



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE RECURSOS
NATURAIS DA AMAZÔNIA – PRODERNA

GABRIELLE DOS ANJOS CURCINO

CUSTO DE OPORTUNIDADE (TRADE-OFF) PARA DIFERENTES
ESTRATÉGIAS DE MANUTENÇÃO DE TRILHOS FERROVIÁRIOS NA
AMAZÔNIA

Belém-PA

2022



PRODERNA

Programa de Pós-Graduação em Engenharia
de Recursos Naturais da Amazônia



Instituto de Tecnologia

CUSTO DE OPORTUNIDADE (TRADE-OFF) PARA DIFERENTES ESTRATÉGIAS DE MANUTENÇÃO DE TRILHOS FERROVIÁRIOS NA AMAZÔNIA

GABRIELLE DOS ANJOS CURCINO

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia de Recursos Naturais do Instituto de Tecnologia da Universidade Federal do Pará, como requisito necessário à obtenção do título de Doutor em Engenharia de Recursos Naturais.

Orientador: Prof^o. Dr. Eduardo de Magalhães Braga

Belém-PA

2022



PRODERNA

Programa de Pós-Graduação em Engenharia
de Recursos Naturais da Amazônia



Instituto de Tecnologia

**“CUSTO DE OPORTUNIDADE (TRADE-OFF) PARA DIFERENTES
ESTRATÉGIAS DE MANUTENÇÃO DE TRILHOS FERROVIÁRIOS NA
AMAZÔNIA”**

Gabrielle dos Anjos Curcino

TESE DE DOUTORADO APRESENTADA AO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE RECURSOS NATURAIS DA AMAZÔNIA, ITEC, DA UNIVERSIDADE FEDERAL DO PARÁ, COMO PARTE DOS REQUISITOS NECESSÁRIOS À OBTENÇÃO DO TÍTULO DE DOUTOR EM ENGENHARIA DE RECURSOS NATURAIS.

Aprovada por:

Prof. Dr. Eduardo de Magalhães Braga
(Orientador - PRODERNA/UFPA)

Prof. Dr. Nélio Teixeira Machado
(Membro Interno – PRODERNA/UFPA)

Prof. Dr. Marcelo Costa Santos
(Membro Interno – PPGEQ/ UFPA)

Prof. Dr. Douglas Alberto Rocha de Castro
(Membro Externo à Instituição – CEULM/ULBRA)

Prof. Dr. Denis Carlos Lima Costa
(Membro Externo à Instituição – IFPA)

**Belém, PA – Brasil
Dezembro de 2022**

" O Universo é o limite."

Gabrielle Curcino.

DEDICATÓRIA

*A toda a minha família, em especial aos meus pais
Edivaldo e Lídia pelo amor incondicional, aos meus
irmãos e sobrinhos e ao meu amado Marcelo Campi por
todo incentivo.*

AGRADECIMENTOS

Agradeço primeiramente, a Deus (a Ele toda Glória para sempre!) que me levantou, me curou e me trouxe a grande oportunidade de recomeçar e realizar grandes conquistas em minha vida. Agradeço também aos meus pais, Edivaldo Curcino e Lídia Curcino que me incentivaram de todas as maneiras e sempre me apoiaram em todos os momentos. Agradeço a toda minha família, em especial meus irmãos, Mônica, Júnior e Alessandra, assim como meus sobrinhos Sarah, Neto, Lídia Evelyn, Gustavo e Alexsander e cunhados Elizete Moraes e Sandro Santos. Agradeço ao meu amado Marcelo Campi Lima, que apareceu em um momento ímpar em minha vida, onde me incentivou e contribuiu para este trabalho, e que tem a minha genuína admiração. Aos amigos, Sheila (Recife), Giscard (Campinas), Maurício Maia (IFPA), Lorrana (aluna), Karla (Natal), Marcos Ferrarini (Pato Branco) e a Família Ministério Fonte de Vida, minha eterna gratidão pelo carinho e por serem a fiel torcida na minha recuperação e sucesso.

Aos Orientadores responsáveis pela elaboração do trabalho teórico, tratamento e análise dos dados, ao Professor Dr. Eduardo de Magalhães Braga, Professor Dr. André Luis Montenegro e ao Dr. Paulo D'Ângelo Costa Assunção.

Agradeço a Universidade Federal do Pará (UFPA) *Campus Belém*, pela oportunidade de ingresso de estudo a nível superior e todas as perspectivas para a continuação na área acadêmica. Agradeço ao Programa de Pós-Graduação em Engenharia de Recursos Naturais da Amazônia (PRODERRNA) por ajudar no desenvolvimento das pesquisas que foram propostas no desenvolvimento do trabalho. Também agradeço a banca por aceitarem o convite para avaliação deste trabalho.

Resumo da Tese apresentada ao PRODERNA/UFPA para obtenção do grau de Doutor em Engenharia de Recursos Naturais. (Dr. Eng.)

**CUSTO DE OPORTUNIDADE (TRADE-OFF) PARA DIFERENTES
ESTRATÉGIAS DE MANUTENÇÃO EM TRILHOS FERROVIÁRIOS NA
AMAZÔNIA**

Gabrielle dos Anjos Curcino

Orientadora: Prof. Dr. Eduardo de Magalhães Braga

Área de Concentração: Engenharia de Processos

A manutenção emergencial dos ativos ferroviários na Amazônia brasileira tem gerado perdas de receita e custos de oportunidade. O objetivo geral deste estudo foi identificar a importância do custo de oportunidade na tomada de decisão para estratégias de manutenção corretiva e preventiva. A metodologia propôs a modelagem das variáveis referentes aos dados econômicos e operacionais da manutenção ferroviária nos últimos dez anos, por aprendizado de máquina não paramétrico *Gradient Boosting Regression Tree*, e hibridizando-o com a análise do custo de oportunidade para o trade-off decisão de uma ferrovia de minério na Amazônia brasileira. Os resultados mostraram que o GBDT foi eficiente em ajustar os dados de treinamento com r_2 igual a um. Da mesma forma, os dados do teste apresentaram valores de r_2 satisfatórios, próximos a um, onde se obteve o grau de importância das variáveis independentes na previsão das variáveis dependentes. O método de Pearson foi utilizado para construir a matriz de correlação para cada par de variáveis. A partir do modelo gerado, foram criados oito grupos de previsão para o ano de 2022. Em seguida, foram estabelecidos níveis de conflito, sugeridos pela literatura econômica, entre os cenários de previsão, onde o custo de oportunidade foi identificado entre as alternativas com melhor benefício às estratégias de manutenção. Dessa forma, o custo de oportunidade aliado ao aprendizado de máquina serve como um instrumento para auxiliar as empresas na busca por melhores decisões de manutenção, o que contribui para o aprimoramento da gestão dos ativos ferroviários. O estudo do custo de oportunidade a partir de uma análise *trade off* baseada em previsões do custo das estratégias de manutenção através da modelagem por *Machine Learning - Gradient Boosting Regression* é um objetivo inédito na literatura.

Palavras-chave: *machine learning*; gestão de manutenção; custo-benefício; *gradient boosting*; tomada de decisão.

Summary of the Thesis presented to PRODERNA/UFPA as part of the necessary requirements to obtain the degree of Doctor in Natural Resources Engineering. (Dr. Eng.)

OPPORTUNITY COST (TRADE-OFF) FOR DIFFERENT MAINTENANCE STRATEGIES AND RAILWAY TRACKS IN THE AMAZON

Gabrielle dos Anjos Curcino

Advisor: Prof. Dr. Eduardo de Magalhães Braga

Concentration area: Process Engineering

The emergency maintenance of railway assets in the Brazilian Amazon has generated revenue losses and opportunity costs. The general objective of this study was to identify the importance of opportunity cost in decision-making for corrective and preventive maintenance strategies. The methodology proposed the modeling of the variables referring to the economic and operational data of railway maintenance in the last ten years, by non-parametric Gradient Boosting Regression Tree machine learning, and hybridizing it with the analysis of the opportunity cost for the trade-off decision-making of an ore railroad in the Brazilian Amazon. The results showed that the GBDT was efficient in fitting the training data with r_2 equal to one. Similarly, the test data presented satisfactory r_2 values, close to one, where the degree of importance of the independent variables in the prediction of the dependent variables was obtained. Pearson's method was used to construct the correlation matrix for each pair of variables. From the generated model, eight forecast groups were created for the year 2022. Then, conflict levels were established, suggested by the economic literature, between the forecast scenarios, where the opportunity cost was identified among the alternatives with the best benefit to maintenance strategies. In this way, the opportunity cost combined with machine learning serves as an instrument to help companies in the search for better maintenance decisions, which contributes to the improvement of rail asset management.

Keywords: maintenance strategies; rails; opportunity cost; machine learning; scikit-learn; railroads.

