hiperplano:

$$w^T x_i + b = 0, \quad (4.6)$$

em que w é o vetor de características do hiperplano, x_i os pontos pertencentes a ele e b o coeficiente linear.

Assumindo que os pontos positivos irão assumir valor $+1$ e os negativos -1 , tem-se a Equação 4.7:

$$\begin{cases} w^T x_i + b \geq 1, \text{ para todo } x_i \in P \\ w^T x_i + b \leq -1, \text{ para todo } x_i \in N, \end{cases} \quad (4.7)$$

sendo P os valores do lado positivo do hiperplano e N , os valores do lado negativo.

Os dados serão classificados conforme a Equação 4.8:

$$f_{w,b}(x) = \text{sgn}(w^T x + b), \quad (4.8)$$

em que sgn é a função sinal.

Para encontrar o vetor w que representa o hiperplano de máxima margem, o modelo de máquina de vetor de suporte torna-se um problema de otimização, cujo objetivo é minimizar a norma do vetor w , conforme a Equação 4.9:

$$\begin{cases} \text{Minimizar}_{w,b} \Phi(w) = \frac{1}{2} \|w\|^2 \\ \text{sujeito a } y_i(w^T x_i + b) \geq 1, i = 1, 2, \dots, l, \end{cases} \quad (4.9)$$

sendo y_i uma variável que vale +1 para pontos positivos e -1 para pontos negativos.

Uma forma de resolver o problema de otimização é introduzir multiplicadores Lagrangianos positivos para cada restrição da desigualdade, conforme a Equação 4.10:

$$L_P \equiv \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i y_i (x_i \cdot w - b) + \sum_{i=1}^N \alpha_i, \quad (4.10)$$

em que N é o número total de pontos (dados de entrada) e α_i são os multiplicadores Lagrangianos positivos.

Minimizar L_P em relação a w torna-se um problema de otimização convexo, ou seja, sem mínimos locais. Na solução, os pontos em que $\alpha > 0$ são chamados de vetores de suporte. O problema de otimização acima gera o chamado classificador *hard margin*, pois não permite que haja pontos dentro das margens do hiperplano, tornando este tipo de modelo de difícil aplicação, pois as classes obrigatoriamente devem ser linearmente separáveis, algo difícil de encontrar em dados reais. Para contornar esse problema, é criada uma solução *soft margin*, que inclui uma restrição ξ que é limitada no modelo a um valor para permitir uma taxa de classificação errada, podendo ser esta dentro da margem ou até mesmo do outro lado do plano. Com isso, o problema toma a forma da Equação 4.11:

$$L_P \equiv \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i [y_i (x_i \cdot w - b) - 1 + \xi_i] - \sum_{i=1}^N \mu_i \xi_i, \quad (4.11)$$

em que ξ é uma variável que computa o erro do modelo, μ são multiplicadores Lagrangianos inseridos para forçar a ξ ser positivo. C é um hiperparâmetro que controla o quanto o modelo é flexível a erros. Quanto maior C , o modelo penalizará mais o erro e tentará realizar a separação perfeita entre as classes; já um valor menor de C faz o modelo aceitar dados classificados erroneamente, evitando o sobreajuste.

Para que a máquina de vetor de suporte possa conseguir classificar corretamente dados que não são linearmente separáveis, faz-se o uso de funções kernel. Essas funções mapeiam os dados em um novo espaço dimensional de maior ordem, onde estes podem

ser linearmente separados. As funções kernel mais utilizadas são: linear (Equação 4.12), polinomial (Equação 4.13), RBF gaussiano (Equação 4.14) e sigmoide (Equação 4.15) [25]:

$$K(a, b) = a^T b, \quad (4.12)$$

$$K(a, b) = (\gamma a^T b + r)^d, \quad (4.13)$$

$$K(a, b) = e^{-\gamma \|a-b\|^2}, \quad (4.14)$$

$$K(a, b) = \tanh(\gamma a^T b + r), \quad (4.15)$$

em que a e b são vetores, d é o grau do kernel polinomial, γ é um parâmetro que regula a distância no qual os dados podem influenciar o hiperplano e r é um coeficiente independente. Demonstrações dos fundamentos teóricos do SVM são encontrados em [9, 43].

4.4.4 Rede Neural

O estudo das redes neurais surgiu inspirado no funcionamento dos neurônios, que quando recebem um número suficiente de neurotransmissores num espaço de milissegundos, realizam a transmissão de impulsos elétricos para o próximo neurônio [25]. O primeiro modelo que representou um neurônio de forma matemática foi proposto por McCulloch e Pitts em 1943 [40]. Neste modelo, tantos os *inputs* quanto os *outputs* eram binários e a ativação ocorria se houvesse a recepção de pelo menos dois estímulos. Em 1957, Rosenblatt [56] apresentou o Perceptron (Figura 10), que era modelado como uma soma ponderada dos *inputs* e pesos, e ao resultado da soma era aplicado uma função degrau.

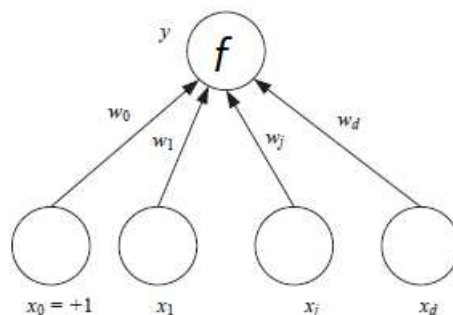


Figura 10 - Perceptron.

Fonte: Adaptado de (Alpaydin, 2014) [3].

O cálculo do Perceptron é feita conforme a Equação 4.16 [18].

$$y = f_h(\sum_{j=0}^P w_j x_j) = f_h(w^T x), \quad (4.16)$$

em que w é o vetor de pesos, x é o vetor com os dados de entrada, P é o número de variáveis em uma amostra, e f_h é a função de ativação.

Conforme Alpaydim (2014) [3], as redes neurais são arquiteturas baseadas em conexões de Perceptrons, em que a saída de um neurônio serve de *input* para a próxima camada (Figura 11). De forma a se tornar derivável, a função degrau foi substituída nos futuros modelos pela função sigmoide (Equação 4.17).

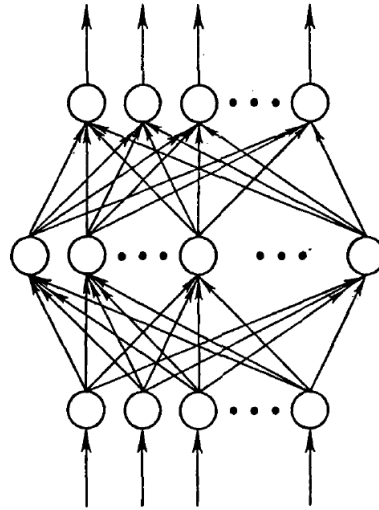


Figura 11 - Exemplo de Rede Neural.
Fonte: Adaptado de (Rumelhart *et al.*, 1986) [57].

$$f(x) = \frac{1}{1 + (e^{-x})}. \quad (4.17)$$

Além da função sigmoide, outras funções de ativação comuns são a tangente hiperbólica (Equação 4.18) e ReLu (Equação 4.19). Esta última muito utilizada em redes com muitas camadas, para evitar o fenômeno do desaparecimento de gradiente. O formato assumido por estas funções se encontra na Figura 12.

$$f(x) = \tanh(x), \quad (4.18)$$

$$\begin{cases} f(x) = x, x \geq 0 \\ f(x) = 0, x < 0. \end{cases} \quad (4.19)$$

Para problemas de classificação binária, a função de erro utilizada para a rede neural é a cross-entropia binária (Equação 4.20).

$$E = -\frac{1}{m} \sum_{i=1}^m [y^i \log(p^i) + (1 - y^i) \log(1 - p^i)], \quad (4.20)$$

em que m é o número de amostras, y é a classe real do dado e p , a classe prevista.

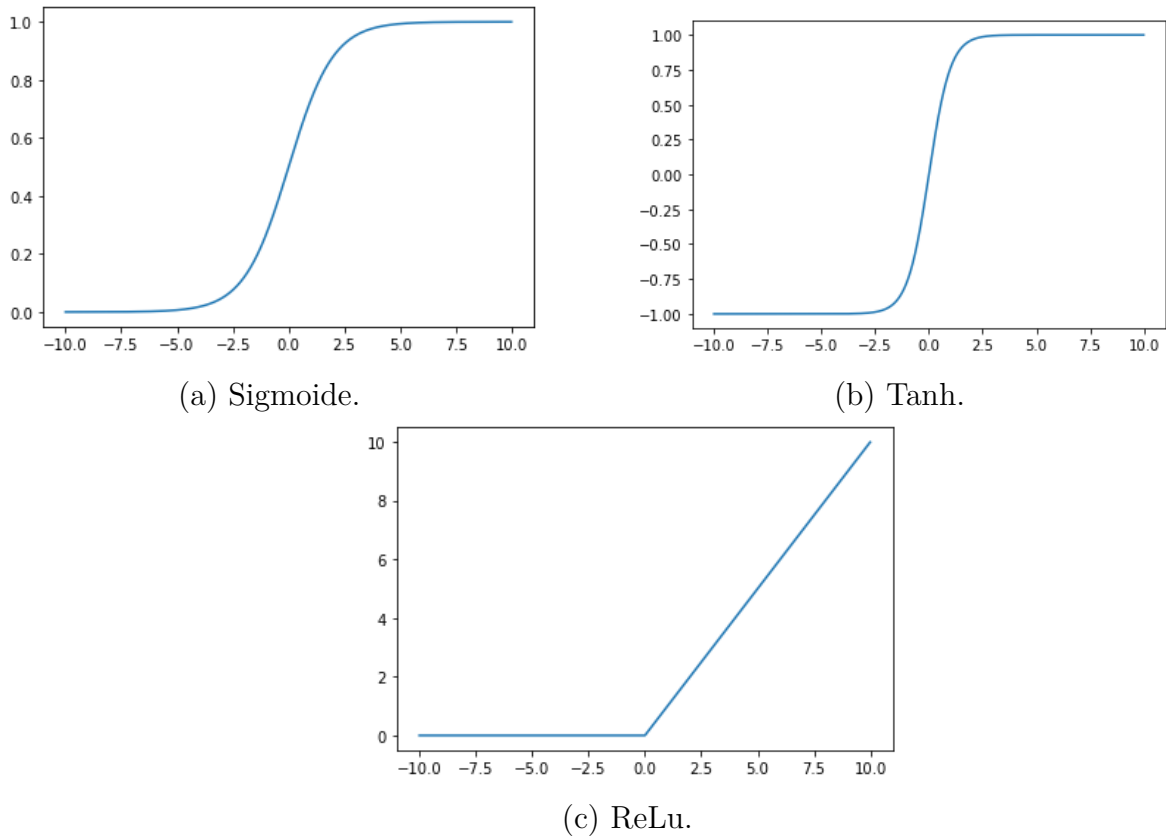


Figura 12 - Funções de Ativação.
Fonte: O autor (2023).

O aprendizado da rede neural dá-se através do *backpropagation* [57], em que os valores dos parâmetros de pesos são atualizados em direção ao negativo da derivada parcial da função de custo em relação ao vetor de parâmetros (Equação 4.21). Para calcular o gradiente da rede, é usada a regra da cadeia a partir da função de erro, passando por todas as camadas, do fim ao início. Ao final, é esperado que a rede convirja para o valor que minimize a função de erro.

$$\Delta w_{h,j} = -\eta \frac{\partial E}{\partial w_{h,j}}, \quad (4.21)$$

em que η é a taxa de aprendizado, E é a função de erro e w o vetor de pesos entre as camadas h e j .

4.4.5 ALMMo

O *Autonomous Learning Multi Model System*, ou ALMMo [5], pode ser entendido como um sistema neuro-*fuzzy* ou como um agrupamento de sistemas baseados em regras *fuzzy* do tipo *AnYa* [6] (Equação 4.22):

$$R_j : \text{SE } \underbrace{x \sim \Xi_j}_{\text{Antecedente}} \text{ ENTÃO } \underbrace{y_j \sim u^T a_j}_{\text{Consequente}}, \quad (4.22)$$

em que x é a amostra, \sim denota similaridade, Ξ_i é a nuvem de dados, y_i o output, a_j os parâmetros do conseqüente e $u = [1, x^T]$

Cada regra *fuzzy* criada pelo ALMMo apontará para uma nuvem de dados, podendo ser representada pelo seu protótipo (ponto de maior densidade discreta), conforme Figura 13. As nuvens de dados são agrupamentos de dados que formam mosaicos de Voronoi [46], não possuindo formas pré-definidas como hiper-retângulos, círculos ou elipsóides.