SUMÁRIO

SUMÁRIO.....	ix
LISTA DE FIGURAS.....	xi
LISTA DE QUADROS E TABELAS.....	xix
NOMENCLATURAS	xiii
CAPÍTULO 1- INTRODUÇÃO.....	14
1.1 - ASPECTOS GERAIS, JUSTIFICATIVA E MOTIVAÇÃO.....	14
1.2 - INEDITISMO E CONTRIBUIÇÃO DA TESE.....	17
1.3 - OBJETIVOS GERAL E ESPECÍFICOS.....	18
1.4 - ORGANIZAÇÃO DO TRABALHO.....	18
CAPÍTULO 2 - REVISÃO BIBLIOGRÁFICA.....	19
2.1 - SETOR DE TRANSPORTE.....	19
2.1.1 - Transporte Ferroviário.....	19
2.1.2 - Ativos Ferroviários: Trilhos.....	22
2.1.3 - Caracterização dos defeitos em trilhos.....	24
2.1.4 - Tensões residuais: causas e ocorrências.....	26
2.2 - MANUTENÇÃO EM TRILHOS FERROVIÁRIOS.....	28
2.2.1 - Definições e conceitos.....	28
2.2.2 - Estratégias de manutenção.....	30
2.3 - CUSTO DE MANUTENÇÃO FERROVIÁRIA.....	35
2.4 - ANÁLISE CUSTO-BENEFÍCIO E CUSTO DE OPORTUNIDADE.....	39
2.4.1 - Custo de oportunidade e a análise de decisão (trade off)	43
2.5 - MECANISMO DE ANÁLISE PARA UMA TOMADA DECISÃO TRADE-OFF...	45
2.5.1 – <i>Machine Learning</i>	45
CAPÍTULO 3 – MATERIAIS E MÉTODOS.....	49
3.1 - MATERIAIS.....	49
3.1.1 – Objeto de Estudo.....	49
3.1.2 – Delimitação do Estudo.....	50
3.2 – MÉTODOS.....	52
3.2.1 - Análise do banco de dados: variáveis (input and output)	52
3.2.2 - Modelagem e Método <i>Gradient Boosting Regression</i>	54
CAPÍTULO 4 – RESULTADOS E DISCUSSÕES.....	58
4.1 - APLICAÇÃO DO APRENDIZADO DE MÁQUINA.....	58

4.1.1 - A escolha do Modelo.....	58
4.1.2 – Características das variáveis.....	59
4.1.3 – Funções de aprendizado aplicadas as <i>splines</i>	60
4.1.4 – Nível de Entropia e Ganho de Informação.....	61
4.1.5 – O método do Gradiente: considerações gerais.....	61
4.1.6 – Restrições na execução e a Função de perda.....	62
4.1.7 – Estimação do erro e Acurácia.....	63
4.1.8 – Correlação de Pearson para as variáveis.....	63
4.1.9 – Análise <i>Trade-Off</i> ou Tomada de Decisão.....	69
CAPÍTULO 5 – CONCLUSÕES E RECOMENDAÇÕES.....	74
REFERÊNCIAS BIBLIOGRÁFICAS.....	76
APÊNDICE	82
ANEXO.....	90

LISTA DE FIGURAS

Figura 2.1 - Produção do Transporte Ferroviário (Bilhões de TKU)	19
Figura 2.2 - Matriz de Transporte Global 2017.....	20
Figura 2.3 - Componentes do trilho Vale S.A, 2009	23
Figura 2.4 - Distribuição de tensões em um componente com uma solda de topo submetido a carregamentos crescentes (curvas 1,2, 3) e distribuição de tensões residuais após a liberação do carregamento.....	27
Figura 2.5 - Estratégias de manutenção e a gestão de ativos.....	29
Figura 2.6 - Manutenção Centrada na Confiabilidade.....	30
Figura 2.7 - Pontos de Falhas.....	31
Figura 2.8 – Processo de otimização - <i>Gradient Boosted Decision Trees</i> (GBDT).....	47
Figura 3.1 - Rede Ferroviária na Amazônia Oriental.....	50
Figura 3.2 - Fluxograma demonstrativo da delimitação do estudo.....	51
Figura 3.3 - Fluxograma da Metodologia de Pesquisa.....	54
Figura 3.4 - Diagrama do modelo <i>Gradient Boosting Regression</i>	56
Figura 3.5 - Dados de treino e teste: Coeficiente de Determinação dos grupos de treino e de teste para o custo de manutenção corretiva (a), preventiva (b) e lucro cessante (c).....	57
Figura 4.1 - Variáveis analisadas na matriz de correlação.....	58
Figura 4.2 - Grau de importância das variáveis para a previsão dos custos de manutenção corretiva (a), preventiva (b) e lucro cessante(c).....	59
Figura 4.3 - Grupos de previsão e cenários construídos para a Manutenção Corretiva.....	61
Figura 4.4 - Grupos de previsão e cenários construídos para a Manutenção Preventiva.....	62

LISTA DE QUADROS

Quadro 2.1 - Tipos de manutenção segundo a Norma NBR 5462.....	32
Quadro 2.2 - “Types of Maintenance Toolk”.....	33
Quadro 2.3 - Considerações entre o custo e benefício.....	41
Quadro 2.4 – Diferença entre custo-benefício e custo efetividade.....	42
Quadro 2.5 – Principais artigos embaixadores para o estudo de aprendizado de máquina e análise <i>trade-off</i>	48

LISTA DE TABELAS

Tabela 2.1 - Especificação do tipo de defeito.....	25
Tabela 3.1 - Dados operacionais e de produtividade, econômico-financeiros e custos de manutenção.....	53
Tabela 4.1: Previsão a curto prazo para os Custos de Manutenção Preventiva e Corretiva.....	66
Tabela 4.2 - Nível de conflito para estabelecer comparações para a Estratégia de Manutenção Corretiva.....	70
Tabela 4.3 - Nível de conflito para estabelecer comparações para a Estratégia de Manutenção Preventiva.....	71
Tabela 4.4 - Seleção do Custo de Oportunidade paras as Estratégias de Manutenção.....	71
Tabela 4.5 - Tomada de Decisão do Custo de Oportunidade paras as Estratégias de Manutenção, simultaneamente.....	73

NOMENCLATURAS

ACB	Análise Custo-Benefício
ANTF	Agência Nacional de Transporte Ferroviário
ANTP	Agência Nacional de Transporte e Pesquisa
ANTT	Agência Nacional de Transporte Terrestres
CMC	Custo de Manutenção Corretiva
CNT	Conselho Nacional de Transporte
CPS	Custo Pessoal e Serviço
DNIT	Departamento Nacional de Infraestrutura e Logística
EFC	Estrada de Ferro Carajás
EPL	Empresa de Planejamento em Logística
GBR	<i>Gradient Boosting Regression</i>
IPEA	Instituto de Pesquisa Econômica Aplicada
MCC	Manutenção Centrada em Confiabilidade
MI	Ministério da Infraestrutura
MRS	Malha Regional Sudeste
MTBF	<i>Mean Time Between Failures</i>
MTTR	<i>Mean Time To Repair</i>
NMS/ NPM	Número de paradas
NBR	Norma Brasileira
OPEX	<i>Operational Expenditure</i>
OP/L /LC	Lucro Cessante
PF	Ponto de Falha
PMC	Custo de Manutenção Preventiva
P. S /C.P. S	Pessoal de serviço
REFER	Rede Ferroviária Nacional
SILCON	Consultoria em Estudos Econômicos
TKU	Tonelada por quilômetro útil
TU/TCT	Total de carga transportada
UFPA	Universidade Federal do Pará
VALE	Empresa de Mineração S. A
ZTA	Zona Termicamente Tratada

CAPÍTULO 1

INTRODUÇÃO

1.1 – ASPECTOS GERAIS, JUSTIFICATIVAS E MOTIVAÇÃO

Os centros mineradores na região Amazônica utilizam substancialmente o transporte ferroviário para a exportação de commodity, demanda muito requisitada no mercado internacional. Entende-se que o desenvolvimento econômico e social impulsiona a infraestrutura ferroviária a aumentar o nível de qualidade, que por sua vez, requer tomadas de decisões estratégicas de manutenção cada vez mais atreladas ao custo-benefício potencial. Como afirma Xavier (2008) quanto mais se produz, mais premente se torna à necessidade de manutenção e menor as oportunidades em executá-la. No entanto, a dificuldade destas mineradoras está em manter o melhor custo-benefício entre a manutenção e a produtividade, principalmente quando persistem os registros de paradas operacionais, acendendo um sinal de alerta para os gestores de manutenção, uma vez que geram custos não programados.

Face às exigências crescentes dos agentes econômicos, o governo brasileiro concedeu a concessão de serviços ferroviários à iniciativa privada, a qual proporcionou maiores investimentos na aquisição de novas tecnologias, capacitação profissional, qualificação das operações e recuperação da malha. Como apontam os dados da Agência Nacional de Transporte Ferroviário – ANTF, em 2022 as ferrovias associadas à agência investiram até 2021 mais de R\$ 85 bilhões (valores correntes), que representam mais de R\$ 141,9 bilhões se atualizados pelo IPCA.

De acordo com o Observatório Nacional de Transporte e Logística – ONTL da EPL (2017), a malha ferroviária brasileira é cerca de 21.244 quilômetros com movimentação de carga e 193 terminais ativos. Enquanto a rede ferroviária na região norte, construída para atender projetos específicos de exploração mineral, é de aproximadamente 2372 Km (inclusive trecho operacional da Norte-Sul), o que equivale a cerca de 11% da malha brasileira.

Os dados apresentados no Painel CNT do Transporte – Ferroviário 2022 mostram que a Estrada de Ferro Carajás – EFC lidera com o volume de 171,2 milhões de toneladas quilometro-útil (TKU) de minério de ferro, equivalente a 46,3% do total das cargas transportadas em 2021. Essa crescente demanda de movimentação sobrecarregou os vagões e, conseqüentemente, levando a degradação dos ativos ferroviários, em alguns casos, o desgaste prematuro dos trilhos. Schneider e Reguly (2014) consideram uma das maiores dessas conseqüências, a ocorrência de

defeitos internos nos trilhos que desafia a operação de mais trens circulando diante de menos janelas de tempo para a realização das inspeções e reparo de defeitos. Souza (2008) também verificou que a incidência contínua de avarias leva às despesas onerosas com ações corretivas, no entanto, um planejamento de manutenção de forma equivocada pode levar a empresa a despender dinheiro com intervenções preventivas desnecessárias.

Os problemas recorrentes encontrados em uma ferrovia de uma concessionária atuante na Região Amazônica, mostram a formação de trincas longitudinal na ZTA - Zona Termicamente Tratada e/ou zona de transição da junta soldada realizada por centelhamento dos trilhos ferroviários, como explica Braga (2018) a formação destas trincas, que aparecem após um determinado tempo de uso da via, ainda não é bem entendida do ponto de vista processual e metalúrgico.

Diante deste cenário, a manutenção de ativos ferroviários, em especial para a correção de falhas provenientes de tensões residuais nos trilhos das ferrovias na Amazônia Brasileira, tem utilizado estratégias de manutenção de emergência, apresentando uma eficiência muito baixa com tempo adicional desperdiçado e custos maiores são incorridos para que isso aconteça, porém não mensurados. Este hiato na gestão de manutenção é explicado pelo desconhecimento do real impacto que o custo gera no processo, como perdas de receita e de custos de oportunidades nas atividades, como explica Talaszka, (2018) apud Medeiros (2005):” o estudo do comportamento dos custos é relevante, haja vista que a base de sustentação de muitas decisões gerenciais está no conhecimento de como os custos podem variar em função do nível de atividade”.

As literaturas existentes sobre os conceitos e definições do custo de oportunidade, trazem interpretações contábeis quantitativas e qualitativas. Leone (1982): “benefício que se deixa de ganhar”; Martins (1987): “quanto à empresa sacrificou em termos de remuneração”. Gray & Johnston (1977): “lucro que se poderia ter conseguido”. Sizer (1974) ao dizer que custo de oportunidade concerne a melhor alternativa da qual se desistiu, isto é, a própria alternativa é o atributo, o que de certa forma, torna o conceito compacto e deixa a cargo do usuário caracterizar o que lhe é conveniente. Morse & Roth (1986): “o custo de oportunidade de uso de um recurso em um modo específico é o recebimento líquido de caixa esperado que poderia ser obtido se o recurso fosse usado na outra ação alternativa mais desejável”; Glautier & Underdown (1986) “o custo de oportunidade pode ser medido como o valor da próxima melhor alternativa abandonada, ou o recebimento líquido de caixa permitido como resultado de preferir uma alternativa ao invés da melhor seguinte.

Na busca por trabalhos anteriores que estudaram o custo de oportunidade como ferramenta de decisão, a pesquisa e análise são relativamente baixas ou quando estudadas individualmente. Pereira (2018) faz uma abordagem econômica do custo de oportunidade, onde o conflito de escolha é conhecido como *trade-off*. Castello Branco (2008) considera o custo marginal como o custo de oportunidade, ou seja, o valor dos recursos adicionais empregados, se estes vierem a ser utilizados numa atividade alternativa. Van Jaarsveld et al., (2013) determinam os custos de oportunidade logística do uso do transporte rodoviário dentro de uma mineradora. Kussano e Batalha (2012) consideram o custo de oportunidade como variável em um modelo de estrutura de custos logísticos do escoamento da safra de soja. Budai-Balke (2009) considera a otimização de manutenção clássica, onde o objetivo é encontrar um trade-off entre manutenção preventiva e corretiva. Sousa et al., (2018) apresentam uma metodologia útil como ferramenta de gestão de ativos, onde aplicaram um modelo linear que disponibiliza uma gama de soluções para análise de *trade-offs* pelo decisor. Por fim, Van Essen et al (2004) explicam que em geral, a avaliação dos custos de oportunidade no setor ferroviário e aéreo continua a ser uma abordagem difícil e pouco pesquisada.

Todavia, quando se trata de implementação dos mecanismos de análise para uma tomada de decisão, o uso de Inteligência Artificial como auxílio já é uma realidade nas empresas. Uma pesquisa elaborada pela empresa Gartner Resecher Circle, realizada em 2017, aponta que o uso de Inteligência Artificial nos últimos quatro anos, foi de 270%. Para 2022, esse número está previsto crescer ainda mais, chegando a 80% dos avanços relacionados ao *Machine Learnig* e à Inteligência Artificial.

A utilização do aprendizado de máquina disponibiliza informações mais claras e cria um terreno fértil para o apontamento de ideias e tomada de decisão. Como apontam Wang e Mamo (2020) “o desafio está em desenvolver um modelo preditivo preciso para um sistema de gerenciamento de bateria”. Para garantir o bom desempenho dos sistemas de micro irrigação, García Nieto et al. (2018) determinam as perdas de pressão produzidas a partir da abordagem de aprendizado de máquina não paramétrica conhecida como árvore de regressão impulsionada por gradiente (GBRT). Esse método é usado para tarefas de classificação, mas existe a possibilidade de aplicá-lo para uma tarefa de regressão (Baturynska e Martinsen, 2020).

Estas pesquisas reforçam a importância em desenvolver estudos do custo de oportunidade no campo da gestão de manutenção ferroviária. Tomando o conhecimento da correlação entre as variáveis econômicas e operacionais em uma série temporal que permitam a identificação do custo de oportunidade nas previsões e gerenciá-los.

O propósito deste trabalho consiste na aplicação realista do conceito de custo em um processo de tomada de decisão (trade-off) frente as estratégias de manutenção, isto é, o custo de oportunidade.

Para a questão de pesquisa, têm-se o questionamento: É possível fazer previsões como suporte para o estudo do custo de oportunidade nas tomadas de decisões gerenciais de manutenção? Frente a este questionamento, se fez necessário desenvolver uma abordagem de predição dos custos das estratégias de manutenção, além de investigar o custo de oportunidade nas decisões de manutenção. Foi utilizado o aprendizado da máquina na modelagem de variáveis quantitativas por *Machine Learning* através do *Gradient Boosting Regression* utilizando *scikit-learn*. O estudo está concentrado em uma Estrada de Ferro, no sudeste do estado do Pará, localizada na Amazônia Oriental Brasileira, a maior reserva mineral do mundo.

Apesar de conceitos econômicos tradicionais e das aplicabilidades do custo de oportunidade serem norteadores para esta pesquisa, um estudo simples foi encontrado, onde a predição de dados operacionais e financeiros de manutenção servem para a análise do custo de oportunidade *trade-off* em estratégias de manutenção ferroviária. Para tanto, é necessária uma análise detalhada das compensações do custo de oportunidade para determinar a tomada de decisão (trade-off) mais adequada.

1.2 - INEDITISMO E CONTRIBUIÇÃO DA TESE

O trabalho abordará o estudo do custo de oportunidade a partir de uma análise *trade off* baseada em predições do custo das estratégias de manutenção através da modelagem por *Machine Learning - Gradient Boosting Regression*. Foi investigada a correlação das variáveis operacionais e econômicas e a previsão do desempenho dos custos de manutenção preventiva e corretiva frente ao custo de oportunidade.

Não foram encontrados na literatura informações a respeito do custo de oportunidade como ferramenta de decisão gerencial em manutenção de trilhos ferroviários o que indica que a presente tese é inédita. Além de que a previsão para as estratégias de manutenção através do uso de aprendizado de máquina foi realizada originalmente pela primeira vez.

Neste sentido este trabalho pretende contribuir de forma inédita com a análise de decisão para a gestão de manutenção ferroviária, avaliando a correlação dos dados operacionais e financeiros de cenários anteriores para a identificação do custo de oportunidade na predição a curto prazo.

1.3 – OBJETIVOS: GERAL E ESPECÍFICOS

Este trabalho teve como **objetivo geral** estudar a aplicação do conceito de custo de oportunidade em um processo de tomada de decisão (*trade-off*) nas estratégias de manutenção de trilhos em uma estrada de ferro na Região Amazônica.