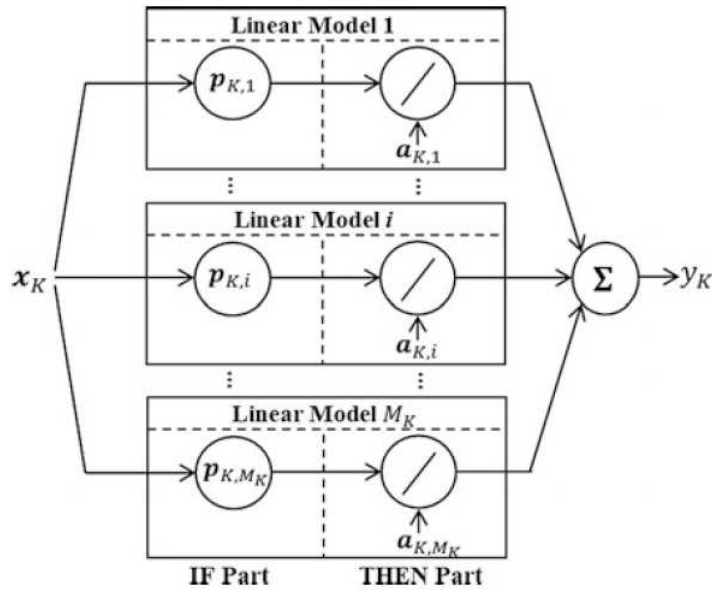


Figura 13 - Regras *Fuzzy* do ALMMo.

Fonte: (Angelov e Gu, 2019) [4].

Após ter seus parâmetros inicializados, o ALMMo computará K . K é uma medida de “tempo” em que será acrescentado em uma unidade a cada dado novo x_K que entrar no sistema. Então, será necessário calcular a média do sistema (Equação 4.23) e o produto escalar médio do sistema (Equação 4.24). Para calcular a próxima medida, é necessário calcular antes a densidade unimodal discreta (Equação 4.25). Esta é uma medida da distribuição espacial no ponto de dado em relação a todos os outros pontos de dados. Após a entrada de um dado novo, a densidade unimodal normalizada deste ponto (Equação 4.26) é calculada em relação a cada nuvem de dados. Pode-se dizer que esta medida é o equivalente ao grau de pertencimento *fuzzy* deste ponto de dado em relação às nuvens formadas.

Média:

$$\mu_K = \frac{K-1}{K} \mu_{K-1} + \frac{1}{K} x_k. \quad (4.23)$$

Produto escalar médio:

$$X_K = \frac{K-1}{K} X_{K-1} + \frac{1}{K} \|x_k\|^2. \quad (4.24)$$

Densidade unimodal discreta:

$$D_K(x) = \frac{1}{1 + \frac{\|x_k - \mu_K\|^2}{S_k - \|\mu_k\|^2}}. \quad (4.25)$$

Densidade unimodal normalizada:

$$\lambda_{l,i} = \frac{D_{l,i}(x_l)}{\sum_{j=1}^{N_j} D_{l,j}(x_l)}. \quad (4.26)$$

Com $\lambda_{K+2,j}$, é possível calcular o output do sistema (Equação 4.27):

$$y_{K+2} = \sum_{j=1}^{N_{K+1}} \lambda_{K+2,j} u_{K+2}^T a_{K+1,j}. \quad (4.27)$$

Condição 1: se a densidade discreta unimodal em $K+1$ do novo dado (Equação 4.28) for maior que todas as densidades discretas unimodais em $K+1$ dos pontos focais, ou menor que a menor densidade discreta unimodal em $K+1$ dos pontos focais, uma nova regra é criada em torno de x_k . Caso esta condição não seja atingida, este dado é atribuído à nuvem mais próxima e os parâmetros da nuvem são atualizados.

$$D_{K+1,i}(x_{K+1}) = \frac{1}{1 + \frac{S_{K,i}^2 \|x_{K+1} - \mu_{K,i}\|^2}{(S_{K,i} + 1)(S_{K,i} X_{K,i} + \|x_{K+1}\|^2) - \|x_{K+1} + S_{K,i} \mu_{K,i}\|^2}}, \quad (4.28)$$

em que $S_{k,i}$ é o número de amostras de dados associados com a nuvem i .

Se a condição 1 for aceita, é necessário verificar a condição 2 (Inequação 4.29): se a nova nuvem de dados se sobrepõe à anterior.

$$D_{K+1,i}(x_{K+1}) \geq \frac{1}{1 + n^2}, \quad (4.29)$$

em que n é um número entre 0 e 1, sendo 0,5 o valor utilizado pelos autores em [5].

Se houver sobreposição, a nuvem (regra) antiga deve ser excluída do modelo e os parâmetros do conseqüente são atribuídos à nova nuvem de dados. Se não sobrescrever, uma nova nuvem (regra) é criada. Após verificadas as condições acima, recalcula-se o densidade unimodal e a utilidade (Equação 4.30).

$$\eta_{K+1,i} = \frac{1}{K + 1 + I_i} \sum_{l=I_i}^{K+1} \lambda_{l,i}, \quad (4.30)$$

em que I_i é o momento em que a nuvem da dados foi estabelecida.

Então uma terceira condição é verificada: se todas as regras possuem uma utilidade maior que η_0 , em que η_0 foi considerada como 0,1 pelo autor [5]. Se houver regra com η_0 menor que o valor de referência, esta regra é excluída, se não, é mantida.

O parâmetro “ a ” dos conseqüentes das nuvens de dados (regras *fuzzy*) é atualizado com as equações 4.31 e 4.32:

$$C_{K+1,j} = C_{K,j} - \frac{\lambda_{K+1,j} C_{K,j} u_{K+1}^T C_{K,j}}{1 + \lambda_{K+1,j} u_{K+1}^T C_{K,j}}, \quad (4.31)$$

$$a_{K+1,j} = a_{K,j} + \lambda_{K+1,j} C_{K+1,j} u_{K+1} (y_{K+1} - u_{K+1}^T a_{K,j}). \quad (4.32)$$

Por último, é feita uma etapa de avaliação das variáveis. $q_{K+1,i,j}$ (Equação 4.33) é o somatório dos valores do parâmetro a ao longo do tempo de existência da regra *fuzzy* analisada. Em $w_{K+1,i,j}$ (Equação 4.34) o valor de $q_{K+1,i,j}$ é normalizado em relação às M variáveis sendo consideradas nesta regra. Finalmente, $\bar{w}_{K+1,j}$ (Equação 4.35) é a média por regra *fuzzy* de $w_{K+1,i,j}$. Se ($w_{K+1,i,j} < \bar{w}_{K+1,j}$) então a variável j deve ser removida da nuvem i . Ou seja, se o parâmetro referente a uma determinada variável tem sido um valor muito pequeno, sinalizando que a variável está sendo muito pouco preditiva, desde a criação da regra, esta variável pode ser excluída da regra (nuvem de dados).

$$q_{K+1,i,j} = \sum_{t=i}^{K+1} |a_{t,i,j}|, \quad (4.33)$$

$$w_{K+1,i,j} = \frac{q_{K+1,i,j}}{\sum_{j=1}^M q_{K+1,i,j}}, \quad (4.34)$$

$$\bar{w}_{K+1,j} = \frac{\sum_{i=1}^{N_{K+1}} w_{K+1,i,j}}{N_{K+1}}. \quad (4.35)$$

4.5 MÉTRICAS DE CLASSIFICAÇÃO BINÁRIA

Para medir o desempenho do modelo de classificação, somente a taxa de dados previstos corretamente não é o suficiente. Por exemplo, caso a base de dados possua uma classe rara e o modelo nunca previr essa classe rara, ele poderá ter uma boa taxa de acertos, porém não pode ser dito que este é um bom modelo. As principais métricas de classificação são baseadas nos valores de VP, VN, FP e FN [31] descritos abaixo:

- VP: verdadeiros positivos. Dados previstos como da classe 1 e realmente pertencentes a esta classe;
- VN: verdadeiros negativos. Dados previstos como da classe 0 e realmente pertencentes a esta classe;
- FP: falsos positivos. Dados previstos como da classe 1 e pertencentes a outra classe;
- FN: falsos negativos. Dados previstos como da classe 0 e pertencentes a outra classe.

A seguir, estão listadas as métricas de classificação mais difundidas. Um resumo completo destas e de outras métricas é encontrado em Hossin e Sulaiman (2015) [31].

- Acurácia (Equação 4.36): taxa total de acerto do modelo, ou seja, quantos dados foram previstos corretamente, independentemente da classe;

$$Acurácia = \frac{VP + VN}{VP + VN + FN + FP}. \quad (4.36)$$

- Erro (Equação 4.37): taxa total de erros do modelo, ou seja, quantos dados foram previstos incorretamente, independente da classe;

$$Erro = \frac{FN + FP}{VP + VN + FN + FP}. \quad (4.37)$$

- Precisão (Equação 4.38): taxa de dados previstos corretamente como da classe 1, comparado com a quantidade total prevista como 1.

$$Precisão = \frac{VP}{VP + FP}. \quad (4.38)$$

- Revocação (Equação 4.39): taxa de dados previstos corretamente como da classe 1, comparado com o número real de dados desta classe.

$$Revocação = \frac{VP}{VP + FN}. \quad (4.39)$$

- *F1 Score* (Equação 4.40): média harmônica entre a precisão e a revocação.

$$F1\ Score = \frac{2 * Precisão * Revocação}{Precisão + Revocação}. \quad (4.40)$$

4.6 TESTE ESTATÍSTICO

Após a coleta dos resultados, é importante realizar um teste estatístico para verificar se a diferença entre os resultados dos modelos é de fato significativa estatisticamente, ou seja, apesar de obter valores distintos, deve ser avaliado se a diferença entre os resultados pode ter sido causada por uma aleatoriedade ou se é de fato representativa.

O teste de Wilcoxon [71] é um teste estatístico não-paramétrico, ou seja, que não necessita que os dados sejam distribuídos de alguma forma específica, e é utilizado em dados dependentes, como, por exemplo, comparando dados de antes e depois de um evento ou comparando resultados de dois algoritmos sobre os mesmos dados. Enquanto o teste-t compara se a diferença da média de dois grupos é próxima de zero, o teste de Wilcoxon testa se a diferença entre o rank da subtração dos dois grupos e o rank esperado para a igualdade dos grupos é próximo de zero. O rank é um valor que inicia-se em 1 e aumenta unitariamente de acordo com a métrica especificada, ou seja, é uma medida de ordenação de valores.

A hipótese nula do teste é que os dados das duas amostras são originados da mesma distribuição. Rejeitar a hipótese nula significa dizer que de fato houve uma diferença entre um grupo de amostras e o outro. Segundo Welkowitz, Cohen e Lea (2021) [70], o teste de Wilcoxon é realizado através das seguintes etapas:

- Os valores das variáveis x e y a serem testadas são colocadas em colunas;
- Em uma terceira coluna, é colocado o módulo da subtração de cada item pareado.
- Na terceira coluna, do menor valor (em módulo) para o maior, é atribuído um valor de rank, começando em 1 e aumentando unitariamente. Caso a subtração seja igual a zero, este par de dados é descartado e em caso de empate é atribuída a média do rank que seria dos dados empatados;
- Na quarta coluna, coloca-se o sinal do resultado da subtração. Um exemplo é exibido na Tabela 6;

Tabela 6 – Exemplo da Tabela do Teste de Wilcoxon

Valor x	Valor y	Módulo da Diferença	Sinal	Rank
2,0	1,0	1,0	+	2,0
0,5	2,0	1,5	-	3,5
3,0	3,0	0,0	x	x
2,0	0,5	1,5	+	3,5
4,0	2,0	2,0	+	5
1,0	1,1	0,1	-	1

Fonte: O autor (2023).

- São computados separadamente o somatório dos valores de rank para os dados com sinal positivo e com sinal negativo;
- Dos dois somatórios, é mantido o de menor valor, chamado de T_1 , e descartado o outro;
- Em um rank com N pares é esperado que a soma de todos os ranks (T) seja igual a Equação 4.41, que é o equivalente a somatório de uma série de N números, começando em 1 e aumentando unitariamente até N ;

$$T = \frac{N(N + 1)}{2}. \quad (4.41)$$

- A hipótese nula considera que o valor esperado (T_E) da soma dos ranks negativos e da soma dos ranks positivos sejam aproximadamente iguais, ou seja, próximos da metade da equação anterior, conforme a Equação 4.42;

$$T_E = \frac{N(N + 1)}{4}. \quad (4.42)$$

- Ainda, é necessário calcular o erro padrão da distribuição dos ranks pareados (σ_{TM}), conforme a Equação 4.43;

$$\sigma_{TM} = \sqrt{\frac{(2N + 1)T_E}{6}}. \quad (4.43)$$

- A estatística z é por fim calculada na Equação 4.44;

$$z = \frac{T_1 - T_E}{\sigma_{TM}}. \quad (4.44)$$

- A hipótese nula será descartada se z for maior, em módulo, que o valor crítico. Considerando uma significância estatística de 5% e uma base de dados suficientemente grande para considerar que a distribuição dos ranks é normalmente distribuída, o valor crítico é considerado como $\pm 1,96$ [70];

5 EXPERIMENTOS COMPUTACIONAIS

Com o objetivo de gerar economia de recursos ao prever a necessidade de duplar um maquinista em escala de viagem longa quando este sobra na primeira parte de seu ciclo, no trecho de Juiz de Fora a Lafaiete da MRS Logística S.A., foi desenvolvido um estudo de análise de dados e aprendizado de máquina dividido em duas etapas. Esta pesquisa, inicia-se em 2019, em uma primeira etapa, diante de um cenário em que todos os maquinistas em excesso eram duplados e não havia ainda uma previsão de conseguir inserir modelos de aprendizado de máquina dentro do cotidiano da programação por restrições de natureza técnica. Nesta primeira fase do trabalho, foi realizada uma análise dos dados disponíveis e foi utilizada uma árvore de decisão para gerar diretrizes aos programadores de equipagens. Na segunda etapa deste trabalho, realizada em 2022, de posse de uma base de dados com mais amostras e mais variáveis, com um cenário corporativo mais adequado à implementação, diversos modelos de inteligência computacional foram estudados, com o objetivo de melhorar o resultado do primeiro experimento e de definir um modelo para aplicação na empresa.

A metodologia utilizada para os experimentos foi adaptada das recomendações de Géron (2019) [25] para realização de um projeto de aprendizado de máquina, e segue conforme a Figura 14:

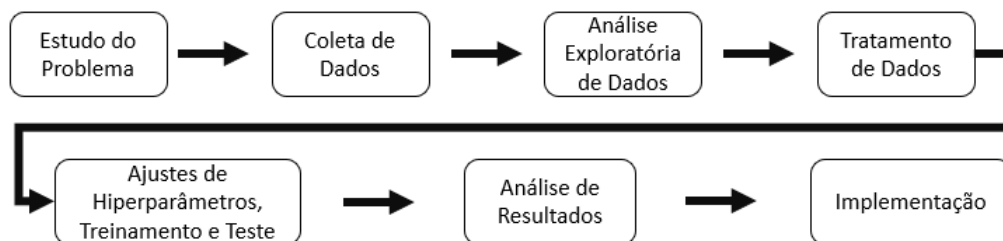


Figura 14 - Projeto de Aprendizado de Máquina.

Fonte: O autor (2023).

- Estudo do Problema: Avaliar se o problema é solucionável através de aprendizado de máquina;
- Coleta de Dados: Conseguir dados relevantes e em quantidade suficiente para aplicação dos modelos;
- Análise Exploratória de Dados: usar técnicas de análise de dados para obtenção de *insights*;
- Tratamento de Dados: aplicar técnicas para preparar a base de dados para utilização dos modelos;

- Ajuste de Hiperparâmetros, Treinamento e Teste: avaliar alguns modelos de aprendizado de máquina, ajustar seus hiperparâmetros e realizar as etapas de treinamento e teste;
- Avaliação dos Resultados: comparar os resultados obtidos por cada modelo nas métricas de maior relevância para o estudo;
- Implementação: a partir dos resultados, escolher o modelo ideal para implementação no ambiente de trabalho;

Os modelos de aprendizado de máquina foram rodados na linguagem *Python* 3.7 e, com exceção do ALMMo, utilizando a biblioteca *Scikit-Learn* [48]. O computador utilizado possui processador AMD Ryzen 7 5800X 8-Core 3.80 GHz, 16.0 GB de RAM e Windows 11 Pro.

5.1 O EXPERIMENTO DE 2019

5.1.1 Estudo do Problema

Em 2019, por não haver nenhuma informação sobre a necessidade do maquinista na etapa de retorno de seu ciclo, todos os maquinistas em excedente na programação de Juiz de Fora eram duplados e não dispensados. Com isso, a empresa perdia oportunidades de economizar com pagamento de horas improdutivas, horas extras, transporte, alimentação, hotel, entre outros. Introduzir novas tecnologias, mudança de processos, além de captação de dados úteis podem ser processos bem demorados para implementar no ambiente corporativo. Visando uma solução paliativa e de rápida implementação para o problema citado de ciclos ociosos, foi proposto um estudo para a definição de diretrizes a serem usadas pelos programadores de equipagem no momento de alocação do excedente. Técnicas de análise exploratória de dados podem ser úteis ao mostrar como as variáveis se relacionam e fornecer *insights* sobre o problema. Ainda, um modelo de árvore de decisão pode apoiar a definição de soluções, mesmo se não aplicado diretamente em ambiente de produção, já que, neste modelo, é possível entender as regras seguidas pela árvore para classificar uma amostra em um grupo.

5.1.2 Base de Dados

A base de dados utilizada foi construída em 2019, coletando dados dos trens sinalizados como duplados no gráfico horário de trens e da base de ciclos ociosos. A partir da experiência técnica do programador de equipagem, foram escolhidas 5 variáveis para o problema:

- *T_Drivers*: Variável numérica que representa o número total de maquinistas no trem;

- *N_Drivers*: Variável numérica que representa o número total de maquinistas no trem que cumprirão a escala de prontidão e estarão futuramente em FJC;
- *Headway_Before*: Variável numérica que representa a quantidade de horas que se passaram entre o último trem e o trem duplado. Desta forma, o modelo avaliará os impactos dos espaçamentos de circulação;
- *Headway_After*: Variável numérica que representa a quantidade de horas que se passaram entre o trem duplado e o próximo trem. Acredita-se que, pela experiência de trabalho do programador de equipagens, duplar quando há um grande espaçamento após o trem é mais eficaz, pois serão enviados maquinistas a mais para suprir o furo de apresentações em FJC, antes que este ocorra;
- *Target*: variável categórica que informa se o trem que foi duplado gerou um ciclo totalmente ocioso pro seu maquinista (*NOT OK*) ou se o maquinista foi necessário (*OK*).

Devido à dificuldade de obter dados específicos, como, por exemplo, o espaçamento de circulação após um trem que foi equipado com mais de um maquinista e que gerou ciclo ocioso, a coleta dos dados se deu de forma manual e lenta. Para sua construção, foi utilizada uma base de ciclos ociosos e para cada trem sinalizado, a informação de espaçamentos era obtida olhando manualmente o gráfico horário de trens. De forma a conseguir implementar a solução ainda em tempo de gerar resultados no mesmo ano, a etapa de coleta de dados encerrou-se com a obtenção de 80 amostras. Um exemplo da citada base segue na Tabela 7.

Tabela 7 – Exemplo da Base de Dados de 2019.

<i>T_Drivers</i>	<i>N_Drivers</i>	<i>Headway_Before</i>	<i>Headway_After</i>	<i>Target</i>
2	2	1,23	2,50	<i>OK</i>
2	2	1,21	3,62	<i>OK</i>
2	2	0,37	1,62	<i>NOT OK</i>
3	3	0,51	0,66	<i>NOT OK</i>
2	1	0,32	0,51	<i>OK</i>

Fonte: O autor (2023).

5.1.3 Análise Exploratória de Dados

Para comparação das correlações, os valores da variável alvo *Target* foram convertidos para número, sendo *OK* transformado em 0 e *NOT OK* em 1. Com isso, uma correlação grande e positiva entre uma variável e o alvo significará que, quanto maior esta variável, maior será a chance de ciclo ocioso.

A matriz de correlações está na Figura 15. A partir dela, é possível observar que o espaçamento depois do trem duplado (*Headway_After*) tem uma correlação maior, em

módulo, que o espaçamento anterior ao trem (*Headway_Before*). Isso reforça a intuição do programador de equipagem sobre a importância maior do *Headway_After* em relação ao *Headway_Before*. Ou seja, quanto maior o espaçamento após o trem que foi duplado (*Headway_After*), menor a probabilidade de haver ciclo ocioso. Observando o ciclo do maquinista, faz sentido pensar que um espaçamento de circulação na Linha do Centro causará um intervalo equivalente de envio de maquinistas para o retorno. Com isso, duplar o trem antes deste espaçamento fará com que haja mais maquinistas no P1-07 antes do intervalo de apresentações e este excedente poderá equipar os trens que chegarem durante o intervalo, sem realizações de ciclos ociosos, pois todos serão necessários para suprir esta falta de chegada de maquinistas. Já equipar após o espaçamento não possui a mesma chance de realização de ciclos úteis, afinal, ao chegar após o problema de intervalo de apresentações, os maquinistas chegarão em um momento junto com vários outros e a chance de algum deles realizar ciclo ocioso é mais alta. Além disso, por pequena diferença, é possível observar que *N_Drivers* possui uma correlação linear maior com *Target* que *T_Drivers*.



Figura 15 - Correlação Linear entre as Variáveis e o Alvo (2019).
Fonte: O autor (2023).

O gráfico da Figura 16 compara o número de maquinistas alocados em trem com o espaçamento de circulação anterior a este trem (*Headway_Before*) e se houve ciclo ocioso ou não. Em nenhum dos dois gráficos é possível observar alguma área de separação clara entre os dados.

Já na Figura 17, ao comparar o número de maquinistas alocados em trem, com o espaçamento de circulação após este trem (*Headway_After*) e se houve ciclo ocioso ou não, é possível ver uma separação melhor dos dados. Há uma relação visualmente linear entre

o $N_Drivers$ e o $Headway_After$, de forma que, para não gerar ciclo ocioso, o número de maquinistas no trem deve ser menor ou igual ao $Headway_After$ em horas. Importante somente ressaltar que para os casos mais extremos há poucos dados para comprovar se pode ser usada essa generalização. De forma semelhante, pode-se obter esta mesma relação ao substituir o $N_Drivers$ pelo $T_Drivers$, porém a divisão não é tão clara quanto com o $N_Drivers$.

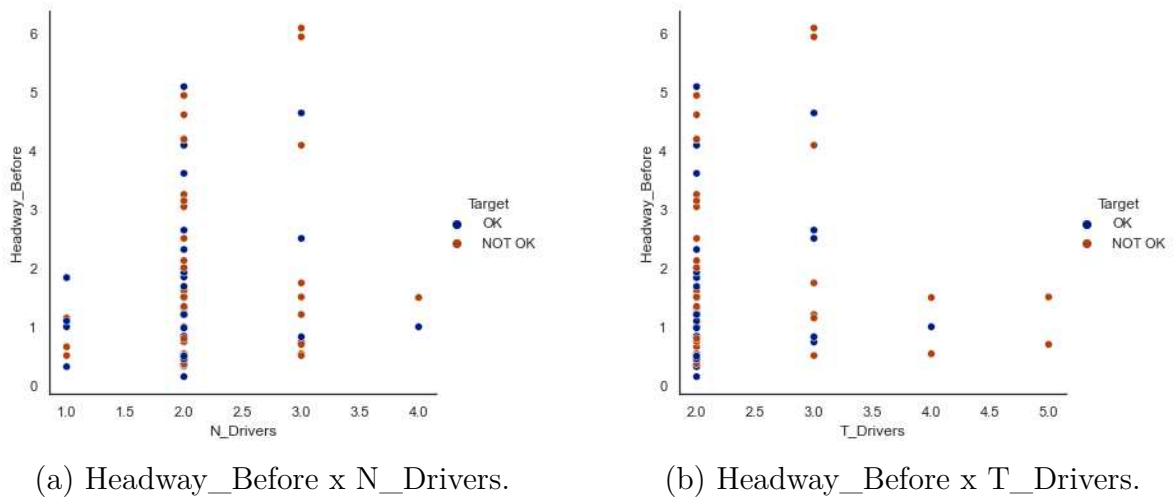


Figura 16 - Número de Maquinistas x Espaçamento Anterior de Trens.

Fonte: O autor (2023).

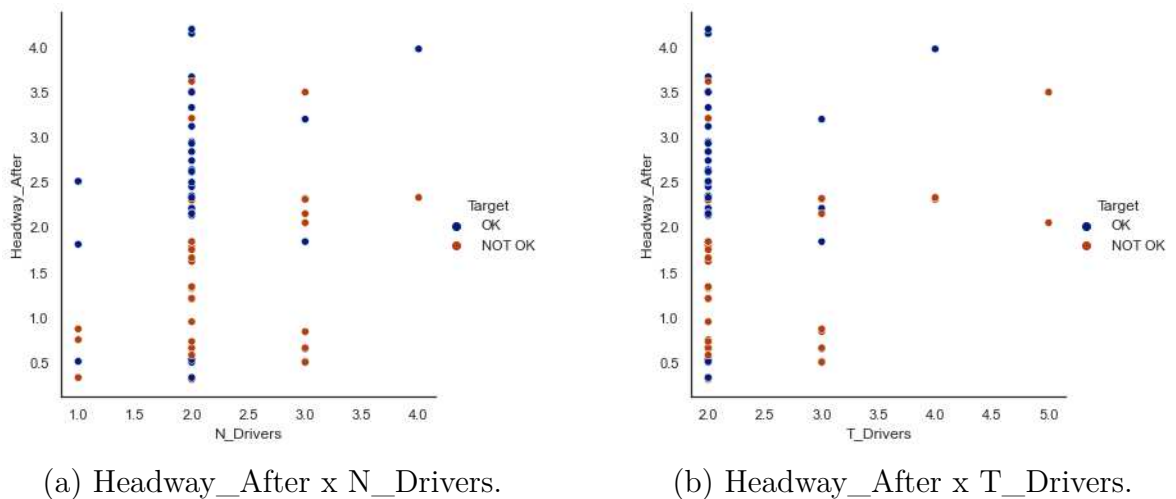


Figura 17 - Número de Maquinistas x Espaçamento Posterior de Trens.

Fonte: O autor (2023).

Concluindo a análise exploratória de dados, foi possível observar que quanto menor o $Headway_After$ maior a probabilidade de haver ciclos ociosos e o ponto de corte é proporcional em um pra um em relação ao $N_Drivers$.

5.1.4 Preparação dos Dados

Para utilizar a base de dados dentro do modelo da árvore de decisão, os valores da variável alvo *Target* foram novamente convertidos para número, sendo *OK* transformado em 0 e *NOT OK* em 1. Para maior clareza das regras a serem explicitadas e, devido aos modelos de árvore não serem sensíveis a variações de grandeza entre as variáveis, não foi aplicada nenhuma técnica de normalização ou de escalamento.