Objetivos específicos:

- levantar dados operacionais e custos financeiros de manutenção;
- identificar as estratégias de manutenção utilizadas;
- desenvolver uma abordagem de predição dos custos das estratégias de manutenção a curto prazo;
- investigar o custo de oportunidade nas decisões de manutenção a partir de uma análise *trade-off*; e
- subsidiar as decisões na gestão de manutenção de trilhos ferroviários;

1.4 - ORGANIZAÇÃO DO TRABALHO

No capítulo 1 buscou-se enfatizar as motivações, os objetivos e o ineditismo e contribuição da tese.

No Capítulo 2 é apresentada a revisão da literatura e o estado da arte acerca do sistema ferroviário, as estratégias de manutenção, custo de manutenção e custo de oportunidade, com ênfase nos ativos ferroviários, especificamente os trilhos. Além do mecanismo de análise utilizando o aprendizado de máquina e *trade-off*.

O Capítulo 3 trata dos materiais e métodos onde o objeto de estudo e sua delimitação serviram para a classificação das variáveis e aplicação dos métodos. E a abordagem metodológica traz a aplicabilidade do aprendizado de máquina e a análise *trade-off*.

Resultados e suas discussões são apresentados no Capítulo 4, incluindo a correlação das variáveis dependentes e a influência na criação de cenários preditivos a curto prazo.

No Capítulo 5, são apresentadas as principais conclusões e recomendações.

CAPÍTULO 2

REVISÃO BIBLIOGRÁFICA

2.1 - SETOR DE TRANSPORTE

2.1.1 - Transporte Ferroviário

Segundo Agência Nacional de Transporte Ferroviário (2020) a concessão no setor também foi primordial para que as ferrovias brasileiras ganhassem eficiência frente ao crescimento da produção ferroviária, que é calculada pela medida que indica o número de toneladas de carga movimentada a cada quilômetro. Em 2019, foram 366 bilhões de TKU (toneladas por quilômetro útil), uma redução — a primeira em 23 anos — de 10% sobre o resultado de 2018 (407 bilhões de TKU).

Desde o início das concessões, contudo, o crescimento em TKU foi de 167%, como mostra Figura 2.1. Isso foi possível, com o apoio de duas frentes de trabalho implementadas nas gestões operacionais e de manutenção, como a recuperação de trechos, com melhoria de vias antigas e de baixa performance, e a construção de novas vias ferroviárias, modernas e eficientes.

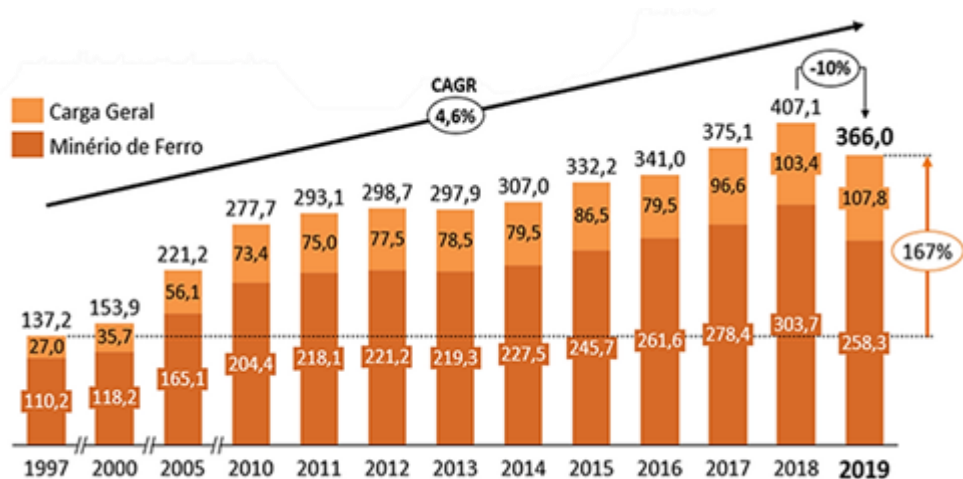


Figura 2.1 - Produção do Transporte Ferroviário (Bilhões de TKU)

Fonte: Agência Nacional de Transporte Ferroviário

De acordo o relatório anual do Projeto infra2038 a densidade ferroviária brasileira de apenas 3,57 quilômetros de ferrovias disponíveis a cada mil quilômetros quadrados de território. Segundo Pezco & LL Advogados (2021) a baixa densidade ferroviária da economia brasileira

está relacionada à extensão do território. Dada a extensão do território, o desafio mais relevante se refere à maximização do uso da infraestrutura existente. Em particular, o planejamento do horizonte próximo deve atentar para a construção e a utilização eficaz dos principais eixos de escoamento de cargas de grande peso no território brasileiro. Notadamente, a densidade ferroviária é maior nos estados das regiões sul e sudeste relativamente aos estados das regiões centro-oeste, norte e nordeste.

O modal ferroviário brasileiro ocupa a segunda melhor posição na matriz de transporte brasileira, possibilitando a intermodalidade logística, fundamental para a retomada do crescimento econômico. O transporte ferroviário de cargas tem representatividade singular na matriz de transporte global, estando presente nas principais economias do mundo, como mostra a Figura 2.2.

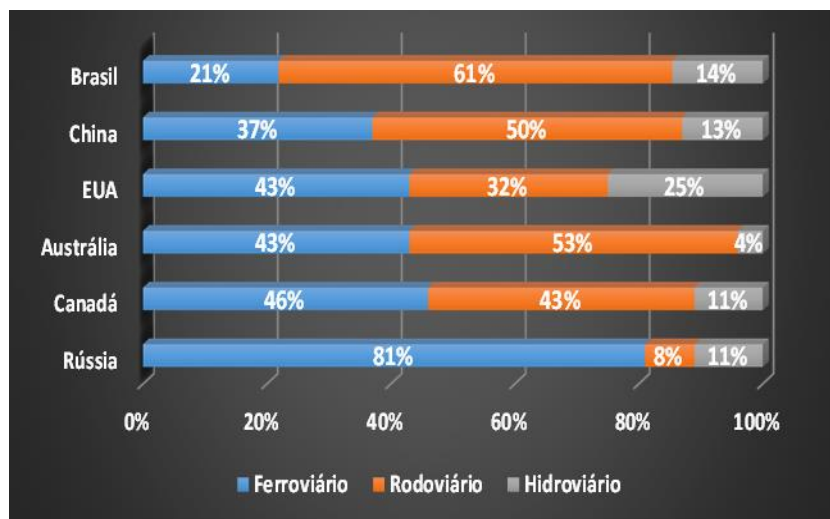


Figura 2.2 - Matriz de Transporte Global. – Ministério Transporte (2017)

Segundo os dados da *Association of American Railroads – AAR* em 2019 as ferrovias norte-americanas com cerca de 140 mil milhas (aproximadamente 225 mil quilômetros) transportaram aproximadamente 1.729.638 milhões de toneladas-milhas, (aproximadamente 1.074.747 milhões de TKU), com destino aos EUA, e cerca de 1.652.970 milhões de toneladas-milhas (aproximadamente 1.027.107 milhões de TKU) com origem nos EUA.

Conforme dados da *Russian Railways*, em 2019 a Rússia movimentou 3.305 bilhões de TKU, onde contam com maior representatividade na matriz de transporte — aproximadamente 80%. Já de acordo com o *China Daily*, as ferrovias chinesas transportaram 3.44 bilhões de toneladas em 2019, um aumento de aproximadamente 8% em relação a 2018 (ANTF, 2020).

De acordo com os dados apresentados no Painel CNT do Transporte – Ferroviário, a maior parte das cargas movimentadas por ferrovias em 2020 tem origem nos estados do Pará (139,6 milhões de TU) e de Minas Gerais (137,0 milhões de TU).

Os produtos que saem do Estado do Pará têm como principal destino o Maranhão; as cargas que partem de Minas Gerais seguem, principalmente, para o Rio de Janeiro e o Espírito Santo. A Estrada de Ferro Carajás - EFC lidera no volume de toneladas úteis transportadas em 2020, com 141 milhões de TU (39,4% do total). Em seguida, vêm a MRS (82,1 milhões de TU, que equivalem a 23% do total) e a EFVM (Estrada de Ferro Vitória a Minas), que transportou 52,8 milhões de TU, 14,8% do total (CNT, 2020).

Apesar de ações pontuais ao longo de décadas de estagnação, o Ministério de Infraestrutura anunciou que o setor ferroviário brasileiro passará por investimentos que garantirão mais espaço na distribuição de insumos e mercadorias dentro do modelo logístico nacional – Plano Nacional de Logística e Transportes, com o intuito de aumentar a capacidade da malha existente nesses 29 mil quilômetros, principalmente os 9 mil não operacionais. Lembra ainda que a velocidade média do transporte de carga por vias ferroviárias é de cerca de 23km/h [quilômetros por hora], que demonstra o primeiro desafio a ser superado para aumentar a eficiência (MI, 2020).