5.1.5 Ajuste de Hiperparâmetros

De forma a implementar regras simples, foi definido que a árvore deveria ter um máximo de dois níveis abaixo da origem. Afinal, uma árvore maior geraria um número grande de regras a serem analisadas pelo programador de equipagens, o que poderia gerar confusão, lentidão na tomada de decisão e até mesmo dificuldade na implementação do processo.

5.1.6 Resultados

A base de dados foi dividida entre treino e teste na proporção de 80% e 20%, respectivamente, e o código foi executado 100 vezes, com reordenação da base, para o cálculo das médias e desvios padrões das métricas de classificação exibidas na Tabela 8.

Tabela 8 – Resultados da Árvore de Decisão.

	Acurácia	Precisão	Revocação	F1 Score	Erro
Árvore	0,70 ± 0,10	0,68 ± 0,10	0,70 ± 0,18	0,67 ± 0,10	0,30 ± 0,10

Fonte: O autor (2023).

Como pode ser observado, a acurácia do modelo ficou em torno de 70%. Isto se deve, possivelmente, ao fato de poucos dados terem sido coletados e também de poucas variáveis terem sido avaliadas para o modelo. Ou seja, os espaçamentos de circulação e números de maquinistas podem não ser os únicos fatores a se levar em conta para realização de ciclo ocioso. Além disso, com uma base pequena, muitos cenários diferentes podem não ter sido contabilizados.

Ainda assim, devido à política anterior de duplar todo o excedente, foi entendido que uma metodologia inicial com 70% de acurácia e revocação traria benefícios e economia ao processo, pois sairia de uma alocação obrigatória e sem critérios para um processo em que somente 30% dos duplados fariam ciclos ociosos.

Analisando a árvore de decisão do modelo, Figura 18, pode-se observar que ela classifica todas as amostras com *Headway_After* menor que 1,795 horas como “ciclo ocioso”. As amostras com *Headway_After* maior que 1,795, se equipados com três ou mais maquinistas em viagem (*N_Drivers*) também serão classificadas como “ciclo ocioso”. Ou

seja, a classificação como “ciclo útil” só ocorre com um *Headway_After* maior que 1,795 e com dois ou menos *N_Drivers* no trem.

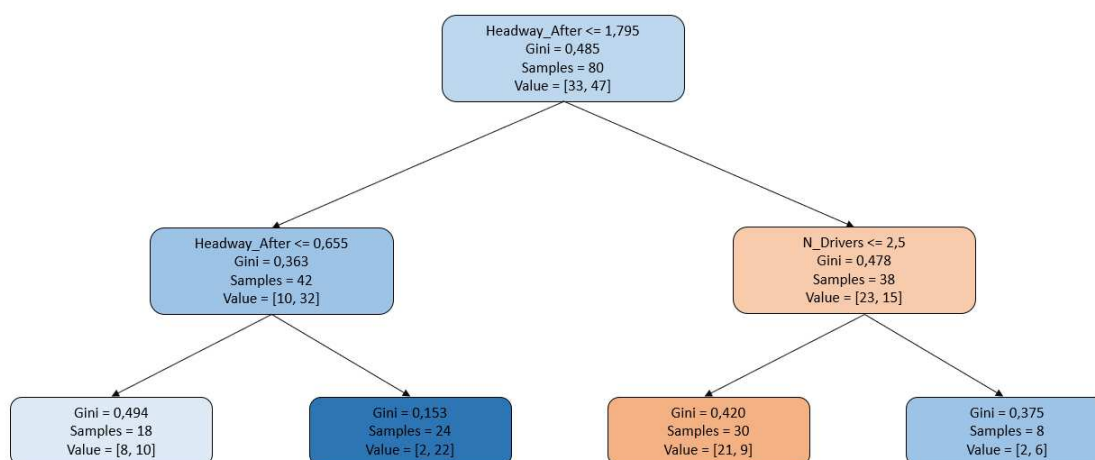


Figura 18 - Árvore de Decisão que Gerou a “Regra de Duplagem”.
Fonte: O autor (2023).

5.1.7 Aplicação na Empresa

As regras de classificação da árvore da Figura 18 foram simplificadas na seguinte diretriz:

- Somente duplar maquinistas em viagem se o espaçamento após o trem for maior que duas horas.

Esta diretriz passou a ser conhecida como “Regra de Duplagem” e começou a ser aplicada na programação de equipagem a partir do segundo trimestre de 2019.

Na Figura 19 é possível acompanhar a quantidade de maquinistas que realizaram ciclos ociosos, por trimestre, desde 2018 até 2022. A pedido da empresa, as quantidades de maquinistas foram representadas como porcentagem do pior resultado, sendo este no primeiro trimestre de 2018. Em vermelho está exibido o valor alvo desejado pela empresa de ciclos ociosos a serem realizados, sendo este reduzido a cada ano.

É possível observar que o uso da “Regra de Duplagem”, vinda da classificação da árvore de decisão, fez o número de ciclos ociosos cair mais da metade do valor inicial. Além disso, ao final de 2019 e 2020 os valores ficaram muito abaixo ao máximo desejado pela empresa. A iniciativa da “Regra de Duplagem” foi um caso de sucesso aplicado no ambiente corporativo.

A partir de 2021, houve um aumento dos ciclos ociosos realizados. Como, desde a criação da “Regra de Duplagem”, não houve uma análise atualizada dos dados e nem atualizações nas regras, com as mudanças de cenários inerentes da ferrovia, a diretriz

passou a ser menos eficiente após alguns anos. Ainda assim, manteve um patamar de redução bastante relevante se comparado aos trimestres em que esta ainda não era aplicada. Vale ressaltar que de 2020 para 2021 o alvo esperado pela empresa foi reduzido em 50%.

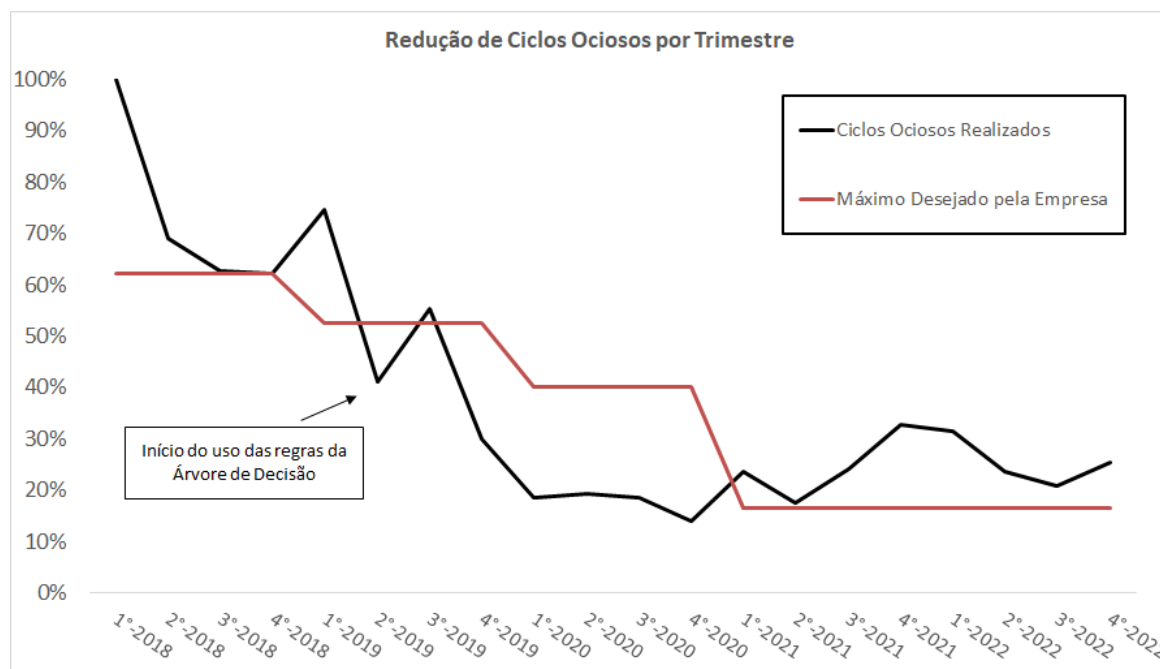


Figura 19 - Ciclos Ociosos vs Redução Esperada.

Fonte: O autor (2023).

5.2 O EXPERIMENTO DE 2022

5.2.1 Estudo do Problema

A partir do final de 2022, com diversas iniciativas internas nas áreas de inovação, a empresa está mais preparada para o uso direto de um modelo de aprendizado de máquina para solução de seus problemas. Uma nova pesquisa foi realizada com o objetivo de propor uma solução de inteligência computacional para definir quando duplar um trem na programação de equipagens de Juiz de Fora, para reduzir ainda mais a ocorrência de ciclos de maquinistas totalmente ociosos. Como, neste estudo de caso, o principal objetivo é prever, dado um cenário variável de trens em Juiz de Fora, se haverá a necessidade de duplar um trem para atender um futuro cenário no P1-07, sendo este mais constante, uma aplicação baseada em aprendizado de máquina torna-se uma boa metodologia para solucionar o problema. Além disso, esta deverá performar melhor que a solução baseada nas regras de 2019.

De posse de uma base de dados histórica ampliada, mostrando em quais cenários duplar o trem foi uma boa solução e em quais cenários não foi, seria possível fazer previsões com situações novas. Desta forma, o problema se encaixa em aprendizado de máquina supervisionado e de classificação (em que a variável alvo representa se houve, ou não, ciclo

ocioso ao duplar um trem). O desempenho do modelo será medido por sua acurácia, ou seja, a proporção de acertos do modelo. Porém, dado que o objetivo geral é evitar ciclos ociosos, a medida de revocação (quantidade de ciclos ociosos previstos corretamente sobre o total de ciclos ociosos realizados) também deve ser avaliada.

5.2.2 Base de Dados

A base de dados foi construída a partir de diversos cenários de trens entre os anos de 2018 e 2022, em que cada amostra representa um trem duplado em Juiz de Fora com a variável alvo representando se o ciclo foi totalmente ocioso ou não. Como descrito por Géron (2019) [25], a utilização da experiência humana para definição das variáveis é muito comum em modelos de aprendizado de máquina. Neste caso, foram utilizadas as seguintes variáveis para construção da base de dados:

- *T_Drivers*: Variável numérica que representa o número total de maquinistas no trem;
- *N_Drivers*: Variável numérica que representa o número total de maquinistas no trem que cumprirão a escala de prontidão e estarão futuramente em FJC;
- *N_Trains_Before*: Variável numérica que representa a quantidade de trens que passou antes do trem duplado, em um horizonte de quatro horas. Este horizonte foi definido devido ao limite do turno do maquinista ser de 12 horas, ou seja, ele realizando um trabalho de oito horas, pode ficar no máximo quatro horas improdutivo. Este seria então o limite em que o cenário de espera poderia ser avaliado para o maquinista. Esta variável e a próxima buscam avaliar o quanto o cenário estava saturado de trens e foram separados em “antes e depois”, devido à experiência prática levar a crer que o cenário de trens posterior teria maior influência que o anterior;
- *N_Trains_After*: Variável numérica que representa a quantidade de trens que passou após o trem duplado, em um horizonte de quatro horas;
- *Headway_Before*: Variável numérica que representa a quantidade de horas que se passaram entre o último trem e o trem duplado. Desta forma, o modelo avaliará os impactos dos espaçamentos de circulação;
- *Headway_After*: Variável numérica que representa a quantidade de horas que se passaram entre o trem duplado e o próximo trem. Semelhante à variável de quantidade de trens, também neste caso acredita-se pela experiência humana que duplar um trem antes de um espaçamento de circulação é o momento em que há maior chance de não gerar um ciclo ocioso;

- *Train_Type*: Variável categórica que representa o tipo de trem, sendo eles minério ou carga-geral. Como o carga-geral é um trem que demora mais no percurso, por fazer paradas intermediárias, o efeito de duplar o carga-geral pode ser diferente do minério;
- *Weekday*: Variável categórica que representa o dia da semana. Como a ferrovia opera de forma diferente a cada dia, seja por motivos de interrupções programadas na circulação, ou até mesmo dias específicos com menos trens, o dia da semana pode impactar na realização de ciclos ociosos;
- *Time*: Variável categórica que representa uma faixa de horário do dia: manhã, tarde ou noite. De forma similar à variável *Weekday*, esta variável busca captar os reflexos do funcionamento da ferrovia em diferentes períodos do dia;
- *Target*: variável categórica que informa se o trem que foi duplado no cenário descrito gerou um ciclo totalmente ocioso pro seu maquinista ou não.

A base de dados contém 997 amostras, sendo 521 corretamente duplados e 476 com a ocorrência de ciclos ociosos. Com isso, é possível afirmar que os dados estão aproximadamente balanceados. A Tabela 9 representa um pequeno exemplo da base de dados utilizada:

Tabela 9 – Exemplo da Base de Dados de 2022.

<i>T</i> <i>Driv.</i>	<i>N</i> <i>Driv.</i>	<i>Trains</i> <i>Before</i>	<i>Trains</i> <i>After</i>	<i>Head.</i> <i>Before</i>	<i>Head.</i> <i>After</i>	<i>Train</i> <i>Type</i>	<i>Day</i>	<i>Time</i>	<i>Target</i>
2	2	3	2	0,75	1,25	HH	Wed	Morn.	NOT OK
2	1	1	3	0,50	1,50	HH	Thu	Morn.	OK
2	2	2	4	2,00	0,50	HH	Thu	After.	OK
2	2	1	5	0,50	0,50	GC	Thu	Night	NOT OK
2	2	5	4	0,50	1,00	HH	Fri	Night	NOT OK

Fonte: O autor (2023).

5.2.3 Análise Exploratória de Dados

Para calcular a correlação entre as variáveis predictoras numéricas e a variável alvo, esta foi convertida de categórica para numérica considerando os eventos “OK” como 0 e os eventos com ciclos ociosos “NOT OK” como 1. Desta forma, uma correlação positiva significaria que quando maior aquela variável, maior sua contribuição para gerar ciclos ociosos em maquinistas duplados.

Confome observado na Figura 20, as correlações mais altas (em módulo) com a variável preditora foram encontradas nas variáveis *N_Trains_After* (quantidade de trens após o trem duplado) e *Headway_After* (tempo entre o trem duplado e o próximo). Ou seja, o cenário que ocorrerá na ferrovia após equipar o trem é mais importante na tomada

de decisão em relação a ciclo ocioso que o cenário que acabou de ocorrer. Desta forma, o maquinista extra chegará na outra sede antes do intervalo da chegada de maquinistas, ajudará suprir o furo de apresentações e não realizará um ciclo de retorno ocioso, estando disponível em um momento em que haverá grandes chances de não haver outro maquinista disponível.

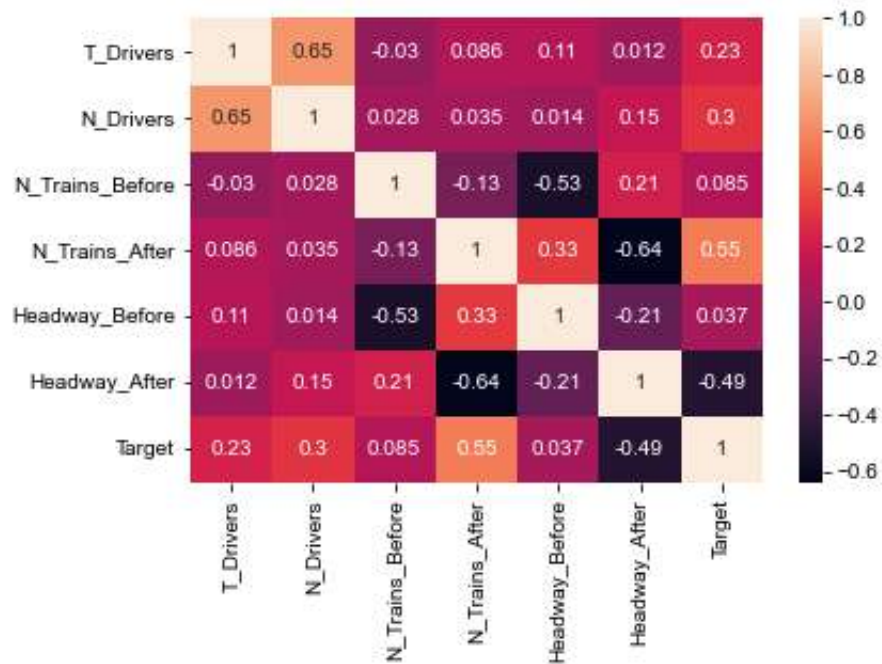


Figura 20 - Correlação Linear entre as Variáveis e o Alvo (2022).
Fonte: O autor (2023).

As variáveis *N_Trains_Before* e *Headway_Before* possuem baixo valor, em módulo, de correlação com a variável alvo. De início, pode se pensar que em um intervalo de circulação na Linha do Centro, duplar o primeiro trem da retomada de circulação poderia também ser útil para ajudar FJC a ter mais maquinistas rapidamente e se recuperar do furo de apresentações. Porém, não parece ser muito o caso. As baixas correlações indicam que este maquinista pode ou não ser útil, duplar a retomada não é garantia nem de utilização do maquinista nem de realização de ciclo ocioso e pode depender de outras variáveis do momento. Outra correlação lógica é com as variáveis que representam o número de maquinista em prontidão *N_Drivers* e *T_Drivers*, já que quanto mais maquinistas em um único trem, maior é chance de um deles fazer ciclo ocioso.

Além disso, algumas variáveis preditoras possuem correlação relevante entre si: *N_Drivers* e *T_Drivers*, já que quanto maior o número de maquinistas de regime de prontidão, maior o número de maquinistas no geral. As variáveis de *headway* e número de trens possuem alta correlação negativa e isso é justificado por ser esperado que quando há um grande espaçamento entre trens, que menos trens circulem nesse período de tempo.

A Figura 21 compara a variável “*N_Trains_After*” (maior correlação com o alvo)

com a ocorrência de ciclo ocioso. É possível observar a partir de três trens a ocorrência de ciclos ociosos aumenta significativamente.



Figura 21 - Ciclos Ociosos x Trens Posteriores.

Fonte: O autor (2023).

Ao avaliar os gráficos de caixa na Figura 22, é possível verificar uma linha de corte bem visível nas variáveis que representam o cenário que veio após o trem duplado. Sem muitos *outliers*, os gráficos de caixa reforçam que os ciclos ociosos acontecem quando há três ou mais trens nas próximas quatro horas (N_Trains_After). De forma semelhante, quando há um espaçamento posterior ($Headway_After$) menor que duas horas também há grande ocorrência de ciclos ociosos, porém este com mais dados *outliers*. Isto reforça o ponto de o N_Trains_After ser mais correlacionado com a variável alvo que o $Headway_After$.

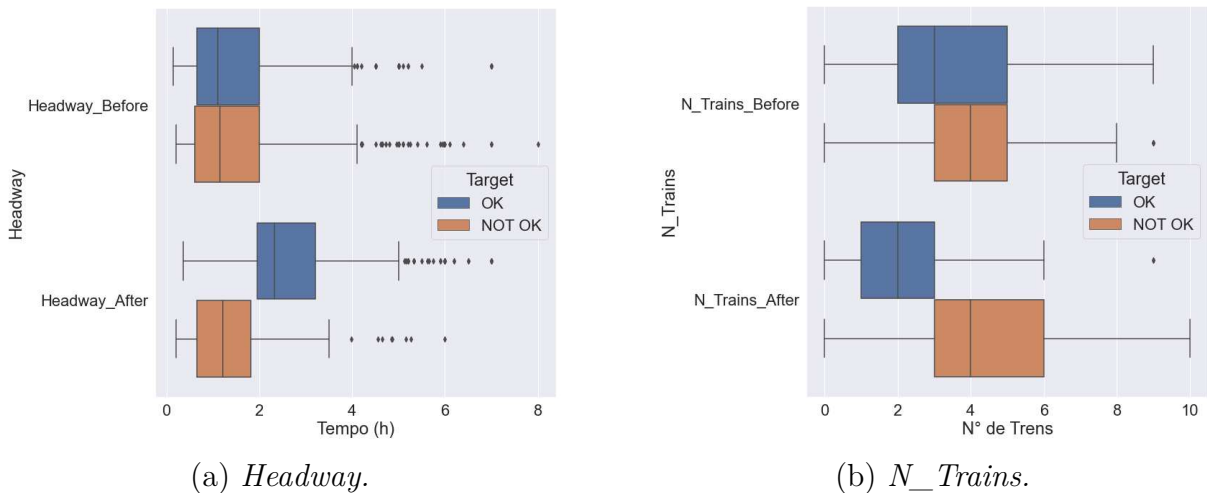


Figura 22 - Gráfico de Caixa para N_Trains e Headway.

Fonte: O autor (2023).

Uma análise das variáveis categóricas é apresentado na Figura 23. Há uma porcentagem um pouco maior de ciclos ociosos em trens GC (*General Cargo*), porém esta diferença não é muito expressiva. O mesmo pode ser dito dos dias da semana, com

pequenas variações entre si. Mas, ao avaliar a variável de horário, é possível perceber que ao duplar de manhã há menos chances de haver ciclos ociosos que duplar à noite. Isso pode ser devido às programações de manutenção que ocorrem na parte da manhã e interrompem a circulação, ou até a algum cenário não percebido de FJC mais comum a noite, em que o reflexo de chegada dos maquinistas da manhã de FDT ajudam a suprir.

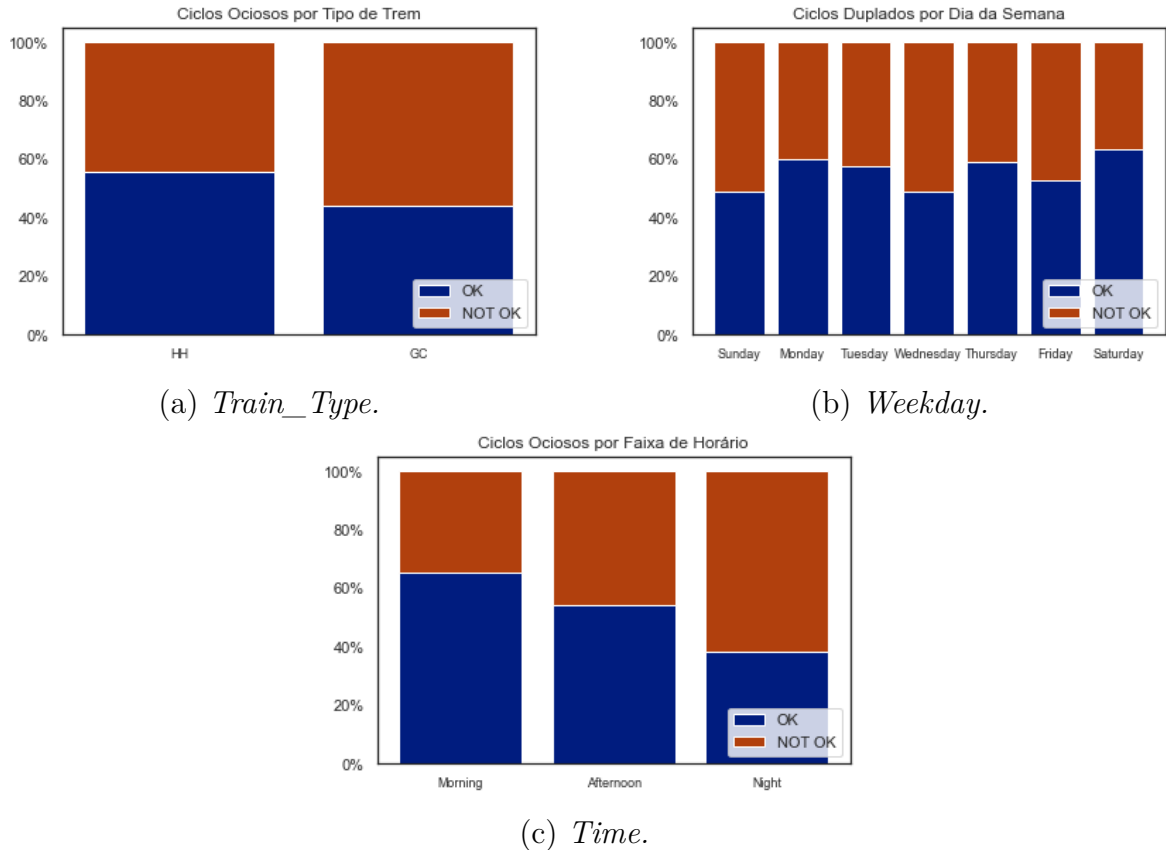


Figura 23 - Histograma das Variáveis Categóricas.
Fonte: O autor (2023).

Concluindo a análise exploratória, foi observado o número de trens posterior ao trem duplado (N_Trains_After) é mais correlacionado com a realização de ciclos ociosos que o espaçamento de trens ($headway$), sendo o número crítico igual a dois trens, ou seja, se duplar um trem com mais do que dois outros trens circulando nas próximas quatro horas, a probabilidade de realização de ciclo ocioso já aumenta significativamente. Ainda, o horário em que foi realizada a sobre-alocação ($Time$) também influencia na ocorrência de ciclos ociosos. Esses fatores não avaliados no primeiro experimento ajudam a explicar o motivo da árvore de decisão de 2019 ter chegado somente a 70% de acurácia: além de uma base de dados reduzida, nem todas as variáveis importantes estavam sendo avaliadas.

5.2.4 Preparação dos Dados

Para aplicação dos algoritmos de aprendizado de máquina, as variáveis categóricas foram convertidas com o *Ordinal Encoder* da biblioteca de *Python Scikit-Learn* [48]. O

Ordinal Encoder foi escolhido por haver uma certa ordem nos valores das classes, como, por exemplo, manhã, tarde ou noite. Além disso, o uso alternativo do *One Hot Encoder* geraria uma coluna para cada classe, o que aumentaria bastante o número de variáveis do problema. A variável *Train_Type* com os valores “HH” ou “GC” viraram 0 e 1, respectivamente. *Target*, a variável preditora, recebeu 0 para “OK” e 1 para “NOT OK”. Cada dia da semana em “*Weekday*” recebeu um número sequencial de 0 a 6, assim como “*Time*” recebeu números de 0 a 2 para cada período do dia. Após a conversão, todos os dados foram escalados de 0 a 1, usando a função *MinMaxScaler* do *Scikit-Learn* [48], pois alguns métodos de aprendizado, como SVM e Rede Neural, são sensíveis à escala das variáveis.