A corrida para implementar o setor ferroviário, frente a demanda do mercado internacional nestes últimos anos, é a resposta para a falta de investimentos às estruturas existentes responsáveis pelos prejuízos econômicos e operacionais. A exemplo, o sobrecarregamento dos vagões versus a degradação dos ativos ferroviários, como em alguns casos, o desgaste prematuro dos ativos cominadores de diversos tipos de falha. Schneider e Reguly (2014) consideram uma das maiores consequências, a ocorrência de defeitos internos nos trilhos, desafiando a operação de mais trens circulando diante de menos janelas de tempo para a realização das inspeções e reparo de defeitos.

Esse esforço por parte das companhias ferroviárias vem ocorrendo em New York - USA, Vaidya (2017) observou que as organizações se esforçam para reduzir os custos, tornando suas operações e manutenção tão eficientes e eficazes quanto possível, agora com mais previsibilidade e resiliência de seus equipamentos ativos. Para atender a essas altas expectativas, as organizações estão recorrendo a uma reforma completa de suas estratégias de manutenção de equipamentos.

No Brasil, Souza (2008) também verificou que a incidência contínua de avarias leva a empresa a arcar com onerosas despesas com ações corretivas. No entanto, um planejamento de manutenção de forma equivocada pode levar a empresa a despender dinheiro com intervenções

preventivas desnecessárias. A exemplo, na MRS Logística S/A, concessionária que opera a chamada Malha Regional Sudeste da Rede Ferroviária Federal S.A, onde o negócio da organização depende diretamente da disponibilidade dos ativos, a responsabilidade do setor de manutenção tem magnitude substancial, onde seu desempenho tem relação direta com as metas e objetivos de produtividade.

2.1.2 - Ativos Ferroviários: Trilhos

Os ativos ferroviários têm significativo impacto econômico no desempenho do sistema, isto é, são responsáveis pela quantificação da eficiência e da eficácia de uma organização. Procaci (2018) complementa que os ativos ferroviários são aperfeiçoados de acordo com a necessidade cada vez maior de inovações para o desenvolvimento industrial. Por este motivo, Neely (2015) explica que a correta medição e avaliação do desempenho dos processos e do sistema produtivo em geral constituem a forma de garantir os objetivos estratégicos da empresa, qualidade, tempo, flexibilidade e custo.

Procaci (2018) sugere a divisão dos ativos ferroviários em dois grandes grupos, denominados Material Rodante (locomotivas, vagões, equipamentos de via e produção industrial) e Malha Ferroviária (via permanente, eletroeletrônica, infraestrutura).

Tratando-se especificamente dos ativos da malha ferroviária, Campos (2018) os subdivide em três grupos, são eles: Via permanente (trilhos, dormentes), Eletroeletrônica (sinalizações de via, comunicação com centro de controle) e Infraestrutura (pontes e túneis). A funcionalidade destes ativos vai além da sustentação para a circulação dos trens, também permitem o deslocamento dos trens de forma produtiva e segura.

Selig e Waters (2018) e Oliveira (2020) descrevem os trilhos como perfis longitudinais de aço fixados sobre os dormentes com a função de guiar o veículo e suportar as rodas dos trens, além de ter rigidez suficiente para servir como vigas, as quais são responsáveis por transferir a carga concentrada das rodas para os dormentes, que estão dispostos ao longo da via, sem causar deflexão excessiva aos mesmos.

O perfil do trilho amplamente utilizado no Brasil é o tipo Vignole com dimensões e características mecânicas definidas pela norma ABNT NBR 7590. Essa mecânica é distribuída em três partes fundamentais: boleto (cabeça), alma e patim, conforme representado na figura 2.3.



Figura 2.3 - Componentes do trilho, Vale S.A (2009)

O Departamento Nacional de Infraestrutura de Transporte – DNIT (2016) ressalva que para satisfazer as funções adequadas, os trilhos devem possuir, dureza, tenacidade, elasticidade e resistência a flexão. Para isso, se faz necessário estudos dos desgastes, levando em conta as normas técnicas e a tolerância tradicionalmente aceitos, evitando ao máximo a substituição devido ao alto custo envolvido.

Assis (2018) equaciona a seguinte definição, os trilhos são projetados para sofrerem tensões bem menores do que sua tensão de escoamento, desde que se faça uso do fator de segurança. A tensão de projeto é a tensão de ruptura do componente dividido por esse coeficiente, ver Equação (1):

$$\sigma_{adm} = \frac{\sigma_r}{v}$$

σ_{adm} = *tensão admissível*

σ_r = *tensão de ruptura*

v = *coeficiente de segurança*

Desse modo, o componente em serviço é submetido a carregamentos muito abaixo do seu limite de resistência, tornando raros os casos de fratura devido à sobrecarga. Leme (2017) enfatiza a obrigatoriedade em trabalhar abaixo do escoamento e ainda acrescentar um fator de segurança quando se tratar de estruturas de cargas.

Ainda dentro deste argumento, Leme ressalta que existem questionamentos sobre o superdimensionamento de estruturas quando determinado aço apresenta uma tensão de ruptura bem maior que a tensão de projeto. Ocorre que a estrutura pode passar por muitas outras solicitações que, na prática são impossíveis de prever: choques, má utilização, má conservação ou falta de manutenção e principalmente a temida fadiga, que reduz absurdamente a resistência da estrutura e causa ruptura repentina, mesmo em tensões abaixo do escoamento.

Vale ressaltar que devido o advento da alta demanda de carregamento, surgiram veículos ferroviários mais potentes e a composições unitárias longas, conseqüentemente, exigindo o

desenvolvimento de trilhos de maior resistência aos desgastes. Coimbra (2008) explica que esses novos trilhos podem ser classificados em dois grupos, são eles:

- trilhos de aço-liga e trilhos tratados termicamente. Para se obter os do primeiro grupo, ao aço é agregado o manganês, o cromo ou silício que contribuem para o aumento da resistência à ruptura.
- trilhos tratados termicamente apresentam uma composição química bastante próxima dos trilhos de aço carbono endurecidos por tratamentos térmicos adequados.

A durabilidade dos trilhos é de grande importância para as ferrovias, sendo os desgastes dos boletos essenciais para representar a vida útil destes componentes.

É fundamental ponderar as normas técnicas e os limites de tolerância existentes nos estudos dos desgastes que geram defeitos e falhas em trilhos, para que se possa evitar ao máximo a substituição que levantaria um alto custo; e a questão da segurança da via que, dependendo do defeito, pode ser comprometida.

Segundo a Vale S.A dentre a gestão dos ativos ferroviários, o trilho é o mais importante da superestrutura, uma vez que é tecnicamente considerado o principal elemento de suporte e guia dos veículos ferroviários e, economicamente detém o maior custo entre os elementos estruturais da via.

2.1.3 - Caracterização dos defeitos em trilhos

Antes de apresentar os defeitos em trilhos, características e ocorrências, os autores Gomes (2016); Macedo (2009), respectivamente, trazem conceitos e tipologias sobre defeitos e discontinuidades que precisam ser discriminados. Entende-se que discontinuidade é uma interrupção da estrutura típica de um material como a falta de homogeneidade na característica mecânica, metalúrgica ou física, ou seja, não é necessariamente um defeito; enquanto o defeito é a condição ou as condições que tornam uma parte incapaz de satisfazer uma mínima aceitação de padrão ou especificação. Quanto a tipologia, são basicamente dois tipos para defeitos e discontinuidades: de fabricação, compreendendo vazios, segregação, inclusões, fissuras transversais e defeitos de laminação; e originados em serviço.

Neste trabalho será utilizada a terminologia para caracterização da direção de propagação dos defeitos de trilhos presentes no Manual Técnico de Via Permanente da Vale S.A. O manual descreve que a maioria dos defeitos de trilhos requer alguma forma de solicitação para iniciar e se desenvolver. Para identificação dos defeitos deve-se utilizar a seguinte convenção:

- Direção Longitudinal Vertical: desenvolve longitudinalmente ao longo do perfil, no plano vertical;
- Direção Longitudinal Horizontal: desenvolve longitudinalmente ao longo do perfil, no plano horizontal;
- Direção Transversal.

Segundo o Manual a identificação destes defeitos é realizada através de Inspeção, onde a identificação e mapeamento dos defeitos são realizados por priorização, classificados em:

- Defeitos internos: Trinca Longitudinal Horizontal; Trinca Longitudinal Vertical; Trinca Transversal o Bolha ou Vazio; Defeitos nas soldas.
- Defeitos externos: são aqueles visíveis, permitindo o acompanhamento de sua degradação ao longo do tempo.

Na identificação dos defeitos é utilizado equipamento de ultrassom que possui nomenclatura para cada defeito, assim como a classificação do tamanho, expresso em unidades, de acordo com a especificação do tipo de defeito, ver Tabela 2.1.