5.2.5 Ajuste de Hiperparâmetros

Com o objetivo de implementar uma solução que pudesse ser aplicada na empresa, cinco algoritmos de aprendizado de máquina foram testados.

- Rede Neural;
- SVM;
- Árvore de Decisão;
- Floresta Aleatória;
- ALMMo.

A escolha dos modelos se deve principalmente às suas características, como o entendimento de suas regras internas, no caso da árvore de decisão, ser evolutivo, para o ALMMo, e serem algoritmos de alto poder preditivo com muitas aplicações realizadas na literatura, como a rede neural, SVM e floresta aleatória. De forma a simplificar a nomenclatura, a rede neural também será identificada como “MLP” (do inglês *Multi-Layer Perceptron*), a árvore de decisão como “*Tree*” e a floresta aleatória como “*Rnd*”.

Para a definição dos hiperparâmetros de cada algoritmo, foi utilizada a biblioteca *Grid Search* do *Scikit-Learn* [48] com validação cruzada de 5 *folds*. Na Tabela 10, estão listados os hiperparâmetros testados. Para o ALMMo, por este modelo ser não paramétrico, não houve etapa de ajuste de hiperparâmetros. Como a base de dados pode ser considerada pequena e com poucas variáveis, os modelos eram treinados em poucos segundos, ou até milésimos de segundos, dependendo dos hiperparâmetros sendo testados. Com este tempo de treinamento, os modelos poderiam ser usados no dia a dia da operação, sem causar atrasos nas tomadas de decisão. O *early stopping* da rede não estava entre os hiperparâmetros avaliados, porém seria uma opção para diminuir ainda mais o tempo total de treinamento.

Tabela 10 – Hiperparâmetros Testados.

Modelo	Parâmetro	Valores
Rede Neural	Arquitetura das camadas escondidas	(9), (8), (7), (6), (5), (4) (3), (9,5), (7,4), (5,3) (3,2), (7,4,2), (6,3,2).
	Ativação da camada escondida	Tanh, ReLu, Sigmoide.
	Solver	Adam, gradiente descendente estocástico e lbfgs.
	Alpha (regularização)	0,0001, 0,001, 0,01, 0,1. 0, 1, 10, 100, 1000.
	Taxa de aprendizado	Constante ou adaptativo.
	Taxa de aprendizado inicial	0,0001, 0,001, 0,01, 0,1, 0, 1, 10.
SVM	C (regularização)	0,001, 0,01, 0,1, 1, 10,100.
	kernel	'linear', 'poly', 'rbf', 'sigmoid'.
	Coefficiente da função kernel	0,001, 0,01, 0,1, 1, 10, 100.
	Coefficiente independente da função kernel	0,001, 0,01, 0,1, 0, 1, 10, 100.
	Grau do kernel polinomial	1, 2, 3, 4, 5.
Árvore de Decisão	Critério de divisão	Gini ou Entropia.
	Máximo de variáveis para realizar divisão	1, 2, 3, 4, 5, 6, 7, 8, 9.
	Profundidade máxima da árvore	1, 2, 3, 4, 5, 6, 7, 8, 9.
	Mínimo de amostras para realizar divisão	30, 25, 20, 15, 10, 5, 3, 2.
	Mínimo de amostras para ser folha	35, 30, 25, 20, 15, 10, 5, 3, 2, 1.
Floresta Aleatória	Número máximo de folhas	10, 9, 8, 7, 6, 5, 4, 3, 2.
	Quantidade de árvores no ensemble	100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 5.
	Critério de divisão	Gini ou Entropia.
	Máximo de variáveis para realizar divisão	1, 2, 3, 4, 5, 6, 7, 8, 9.
	Profundidade máxima da árvore	1, 2, 3, 4, 5, 6, 7, 8, 9.
	Mínimo de amostrar para realizar divisão	30, 25, 20, 15, 10, 5, 3, 2.
	Mínimo de amostras para ser folha	35, 30, 25, 20, 15, 10, 5, 3, 2, 1.
Bootstrap	Verdadeiro ou Falso.	

Fonte: O autor (2023).

As configurações escolhidas para cada um dos algoritmos estão na Tabela 11, sendo esta escolha baseada no melhor resultado de acurácia da validação cruzada de cada combinação do *Grid Search*. A rede neural convergiu com a taxa de aprendizado e o número máximo de iterações selecionados.

5.2.6 Resultados

A base de dados foi dividida em 80% para treino e 20% para testes e treinado nos cinco modelos de aprendizado citados. O experimento foi repetido 100 vezes, reordenando a base aleatoriamente. Após, foram computadas a média e desvio padrão para as métricas: acurácia, precisão, revocação, *F1 Score* e o erro médio quadrado. Os resultados foram compilados na Tabela 12.

Na Figura 24 é mostrada a função densidade de probabilidade para as métricas acurácia e revocação. Como pode ser observado, as curvas referentes à floresta aleatória

aparentam ser bimodais, não seguindo uma distribuição normal.

Tabela 11 – Hiperparâmetros Seleccionados.

Modelo	Parâmetro	Valores
Rede Neural	Camadas escondidas	(3,2)
	Ativação da camada escondida	Tanh
	Solver	Adam
	Alpha (regularização)	0,001
	Taxa de aprendizado	Constante
	Taxa de aprendizado inicial	0,001
SVM	C (regularização)	0,01
	kernel	'poly'
	Coefficiente da função kernel	10
	Coefficiente independente da função kernel	10
	Grau do kernel polinomial	3
Árvore de Decisão	Critério de divisão	Entropia
	Máximo de variáveis para realizar divisão	5
	Profundidade máxima da árvore	5
	Mínimo de amostras para realizar divisão	10
	Mínimo de amostras para ser folha	2
Floresta Aleatória	Número máximo de folhas	9
	Quantidade de árvores no ensemble	40
	Critério de divisão	Gini
	Máximo de variáveis para realizar divisão	6
	Profundidade máxima da árvore	8
	Mínimo de amostrar para realizar divisão	2
	Mínimo de amostras para ser folha	3
Bootstrap	Verdadeiro	

Fonte: O autor (2023).

Tabela 12 – Desempenho dos Modelos (com Desvio Padrão).

	Acurácia	Precisão	Revocação	F1 Score	Erro
ALMMo	0,86 ± 0,02	0,85 ± 0,03	0,86 ± 0,03	0,85 ± 0,02	0,14 ± 0,02
MLP	0,88 ± 0,03	0,87 ± 0,04	0,87 ± 0,06	0,87 ± 0,04	0,12 ± 0,03
SVM	0,88 ± 0,02	0,88 ± 0,03	0,87 ± 0,03	0,87 ± 0,02	0,12 ± 0,02
Tree	0,86 ± 0,02	0,85 ± 0,04	0,86 ± 0,04	0,85 ± 0,02	0,14 ± 0,02
Rnd	0,90 ± 0,02	0,89 ± 0,03	0,91 ± 0,03	0,90 ± 0,02	0,10 ± 0,02

Fonte: O autor (2023).

Para verificar a significância estatística dos resultados, o teste não paramétrico de Wilcoxon foi utilizado, em que a hipótese nula representa que os dados são de mesma população e rejeitar a hipótese nula significa que os dados são de fato vindos de distribuições diferentes. O teste de Wilcoxon foi escolhido por não necessitar que os dados estejam distribuídos normalmente e por trabalhar com dados emparelhados, podendo ser usado para comparar os resultados de diferentes algoritmos sobre o mesmo dado. A significância estatística utilizada foi de 5%. Na Tabela 13 estão os valores p obtidos, arredondados na quarta casa decimal, ao comparar os resultados das acurácias dos modelos. É possível

observar que o modelo que obteve o melhor resultado, a floresta aleatória, rejeitou a hipótese nula contra todos os outros modelos. Desta forma, pode-se dizer que os resultados dela não são originados de mesma distribuição que os outros modelos. Já o MLP não obteve significância quando comparado ao SVM, e o ALMMo também não obteve significância quando comparado à árvore. Com isso, não é possível afirmar que o MLP performou melhor que o SVM e nem o ALMMo em relação à árvore.

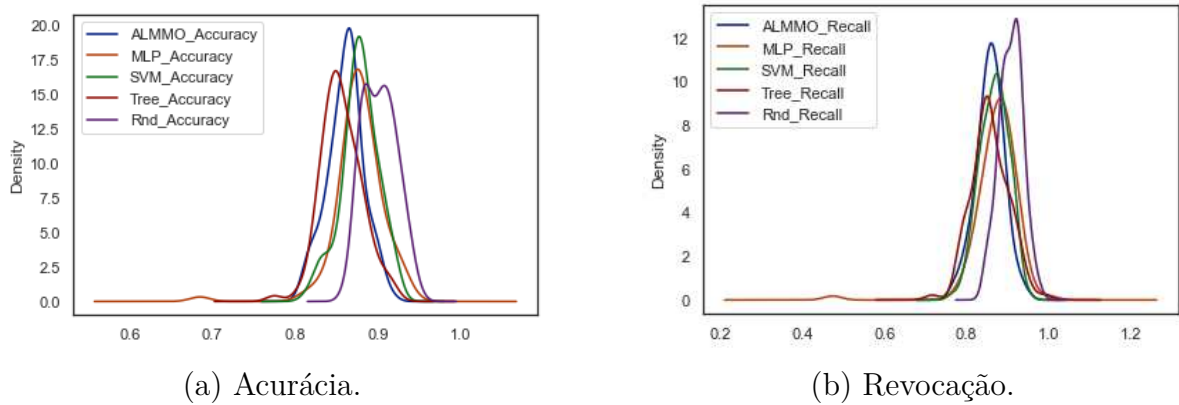


Figura 24 - Curva de Densidade para os Modelos.

Fonte: O autor (2023).

Tabela 13 – Valores P do Teste de Wilcoxon para a Acurácia dos Modelos.

	ALMMo	MLP	SVM	Tree	Rnd
ALMMo	-	0,0000	0,0000	0,0861	0,0000
MLP	0,0000	-	0,7629	0,0000	0,0000
SVM	0,0000	0,7629	-	0,0000	0,0000
Tree	0,0861	0,0000	0,0000	-	0,0000
Rnd	0,0000	0,0000	0,0000	0,0000	-

Fonte: O autor (2023).

O teste também foi realizado para a revocação, conforme Tabela 14, com resultados semelhantes à acurácia. Porém, além da floresta aleatória, o SVM e o MLP rejeitaram a hipótese nula em relação a todos os outros modelos.

Tabela 14 – Valores P do Teste de Wilcoxon para a Revocação dos Modelos.

	ALMMo	MLP	SVM	Tree	Rnd
ALMMo	-	0,0000	0,0017	0,8163	0,0000
MLP	0,0000	-	0,0041	0,0002	0,0000
SVM	0,0017	0,0041	-	0,0362	0,0000
Tree	0,8163	0,0002	0,0362	-	0,0000
Rnd	0,0000	0,0000	0,0000	0,0000	-

Fonte: O autor (2023).

A Figura 25 exibe a quantidade de eventos ociosos que a floresta aleatória geraria, mostrando um ganho perceptível em relação a atual “regra de duplagem”. Pode-se notar

que com o uso do modelo, os resultados ficam bem abaixo do valor máximo (alvo) esperado pela empresa, ou seja, atingindo com folga os resultados esperados.

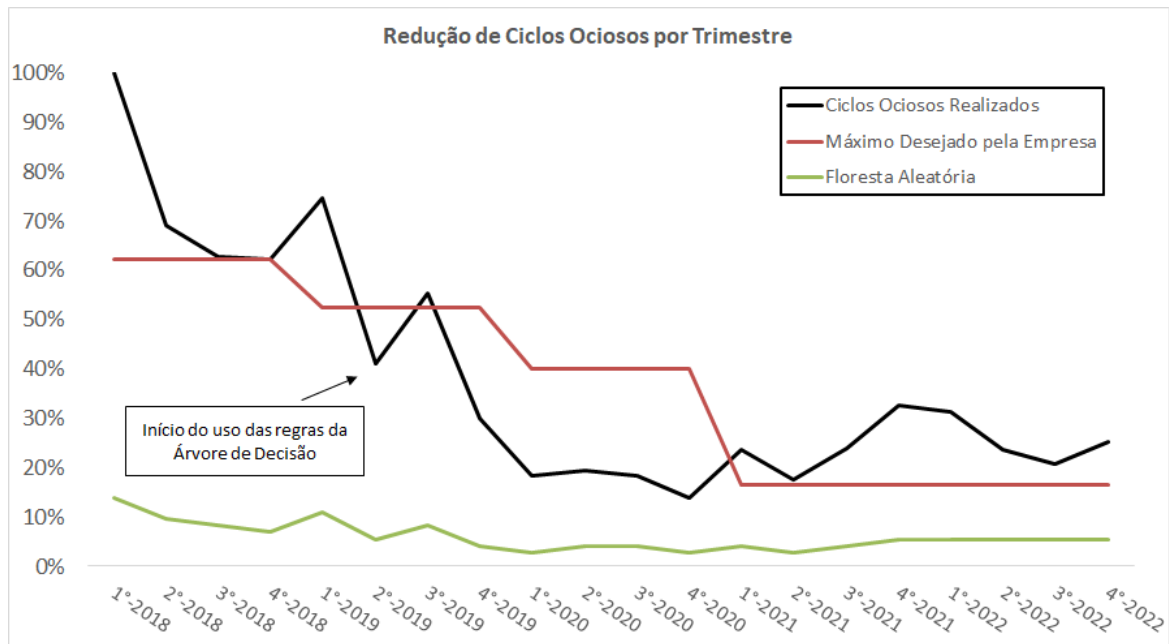


Figura 25 - Ciclos Ociosos com Floresta Aleatória.

Fonte: O autor (2023).

É interessante notar que na outra aplicação de aprendizado de máquina na literatura [24], foi proposta também uma floresta aleatória para sua solução, em que o modelo possuía 86,7% de precisão e 87,9% de revocação. Mas vale ressaltar que, em [24], a intenção era prever se os maquinistas estariam satisfeitos com a alocação e não era um estudo focado em economia de recursos, como este. Então, as similaridades são resumidas em somente serem duas pesquisas de inteligência computacional aplicadas em alocação de equipagem ferroviária que obtiveram melhores resultados com florestas aleatórias. Isso pode ser devido à característica deste ser um modelo *ensemble* que avalia diversas árvores e funciona fazendo a previsão como a resposta mais votada.

5.2.7 Aplicação na Empresa

O modelo de floresta aleatória foi o que atingiu os melhores valores em todas as métricas utilizadas. Logo, este é o modelo recomendado que corresponderá à melhor performance de classificação.

Contudo, considerando que os modelos de árvore e ALMMo tiveram um desempenho somente 4% pior do que a floresta, estes podem ser recomendados por possuírem outras características de interesse. A árvore de decisão é um modelo que permite que suas regras sejam explícitas, desta forma, torna-se um modelo indicado para o meio corporativo, no qual as tomadas de decisão são melhores aceitas quando baseadas em regras explícitas do que em algo “caixa-preta”.

Já o ALMMo tem uma outra característica importante para o dia a dia da operação. Como observado na criação da “regra de duplagem”, esta começou a ficar obsoleta, pois foi baseada em um cenário e com o passar do tempo, ajustes necessitariam ser feitos para que continuasse valendo. O mesmo pode-se dizer dos modelos de aprendizado de máquina, estes devem ser retreinados com dados mais recentes de tempos em tempos para que continuem com uma boa performance. Com isso, a grande vantagem do ALMMo é ser evolutivo, ou seja, não precisar de retreinar a base inteira, somente adicionar os dados novos.

Na Figura 26, vemos quantos ciclos ociosos seriam gerados por mês se fosse utilizada a árvore de decisão ou o ALMMo. É possível observar que os ciclos ociosos ainda ficariam em um patamar bem abaixo do alvo máximo definido pela empresa. Com isso, é bastante provável que a empresa prefira um modelo mais explícito como a árvore para justificar suas decisões, ao custo de uma pequena perda de acurácia, que usar um método *ensemble*, “caixa-preta”, como a floresta aleatória.

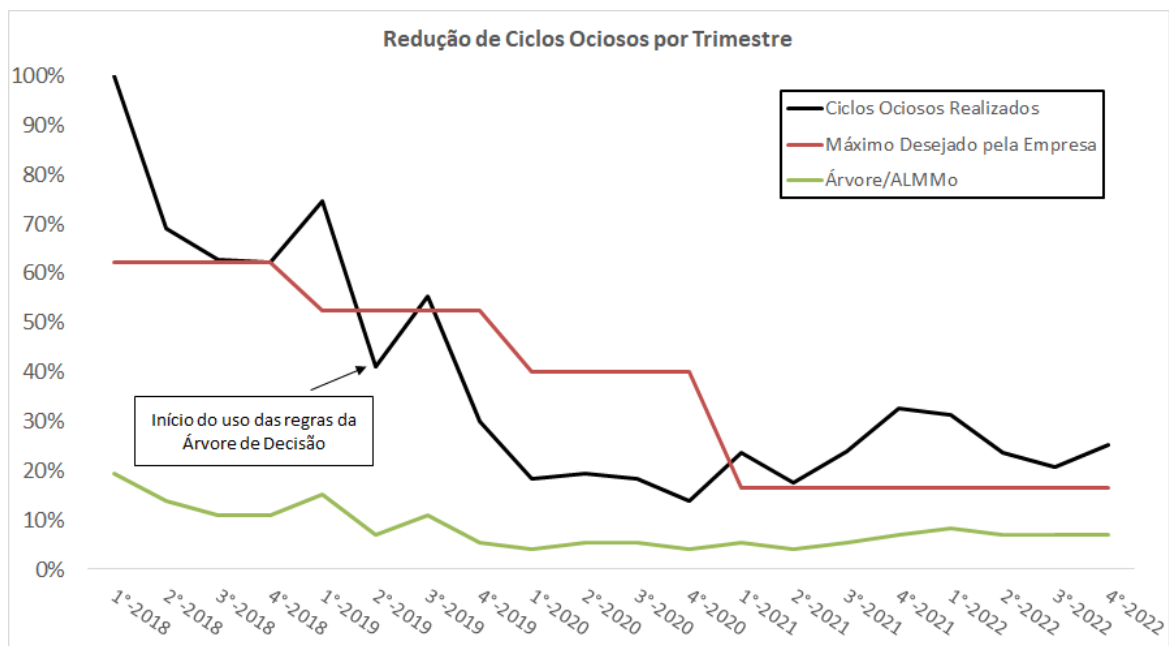


Figura 26 - Ciclos Ociosos com Árvore/ALMMo.
Fonte: O autor (2023).

Como todos os modelos identificaram ciclos ociosos dentro do nível de serviço (alvo) esperado pela empresa, ficará a cargo desta definir qual será sua prioridade ao implementar o modelo de previsão. Caso decida adotar o modelo de melhor acurácia, deverá utilizar a floresta aleatória. Se sua prioridade for ter um modelo mais explícito em suas decisões, a árvore ainda performa dentro esperado. Por último, caso escolha ir para o caminho de facilitar o treinamento de novos dados, sem armazenamento de grandes bases, permitindo o algoritmo ir evoluindo sem necessidade de parada para retreino, o ALMMo também obteve resultados satisfatórios.

Ainda, até que a empresa inicie a aplicação do modelo, vale ressaltar a necessidade de atualização da “regra de duplagem”. Pela análise exploratória de dados, foi visto que a variável N_Trains_After está mais correlacionada com ciclos ociosos, e começa a gerá-los a partir do valor de três trens. Então, ao duplar somente se o cenário tiver dois ou menos trens após o trem com sobre-alocação, é esperado um resultado melhor que a regra antiga, que tinha como referência valores de *headway*.

6 CONCLUSÃO

Com o desenvolvimento deste trabalho foi proposta uma solução de aprendizado de máquina para diminuir o número de maquinistas realizando ciclos ociosos na MRS Logística S.A. Para iniciar o estudo, foi feita uma revisão dos conceitos, literatura e legislações envolvidas no processo, incluindo vários estudos dentro da própria empresa. No cenário anterior à pesquisa, inicialmente todos os maquinistas em excendente eram usados em tarefas ociosas, com a expectativa que fossem necessários em algum momento de seu ciclo. Através de um estudo inicial com uma base de dados bem reduzida, foi criada uma regra interna para somente sobre-alocar uma tarefa se houvesse um espaçamento de circulação igual ou maior que duas horas após o trem. Esta regra foi aplicada e gerou bons resultados nos primeiros anos de aplicação, reduzindo os ciclos ociosos em mais da metade do valor inicial, conforme Figura 19. Com o passar dos anos e mudanças do cenário foi percebido que era necessária uma alternativa a esta diretriz rígida.

Ao treinar cinco algoritmos diferentes de inteligência computacional, foi obtido um percentual de redução ainda maior que o obtido pela “Regra de Duplagem”, conforme Figura 25. Foi visto ainda que a floresta aleatória obteve o melhor desempenho nas métricas de classificação, porém os outros modelos tiveram um resultado somente 4% inferior, no pior dos cenários, não representando uma perda financeira expressiva em gastos mensais de equipagem, conforme Figura 26. Com isso, a empresa fica com três opções de utilização de acordo com o que atender melhor suas futuras necessidades. O uso da floresta aleatória é recomendado em caso da empresa preferir a melhor acurácia e revocação, mesmo os ganhos sendo pequenos frente às outras soluções. O uso da árvore de decisão pode ser o ideal, pois esta gera melhor clareza de suas regras internas, sendo de mais fácil aplicação no meio corporativo. Por último, o ALMMo ainda é vantajoso por fornecer uma solução que evolui sem precisar ser retreinada na base de dados inteira.

Além disso, enquanto o modelo não é implementado, a partir da análise exploratória de dados, podemos concluir que a regra anterior que tinha como parâmetro o espaçamento de circulação não era a mais adequada, pois ao se ter uma base de dados maior e com mais variáveis, concluiu-se que o ciclo ocioso estava mais relacionado à quantidade de trens posteriores (N_Trains_After) que ao espaçamento de circulação ($Headway_After$) em si. É proposta uma atualização da regra, até que o modelo de aprendizado seja aplicado, em que só seja possível duplar um maquinista se só houver no máximo dois trens previstos nas próximas quatro horas.

Vale ressaltar que aplicações de inteligência computacional dentro da alocação de equipagem ferroviária é um tópico pouco estudado. Na revisão bibliográfica, somente um artigo foi encontrado e com um escopo bem diferente deste trabalho, em que o objetivo era ganhar poder computacional na otimização, enquanto este trabalho propõe-se a prever

se o maquinista será necessário em seu ciclo completo, ressaltando a contribuição deste para a literatura.

Os objetivos deste trabalho foram alcançados. Através de estudo dos dados e criação de árvore de decisão, a “Regra de Duplagem” levou a uma redução no uso de maquinistas duplados. Com a avaliação do desempenho de modelos de inteligência computacional, foi definido que a floresta aleatória possui maior acurácia, mas a árvore de decisão e o ALMMo também são boas opções para serem implementadas na empresa, devido às suas características.

6.1 TRABALHOS FUTUROS

Com este trabalho foi possível propor um modelo de aprendizado de máquina que auxiliasse na tomada de decisão a respeito de quando sobre-cobrir uma tarefa durante a alocação de equipagem de ultra-curto prazo. A otimização da alocação dos maquinistas é um tópico recorrente de pesquisas, inclusive dentro da MRS Logística S.A. Com isso, a primeira sugestão de trabalho futuro é o desenvolvimento de um modelo de otimização que aloque os maquinistas da maneira ótima e que utilize o modelo de aprendizado discutido neste trabalho para fazer a alocação ou dispensa do excedente.

Há ainda muitas outras oportunidades dentro da programação de equipagem ferroviária. Por exemplo, o horário de chegada previsto dos trens, principalmente os mais distantes, sofrem bastante alteração em seu percurso. Uma oportunidade de estudo seria avaliar um modelo que conseguisse melhorar a previsão dos trens, pois com uma confiabilidade maior da previsão recebida, o programador de equipagem pode tomar decisões melhores com mais antecedência.

Por último, como foi visto, a programação de equipagens era feita localmente em cada sede e foi posteriormente transferida para ser mais centralizada no centro de controle, o que gerou uma visão macro do sistema e redução de horas improdutivas dos maquinistas. Sugerimos o estudo de um otimizador global de equipagem, que não só otimizasse o cenário de sedes próximas, como de toda a malha. Desta forma, seria encontrado a solução ótima do cenário inteiro da malha, incluindo necessidade de ir com um maquinista de uma sede para a outra. Juntamente com este otimizador global, aplicações com aprendizado de máquina para alocar a sobra e prever trens seriam adicionais que contribuiriam para a melhoria da solução.