Tabela 2.1 - Especificação do tipo de defeito

Orientação	Nome	Sigla	Un.	P	M	G
T	Trinca de patinagem de roda	EBF	% boleto	< 15	15-30	> 30
LH	Trinca horizontal no boleto	HSH	mm	< 50	50-100	> 100
LV	Trinca vertical no boleto	VSH	mm	< 50	50-100	> 100
T	Trinca Transversal	TDT	% boleto	< 15	15-30	> 30
T	Trinca de fragmentação	TDD	% boleto	< 15	15-30	> 30
C	Trinca composta	TDC	mm	< 25	25-100	> 100
T	Trinca em solda elétrica	DWP	% boleto	< 15	15-30	> 30
T	Trinca em solda aluminotérmica	DWF	% boleto	< 15	15-30	> 30
C	Trinca na alma	SWO	mm	< 25	25-50	> 50
C	Trinca nos furos na junta	BHJ	mm	< 25	25-50	> 50
C	Trinca nos furos fora da junta	BHO	mm	< 25	25-50	> 50
LV	Trinca vertical na alma em junta	PRJ	mm	< 25	25-100	> 100
LV	Trinca vertical na alma fora da junta	PRO	mm	< 25	25-100	> 100
LH	Trinca no filete	HWS	mm	< 25	25-50	> 50
LH/LV	Inclusão	INC	mm	< 50	50-100	> 100

Fonte: Manual Técnico de Via Permanente – VALE S.A, 2009

A nível de manutenção são consideradas fraturas em situações que houver ruptura total da seção transversal do trilho ou em casos que houver fragmentação da seção com perda de material. Para os demais casos, é considerada a ocorrência de trinca.

De acordo com Alvarez (2019) em sua análise mecânica, os trilhos ferroviários são componentes susceptíveis aos altos níveis de tensões decorrentes da pressão de contato cíclica imposta pela carga vertical distribuída entre as rodas. Também estão sujeitos a efeitos tais como temperatura, processo de fabricação, assentamento da via, esforço de flexão etc. que quando

combinados podem agravar o nível de tensões principalmente na região do boleto, levando o componente ao colapso.

Os autores Schneider e Reguly (2014); Neto (2012); Queiroga et al. 2015, respectivamente, fazem uma abordagem referente aos defeitos de fabricação como os defeitos metalúrgicos, formados durante a solidificação do lingote (como segregações, inclusões e trincas de solidificação) ou no processo de laminação (dobras).

2.1.4 - Tensões residuais: causas e ocorrências

A diversidade de defeitos em trilhos está relacionada as tensões residuais trativas, as chamadas de “assassinias silenciosas”. AWS (2010) explica que estas tensões existem em materiais e estruturas, independente da presença de quaisquer cargas externas ou gradientes térmicos. Uma estrutura projetada para falhar em um certo nível de tensões pode falhar em um nível muito menor à medida que um componente de tensões residuais é adicionado a ele.

Os efeitos deletérios das tensões residuais, em uma estrutura, geralmente são observados somente após as cargas ou condições ambientais externas serem introduzidas, de modo que os efeitos combinados causam danos visíveis à estrutura. As tensões residuais influenciam significativamente na estabilidade estrutural e dimensional, bem como na vida útil de diferentes produtos laminados, como os trilhos. Thakur et al. (2006) complementam, tensões residuais combinadas com cargas estáticas e dinâmicas externas podem causar falhas catastróficas prematura de trilhos, o que pode levar a grandes acidentes ou perda de propriedades.

Por fim, Melado (2014) apud Kandil et al., (2001) explicam que as tensões residuais trativas na superfície do material podem ser a maior causa frequente de fratura por fadiga, trincas de têmpera e trincas de corrosão sob tensão. Por outro lado, as tensões residuais de compressão na superfície das peças tendem a reequilibrar as tensões trativas.

Quando um componente soldado, contendo uma distribuição inicial de tensões residuais (Fig. 5, curva 0), é submetido a um carregamento de tração, as tensões residuais se somam diretamente às tensões do carregamento, enquanto não ocorrerem deformações plásticas no componente. Assim, as regiões da solda, nas quais as tensões residuais de tração são mais elevadas, atingem condições de escoamento plástico antes do resto do componente (curvas 1, 2 e 3, Fig. 2.4).

O desenvolvimento de deformações plásticas, localizadas principalmente na região da solda, tende a diminuir as variações dimensionais que eram as responsáveis pela existência das tensões residuais. Desta forma, quando o carregamento externo é retirado, o nível dessas tensões

fica reduzido (curva 4, Fig. 2.4). Isto é, as variações dimensionais ocorridas na soldagem e responsáveis pelas tensões residuais são, pelo menos parcialmente, removidas pela deformação plástica causada pelo carregamento posterior.

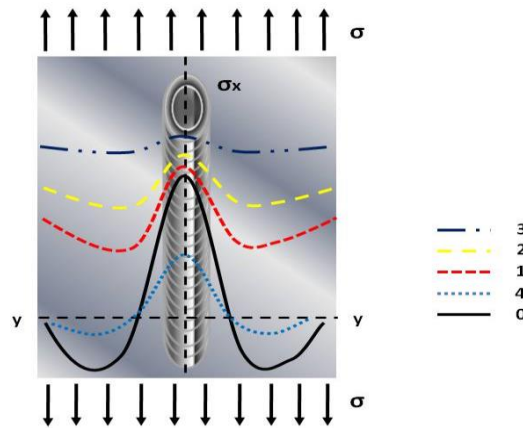


Figura 2.4 - Distribuição de tensões em um componente com uma solda de topo submetido a carregamentos crescentes (curvas 1,2, 3) e distribuição de tensões residuais após a liberação do carregamento.

Alguns efeitos específicos da presença de tensões residuais em um componente soldado são abordados por Marques (2005), os quais:

- Flambagem;
- Falha por fadiga;
- Fratura frágil;
- Formação de trincas;
- Corrosão sob tensão.

Algumas concessionárias no Brasil observaram, entre os efeitos das tensões residuais, o comportamento mais frequente de formação de fraturas e falhas em soldas e juntas soldadas.

A exemplo, no estudo de caso para a ferrovia MRS Logística, Assis (2018) apurou a ocorrência de fraturas no 1º semestre de 2017 nas seguintes proporções: 59% em trilhos, 33% em soldas alumino térmicas e 8% em soldas elétricas.

Os problemas recorrentes encontrados nas ferrovias da Empresa VALE S.A, escopo deste estudo, mostram a formação de trincas longitudinal na ZTA-Zona Termicamente Tratada e/ou zona de transição da junta soldada realizada por centelhamento dos trilhos ferroviários. Essas falhas geram constantes intervenções para reparos gerando grandes prejuízos. Estas trincas aparecem após um determinado tempo de uso da via. Neste sentido, a formação dessas

trincas ainda não é bem entendida do ponto de vista processual e metalúrgico (BRAGA et al, 2018).

2.2 - MANUTENÇÃO EM TRILHOS FERROVIÁRIOS

2.2.1 - Definições e conceitos

Por gerações a manutenção foi vista como uma atividade onerosa e maçante, o que levava os gestores a não ter uma visão do impacto dessa atividade sobre as relações de produtividade e mercado. Este cenário mudaria com a nova visão holística de um mercado que busca confiabilidade, segurança e disponibilidade a custos adequados. Como explica Campos (2018) “...ou seja, é um processo indispensável para uma organização que pretende se diferenciar no atual mercado competitivo...”

De posse de uma correta gestão de manutenção, é possível promover a maior disponibilidade e desempenho dos equipamentos garantindo o perfeito estado de funcionamento da produção e contribuindo para o retorno sobre o investimento (Pereira, 2010).

Campos (2018) apud Filho (2008) traz uma abordagem da gestão de manutenção como a adequação de atos, normas e instruções de procedimentos ligados a um sistema de manutenção, definindo objetivos que a equipe de manutenção deve cumprir.

A relação produtividade versus a infraestrutura ferroviária está condicionada, entre outros aspectos, à gestão da manutenção de ativos ferroviários, que requer considerável investimento para manter o sistema em condições estruturais adequadas.

Todavia, é importante definir a gestão de ativos, para diferir os objetivos da gestão de manutenção. Primeiramente a norma ABNT NBR ISO 55000 define ativo como um item, algo ou entidade que tem valor real ou potencial para uma organização. O controle efetivo e a governança de ativos das organizações são essenciais para perceber o valor através da gestão de riscos e oportunidades, para alcançar o equilíbrio desejado de custo, risco e desempenho, isto é, a gestão dos ativos.

Procaci (2018) define gestão de ativos como uma atividade por meio da qual toda a organização gerencia de maneira sustentável os seus ativos, levando em consideração tanto o desempenho, quanto seu custo durante o ciclo de vida. Complementando o conceito, Dantas (2014) apud OCDE, (2001) afirma que qualquer gestão de ativos está alicerçada em três pontos fundamentais: reconhecimento do valor económico dos ativos, eficiência económica e otimização das despesas ao longo do ciclo de vida dos ativos e o perfeito entendimento do papel da organização como vigilante e gestor dos ativos.

O transporte ferroviário deve oferecer confiabilidade, pontualidade e segurança para o mercado logístico. O adequado monitoramento dos ativos ferroviários influencia diretamente a confiabilidade do seu sistema como um todo.