REFERÊNCIAS

- [1] ABBINK, E.; HUISMAN, D.; KROON, L. Railway Crew Management. In: BORNDÖRFER, R. et al. **Handbook of Optimization in the Railway Industry**. 1 ed. Cham: Springer, 2018. p.243-264.
- [2] ALBERS, M. **Freight railway Crew Scheduling: Models, Methods and Applications**. 1 ed. Köln: Logos Verlag, 2009.
- [3] ALPAYDIM, E. **Introduction to Machine Learning**. 3 ed. Cambridge: MIT Press, 2014.
- [4] ANGELOV, P. P.; GU, X. **Empirical Approach to Machine Learning**. 1 ed, Cham: Springer, 2019.
- [5] ANGELOV, P. P.; GU, X.; PRÍNCIPE, J. C. Autonomous Learning Multimodel Systems From Data Streams. **IEEE Transactions on Fuzzy Systems**, v.26, n.4 p.2213–2224, 2018.
- [6] ANGELOV, P. P.; Yager, R. A new type of simplified fuzzy rule-based system. **International Journal of General Systems**, v.41, p.163-185, 2012.
- [7] ANTF. Informações gerais. **Associação Nacional dos Transportadores Ferroviários**. Disponível em: <https://www.antf.org.br/informacoes-gerais/>. Acesso em 18 de maio de 2022.
- [8] BIGONI, C. Optimization of phase prediction for brain-state dependent stimulation: a grid-search approach. **Journal of Neural Engineering**, v.20, 2023.
- [9] BOSER, B. E.; GUYON, I.M.; VAPNIK, V.N. A Training Algorithm for Optimal Margin Classifiers. **Proceedings of the Fifth Annual Workshop on Computational Learning Theory**, p.144-152, 1992.
- [10] BRASIL. Concessões, renovações antecipadas e autorizações formam os pilares do crescimento ferroviário, diz Tarcísio. **Ministério da Infraestrutura**. Brasília, DF, 2021. Disponível em: <https://www.gov.br/infraestrutura/pt-br/assuntos/noticias/2021/12/concessoes-renovacoes-antecipadas-e-autorizacoes-formam-os-pilares-do-crescimento-ferroviario-diz-tarcisio>. Acesso em 12 de junho de 2022.
- [11] BRASIL. Decreto-lei nº 5.452, de 1 de maio de 1943. Aprova a consolidação das leis do trabalho. **Lex: coletânea de legislação: edição federal**, São Paulo, v. 7, 1943.
- [12] BRASIL. Pro Trilhos chega a 89 propostas e R\$ 258 bilhões de investimentos projetados em um ano. **Ministério da Infraestrutura**. Brasília, DF, 2021. Disponível em: <https://www.gov.br/pt-br/noticias/transito-e-transportes/2022/10/pro-trilhos-chega-a-89-propostas-e-r-258-bilhoes-de-investimentos-projetados-em-um-ano>. Acesso em 03 de outubro de 2022.
- [13] BRASIL. **Resolução nº 5.831, de 23 de outubro de 2018**. Regulamenta o Estabelecimento, a Revisão e a Apuração das Metas de Produção e das Metas de Segurança das Concessionárias que exploram a Infraestrutura e o Serviço Público de

Transporte Ferroviário de Cargas. Diário Oficial da República Federativa do Brasil. Brasília, DF, 2018. Disponível em: https://www.in.gov.br/materia/-/asset_publisher/Kujrw0TZC2Mb/content/id/47328782/UCEQ1TzKXPYVi6cWuD3q0ksQ. Acesso em 18 de maio de 2022.

- [14] BRASIL. Transporte de cargas pelas ferrovias do país cresceu 30% em março. **Ministério da Infraestrutura**. Brasília, DF, 2021. Disponível em: <https://www.gov.br/infraestrutura/pt-br/assuntos/noticias/2021/5/transporte-de-cargas-pelas-ferrovias-do-pais-cresceu-30-em-marco>. Acesso em 12 de junho de 2022.
- [15] BREIMAN, L. et al. **Classification and Regression Trees**. 1 ed. Londres: CRC Press, 1984.
- [16] BREIMAN, L. Random Forests. **Machine Learning**, v.45, p.5–32, 2001.
- [17] CAPRARA, A. et al. Algorithms for Railway Crew Management **Mathematical Programming**, v.79, 2000.
- [18] CHENG, B; TITTERINGTON, D. Neural Networks: A Review from Statistical Perspective. **Stat. Sci.**, v.9, 1994.
- [19] CNT. Boletim Unificado. **Confederação Nacional dos Transportes**. Brasília, DF, 2022. Disponível em: <https://www.antf.org.br/informacoes-gerais/>, Acesso em 18 de maio de 2022.
- [20] DAWSON, S. Rail crew shortages lead to 35,000 train cancellations. **The Times**. 2019. Disponível em: <https://www.thetimes.co.uk/article/rail-crew-shortages-lead-to-35-000-train-cancellations-bwk05lmp2>. Acesso em 18 de maio de 2022.
- [21] FERNANDES, L. F. D. **Novo Modelo de Gestão para Alocação e Controle de Equipagens - Caso MRS Logística SA**. 2012. Monografia (Especialização em Transporte Ferroviário de Carga) - Instituto Militar de Engenharia, Rio de Janeiro, 2012.
- [22] FUZETO, P. U.; LOPES, P. A. S.; CASTRO, C. D. **Método Eficiente de Dimensionamento de Tripulações para Processos com Baixa Frequência de Trens e Transit Time Irregular**. 2017. Monografia (Especialização em Transporte Ferroviário de Carga) - Instituto Militar de Engenharia, Rio de Janeiro, 2017.
- [23] GALWAY, L.; CHARLES, D.; BLACK, M. Machine learning in digital games: A survey. **Artif. Intell. Rev.**, v.29, p.123-161, 2008.
- [24] GATTERMANN-ITSCHERT, T.; PORESCHACK, L.; THONEMANN, U. Using Machine Learning to Include Planners' Preferences in Crew Scheduling Optimization. **Transportation Science**, 2022.
- [25] GÉRON, A. **Mãos à Obra: Aprendizado de Máquina com Scikit-Learn, Keras & TensorFlow**. 2 ed. Alta Books, 2019.
- [26] GOODFELLOW, I. J. et al. Generative Adversarial Networks. **Advances in Neural Information Processing Systems**, v.03, 2014.

- [27] GOODFELLOW, I. J.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. 1 ed. Cambridge: MIT Press, 2016.
- [28] GOPALAKRISHNAN, B.; JOHNSON, E. Airline Crew Scheduling: State-of-the-Art. **Annals of Operations Research**, v.140, p.305-337, 2005.
- [29] HOLLAND, D. Metro driver shortage and train cancellation fears after training axed because of Covid-19 pandemic. **Chronicle Live**. 2020. Disponível em: <https://www.chroniclelive.co.uk/news/north-east-news/metro-driver-shortage-training-covid-18572910>. Acesso em 18 de maio de 2022.
- [30] HOPFIELD, J. Neural Networks and Physical Systems with Emergent Collective Computational Abilities. **Proceedings of the National Academy of Sciences of the United States of America**, v.79, p.2554-2558, 1982.
- [31] HOSSIM, M.; SULAIMAN, M.N. A Review on Evaluation Metrics for Data Classification Evaluations. **International Journal of Data Mining & Knowledge Management Process**, v.5, p.01-11, 2015.
- [32] JIANG, X.; XU, C. Deep Learning and Machine Learning with Grid Search to Predict Later Occurrence of Breast Cancer Metastasis Using Clinical Data. **Journal of clinical medicine**, v.11, 2022.
- [33] JÜTTE, S. **Large-Scale Crew Scheduling: Models, Methods, and Applications in the Railway Industry**. 1 ed. Köln: Springer Gabler, 2019.
- [34] KASIRZADEH, A.; SADDOUNE, M.; SOUMIS, F. Airline crew scheduling: models, algorithms, and data sets. **EURO Journal on Transportation and Logistics**, v.6, n.2, p.111-137, 2017.
- [35] KOMOROWSKI, M. et al. Exploratory Data Analysis. **Secondary Analysis of Electronic Health Records**, p.185-203, 2016.
- [36] KOTSIANTIS, S. B. Supervised Machine Learning: A Review of Classification Techniques. **Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering**, p.3-24, 2007.
- [37] LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v.86, n.11, p.2278-2324, 1998.
- [38] LEGGATE, A. et al. Modelling crew scheduling in offshore supply vessels. **Journal of the Operational Research Society**, v.69, p.959-970, 2018.
- [39] Loh, W. Classification and Regression Trees. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, v.1, p.14-23, 2011.
- [40] MCCULLOCH, W.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, v.5, p.115-133, 1943.
- [41] MITCHELL, T. **Machine Learning**. 1 ed. McGraw-Hill Science, 1997.

- [42] MRS Logística S.A. Demonstração de Resultado de Exercício, 1º Trimestre de 2022. **MRS Logística S.A.** 2022. Disponível em: <https://api.mziq.com/mzfilemanager/v2/d/e14f1680-635b-4328-8a06-b799b1574b43/94b30acd-e9a8-194c-53c8-033c2b589095?origin=1>. Acesso em 21 de junho de 2022.
- [43] MURPHY, K. P. **Machine Learning: A Probabilistic Perspective** 1 ed. MIT Press, 2012.
- [44] NABAIS, R. J. S. **Manual Básico de Engenharia Ferroviária**. 1 ed. São Paulo: Oficina de Textos, 2014.
- [45] NGUYEN, H. et al. Deep learning methods in transportation domain: A review. **IET Intelligent Transport Systems**, v.12, 2018.
- [46] OKABE, A. et al. Spatial Tessellations: Concepts and Applications of Voronoi Diagrams. **Wiley Series in Probability and Mathematical Statistics**, v.43, 2000.
- [47] PARKINSON, J. Why do crew shortages cause one in 10 train delays? **BBC**. 2016. Disponível em: <https://www.bbc.com/news/magazine-35293307>. Acesso em 18 de maio de 2022.
- [48] PEDREGOSA, F. et al. Scikit-learn: machine learning in Python. **Journal of Machine Learning Research**, v.12, p.2825–2830, 2011.
- [49] PEREIRA, V. B. **Transportes: História, Crises e Caminhos**. 1 ed. Rio de Janeiro: Civilização Brasileira, 2014.
- [50] POTDAR, K.; PARDAWALA, T.; PAI, C. A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers. **International Journal of Computer Applications**, v.175, p.7-9, 2017.
- [51] PRABU, S. et al. Grid Search for Predicting Coronary Heart Disease by Tuning Hyper-Parameters. **Computer Systems Science and Engineering**, v.43, p.737-749, 2022.
- [52] RANA, K.; KAUR, P.; MOHAMMED, S. Review on Machine Learning Based Algorithms Used in Autonomous Cars. **SSRN Electronic Journal**, v.5, p.114-118, 2018.
- [53] RODRIGUES, R.S. **Alocação Eficiente da Tripulação nos Trens da MRS Logística**. 2015. Monografia (Especialização em Transporte Ferroviário de Carga) - Instituto Militar de Engenharia, Rio de Janeiro, 2015.
- [54] ROKACH, L.; MAIMON, O. Decision Trees. In: MAIMON, O.; ROKACH, L. (eds) **Data Mining and Knowledge Discovery Handbook.**, Boston: Springer, 2005.
- [55] ROSA, R. A. **Operação Ferroviária: Planejamento, Dimensionamento e Acompanhamento**. 1 ed. São Paulo: LTC, 2016.
- [56] ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, v.65, n.6, p. 386-408, 1958.

- [57] RUMELHART, D. E.; HINTIN, G. E.; WILLIAMS, R. J. Learning Representations by Back-propagating Errors. **Nature**, v.323, p.533-536, 1986.
- [58] SANTOS, M. V. S. **Estudo da Alocação dos Postos de Equipagens - O Caso da MRS Logística SA**. 2007. Monografia (Especialização em Transporte Ferroviário de Carga) - Instituto Militar de Engenharia, Rio de Janeiro, 2007.
- [59] SANTOS, S. **Transporte Ferroviário: História e Técnicas**. 1 ed. São Paulo: Cengage Learning, 2012.
- [60] SETTI, J. B. **Ferrovias no Brasil: um século e meio de evolução**. 1 ed. Rio de Janeiro: Memória do Trem, 2008.
- [61] SHAHRIAR, S. GAN computers generate arts? A survey on visual arts, music, and literary text generation using generative adversarial network. **Displays**, v.73, 2022.
- [62] SIDEY-GIBBONS, J.; SIDEY-GIBBONS, C. Machine Learning in Medicine: a Practical Introduction. **BMC Medical Research Methodology**, v.19, 2019.
- [63] SIL, R. et al. Artificial Intelligence and Machine Learning based Legal Application: The State-of-the-Art and Future Research Trends. **International Conference on Computing, Communication, and Intelligent Systems**, India, p.57-62, 2019.
- [64] SILVA, G.; REIS, A. A study of different metaheuristics to solve the urban transit crew scheduling problem. **Journal of Transport Literature**, v.8, p.227-251, 2014.
- [65] SILVA, N. R. **Eficiência de Cenários de Transporte de Maquinistas para Redução de Custos**. 2013. Monografia (Graduação em Administração) - Universidade Federal Fluminense, Volta Redonda, 2013.
- [66] TANG, R. et al. A Literature Review of Artificial Intelligence Applications in Railway Systems. **Transportation Research Part C Emerging Technologies**, 2022.
- [67] TELLES, P. C. S. **História da Engenharia Ferroviária no Brasil**. 1 ed. Rio de Janeiro: Notícia & Cia, 2011.
- [68] TRINCHINATO, L. F. **Planejamento Operacional de Trens de Carga Geral no Interior de SP em Busca da Excelência**. 2017. Monografia (Especialização em Transporte Ferroviário de Carga) - Instituto Militar de Engenharia, Rio de Janeiro, 2017.
- [69] WAN, X. Influence of feature scaling on convergence of gradient iterative algorithm. **Journal of Physics: Conference Series**, v.1213, p.1-5, 2019.
- [70] WELKOWITZ, J.; COHEN, B. H.; LEA, R. B. **Introductory Statistics for the Behavioral Sciences**. 7 ed. Nova York: Wiley, 2021.
- [71] WILCOXON, F. Individual Comparisons by Ranking Methods. **Biometrics Bulletin**, v.1, n.6, p.80-83, 1945.
- [72] YAAKOUBI, Y.; SOUMIS, F.; LACOSTE-JULIEN, S. Machine Learning in Airline Crew Pairing to Construct Initial Clusters for Dynamic Constraint Aggregation. **EURO Journal on Transportation and Logistics**, v.9, n.4, p.227-251, 2020.

- [73] YU, T.; ZHU, H. Hyper-Parameter Optimization: A Review of Algorithms and Applications. **ArXiv**, v. abs/2003.05689, 2020.