Partindo deste princípio, a NBR 5462 – Confiabilidade e Mantenabilidade define a manutenção, como a combinação de todas as ações técnicas e administrativas, incluindo as de supervisão, destinadas a manter ou recolocar um item em um estado no qual possa desempenhar uma função requerida (ABNT, 1994). Dessa forma, a manutenção dos equipamentos e componentes de uma superestrutura ferroviária tem como objetivo manter a disponibilidade das ferrovias, diminuindo as interdições e restrições de velocidade (Moura, 2018).

É importante ressaltar que a Gestão de Ativos incorpora a Manutenção, e auxilia na definição das melhores estratégias a serem aplicadas para que seja atingido o objetivo final (TROMBETA, 2016). Na figura 2.5 estão representados os objetivos da gestão de ativos e os tipos e estratégias de manutenção.

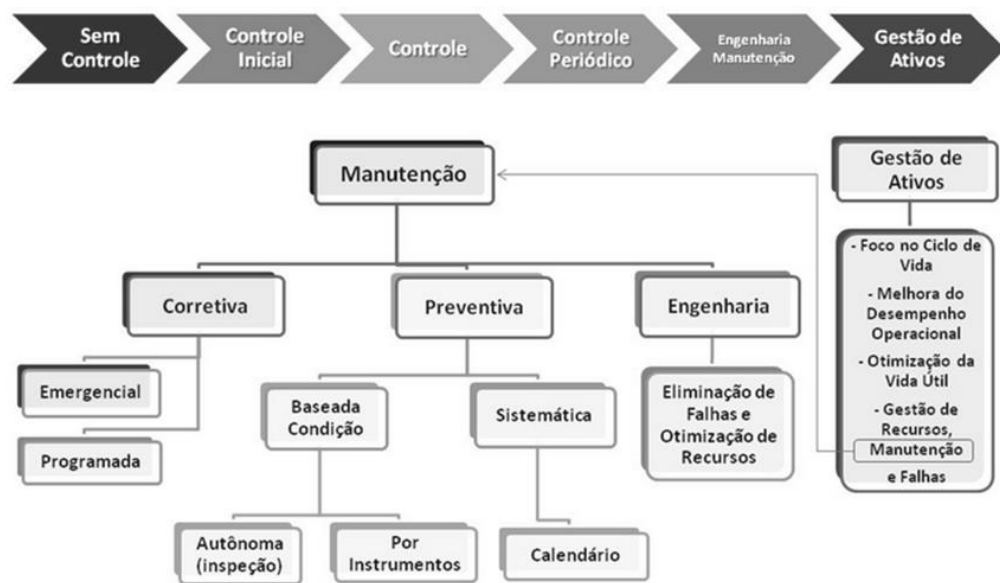


Figura 2.5 - Estratégias de manutenção e a gestão de ativos; (Trombeta, 2016).

2.2.2 - Estratégias de manutenção

A estratégia de manutenção de uma ferrovia depende de inúmeras variáveis, mas principalmente do orçamento disponível, das características da via e da carga transportada e das metas da empresa quanto ao volume de transporte e disponibilidade da via. Para isso, é primordial a análise do desempenho dos equipamentos e dos tipos de manutenção mais adequados para cada cenário.

Corrêa (2022) conceitua as estratégias de manutenção como o conjunto de ações através de um plano de manutenção, o qual é definido pelo gestor de manutenção, para organizar e

trazer qualidade aos serviços prestados, considerando o tipo de manutenção mais adequado para o modelo e idade do equipamento.

Para definir qual a estratégia de manutenção a ser utilizada, a Curva PF – falha potencial, é uma ferramenta que auxilia a gestão de manutenção e auxilia a gestão a entender e analisar o comportamento dos equipamentos perante sua vida útil. Como define Teles (2017), a Curva PF é uma ferramenta analítica essencial para um plano de manutenção que seja baseado em confiabilidade e esteja seguindo os padrões RCM - *Reliability Centered Maintenance* ou Manutenção Centrada na Confiabilidade (Figura 2.6).

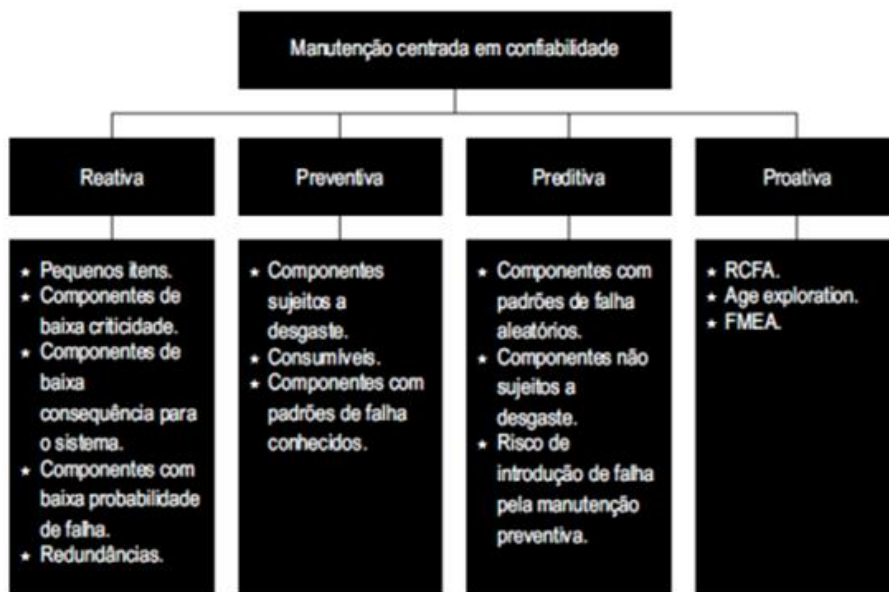


Figura 2.6 - Manutenção Centrada na Confiabilidade, segundo Pinto (2004)

De forma funcional, a Curva PF é um gráfico que conflita em um plano cartesiano simples a performance do equipamento sobre o seu tempo de funcionamento. Com o objetivo principal de identificar o intervalo PF, que seria o tempo entre a falha potencial e falha funcional, como mostra a Figura 2.7:

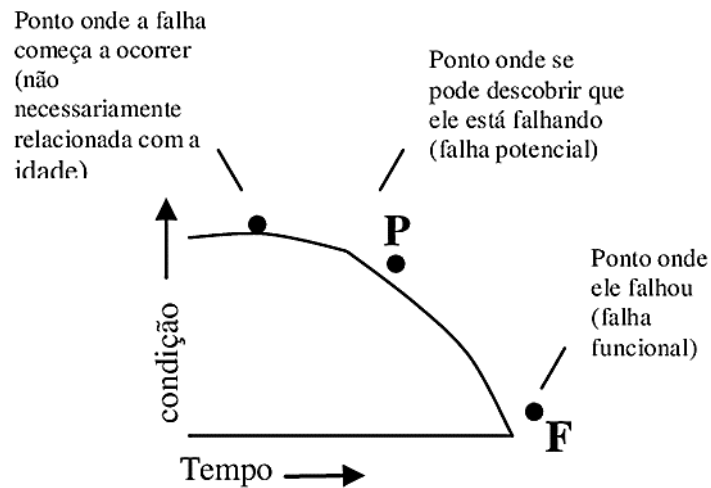


Figura 2.7 - Pontos de Falhas. MOUBRAY, John. SPES, 2000.”

A falha potencial é um sinal que indica que a máquina está com um defeito, mas não para de funcionar, como um pequeno vazamento, por exemplo. Se o problema for solucionado neste momento, o custo com reparo e o período de inércia da máquina será menor. Enquanto a falha funcional é quando o vazamento prejudica outros componentes e a falha interrompe a produção da máquina, gerando custos mais altos, uma vez que estamos tratando de manutenções corretivas (CORRÊA, 2022).

Baran (2015) explica que a análise de criticidade é uma técnica que identifica e classifica efeitos e eventos potenciais baseados nos impactos e na importância dos itens para o processo. Assim, a criticidade é diretamente proporcional ao impacto que esse equipamento pode causar no processo. Essa análise pode posteriormente ser aplicada concomitantemente com estudos de risco e de confiabilidade de projetos.

No Brasil, os planos de manutenção estão alicerçados na norma NBR 5462, que define a estratégia de manutenção, como a forma de assegurar eficiência e produtividade, controlando custos, otimizando recursos e ampliando oportunidades, a partir dos três tipos de manutenção diferenciados como mostra a Quadro 2.1.

Quadro 2.1 - Tipos de manutenção segundo a Norma NBR 5462

Corretiva	Efetuada após a ocorrência de uma pane destinada a recolocar um item em condições de executar uma função requerida.	Altos custo e tempo, e baixo investimento Interrupção do processo ou redução da performance. Lucro cessante (mau funcionamento do equipamento que impediu o lucro esperado para esse período)
Preventiva	Efetuada em intervalos predeterminados, ou de acordo com critérios prescritos, destinada a reduzir a probabilidade de falha ou a degradação do funcionamento de um item	Custo, tempo e investimento intermediários Frequência atrelada e ou ao tempo, à quilometragem, à produtividade.
Preditiva	Aplicação sistemática de técnicas de análise, para reduzir ao mínimo a manutenção preventiva e diminuir a manutenção corretiva.	Falhas em estágio inicial, quando ainda não são prejudiciais ao equipamento e/ou processo de produção. Custo e tempo baixos e investimento alto.

Apesar dos tipos de manutenção apresentarem significados, ações e objetivos completamente diferentes, como citam Apolinário Júnior (2019 pag.21); Modesto et al. (2019 pag. 20); Felix et al (2018), a norma viabiliza a análise da estratégia a partir dos parâmetros, custo, tempo e investimento.

As organizações ferroviárias europeias buscam reduzir ainda mais os custos gerais de vida de seus equipamentos e melhorar o seu desempenho. O autor Vaidya (2017) relata que eventualmente, as despesas de manutenção, os horários de reparação, a eficiência e a relevância dos reparos, produtividade e competitividade planejados tornaram-se mais exigentes. Por outro lado, Li (2017) reforça que os altos custos de instalação, recursos limitados e efeitos negativos das intervenções de manutenção tem motivado os gerentes de infraestrutura ferroviária e empreiteiros de manutenção a procurar estratégias de manutenção econômicas, além da manutenção tradicional baseada em tempo estratégias, que executam a manutenção em um intervalo ideal de tempo/uso.

De acordo com os padrões europeus (EN 13306: 2001 2001), a manutenção constitui todas as ações necessárias durante o ciclo de vida operacional de um ativo [...] essas ações abrangem o trabalho de manutenção específico para os próprios ativos, bem como abordagens estratégicas para garantir que os programas de trabalho, tecnologias e procedimentos adequados estejam em vigor para uma manutenção confiável (KEFALIDOU et a., 2018).

Segundo Dantas (2014) a Companhia portuguesa REFER classifica a Manutenção em dois tipos: Manutenção Corretiva (MC) - Efetuada quando a condição dos ativos ultrapassou já os limites para intervenção; e a Manutenção Preventiva que subdivide-se em Manutenção

Preventiva Sistemática (MPS) - Efetuada regularmente independentemente da condição dos ativos e Manutenção Preventiva Condicionada (MPC) - Efetuada em função da condição dos ativos quando necessário, antes de os limites dos respectivos parâmetros serem ultrapassados.

As normas e literaturas existentes sobre as estratégias de manutenção apresentam pontos incomuns como, tempo, custo e risco, então de forma resumida segue o quadro 2.2 adaptados do autor Hupje (2020) em “Types of Maintenance Toolk”, onde são apresentados um breve resumo da tipologia, tarefas envolvidas, estratégias e intervalo de tarefas.

Quadro 2.2 - “Types of Maintenance Toolk”, Hupje (2020).

Manutenção Preventiva						Manutenção Corretiva	
Tipo de Manutenção	Baseado no tempo	Descoberta de falhas	Baseado em risco	Baseado em condição	Preditiva	Adiada	Emergência
Tipo de tarefa	Substituição de revisão programada	Teste funcional	Medição de condição	Cálculo e extrapolação de parâmetros	Inspeção ou teste	Reparar Substituir	Reparar Substituir
Estratégia	Restaurar para substituir independente da condição	Determinar se ocorreu uma falha oculta	Restaurar ou substituir com base em uma condição medida em comparação com um padrão de definição	Determinar se a falha é iminente e a intervenção é necessária	Determinar a condição e conduzir a avaliação de risco para determinar quando a próxima inspeção, teste ou intervenção será necessária	Restaurar ou substituir após falha Resultado de uma estratégia de execução até a falha ou uma falha não planejada	Restaurar ou substituir após falha não planejada
Intervalo	Tempo fixo ou intervalo de uso, ex. 1 mês, 1000hr. ou 10000 km	Intervalo de tempo fixo (pode ser definido com base na avaliação de risco)	Intervalo de tempo fixo para medições/inspeções de condição	Monitoramento online contínuo de parâmetros, intervenção conforme necessário	Intervalo baseado no tempo entre tarefas e escopo da tarefa é baseado na avaliação de risco	Não aplicável. A intervenção é adiada para permitir o planejamento adequado	Intervenção imediata necessária

Segundo Hupje (2020) a manutenção de emergência deve ser evitada ao máximo, e isso as organizações de classe mundial já garantem que menos de 2% de sua manutenção total seja de manutenção de emergência. O autor ainda afirma que geralmente esta manutenção não é bem planejada (preparada) e, portanto, é altamente ineficiente e cara (normalmente pelo menos 3 a 5 vezes mais cara que uma manutenção preventiva bem planejada).

A manutenção de ativos ferroviários por parte das companhias, em especial para a correção das falhas provenientes de tensões residuais nos trilhos, tem utilizado estratégias de emergência, apresentando uma eficiência muito baixa com tempo adicional desperdiçado procurando materiais, organizando o acesso ao equipamento, aguardando outros negócios etc. e custos maiores são incorridos para que isso aconteça. Partindo desta premissa, na análise do custo-benefício, o custo de oportunidade não é considerado.

Assim explicam Souza (2008) a manutenção por oportunidade, ou manutenção oportunista, o tempo é aproveitado na paralisação do equipamento quando ela ocorre, em contraste à paralisação para manutenção planejada. A investigação minuciosa dessas oportunidades e suas ocorrências, sugerem questionamento quanto as oportunidades surgem, quais máquinas permitem reparos simultâneos e qual o tempo de duração.

A utilização da manutenção oportunista afeta diretamente no tempo total de disponibilidade dos equipamentos, pois uma vez aproveitado o tempo de paralisação por conta de uma avaria para a realização de uma intervenção preventiva, poupa o equipamento de parar posteriormente para cumprir o plano de manutenção preventiva. (Souza 2008 apud Osada 1993).

Os conceitos de confiabilidade e manutenibilidade são amplamente abordados na literatura de manutenção. Segundo Oliveira 2012 a confiabilidade é a probabilidade de um equipamento operar, sem falhas, durante um período predeterminado, isto é, à medida que se aumenta o tempo de avaliação, maior é a chance de acontecerem falhas, ou seja, menor será a confiabilidade da máquina ou do ferramental. Enquanto o conceito de manutenibilidade está relacionado ao grau de facilidade para se fazer o reparo em um equipamento, quando este é realizado, de acordo com os procedimentos definidos.

Quando o equipamento está operando, a avaliação da confiabilidade normalmente é feita após a estabilização de sua taxa de falhas, ou seja, quando se encontra na fase de Vida Útil.

O indicador utilizado para se fazer essa avaliação é o MTBF (*Mean Time Between Failures* – Tempo Médio Entre Falhas). A Manutenibilidade é medida por meio de um indicador chamado MTTR (*Mean Time To Repair* - Tempo Médio para Reparo). Esses indicadores são obtidos a partir dos registros de manutenção (PEREIRA, 2009).

Devido as demandas específicas do segmento industrial, a gestão da manutenção utiliza para análise de criticidade de equipamentos, métodos como a Manutenção Centrada em Confiabilidade (MCC), Classificação ABC, Análise dos Modos de Falha e Efeitos (FMEA) e Análise Crítica dos Modos de Falha e Efeitos (FMECA) adaptados para necessidades

específicas, sob a forma de ferramentas específicas para análise ou integradas umas as outras em programas de manutenção (BARAN, 2015).

De modo geral, a busca por otimizar a manutenção nas empresas partiu de uma melhoria nas estratégias associadas, para que estas se adaptassem melhor à realidade de cada empresa. Existem muitas ferramentas e informação, de acordo com Vasconcellos e Fregma (2022) e Ribeiro (2017):

- Técnicas de manutenção baseada na confiabilidade (RCM);
- Técnicas de Stock baseado na confiabilidade (RCS);
- Técnicas de manutenção baseada no negócio (BCM);
- Técnicas de manutenção baseada nas pessoas (PCM);
- Técnicas de manutenção baseada no risco (RBM);
- Manuais do fabricante;
- Planos de manutenção;
- Políticas de manutenção; e
- Experiência ou know-how.

O método mais utilizado na gestão de manutenção de ferrovias é a Manutenção Centrada em Confiabilidade (MCC), uma vez que Igba et. al (2013); Garza (2002); Oliveira (2012); Guedes (2017) afirmam que o objetivo da MCC é preservar a função mais importante do equipamento ou sistema, assegurar a confiabilidade e disponibilidade necessária com o menor custo possível. Assim como consideram o método mais efetivo para o tratamento de falhas potenciais.

2.3 - CUSTO DE MANUTENÇÃO FERROVIÁRIA

A Agência Nacional de Transporte Ferroviário - ANTF (2020) apontou que as ferrovias associadas à agência investiram mais de R\$ 75 bilhões entre 1997 e 2019 (ou R\$ 113,37 bilhões, atualizados com base no IPCA), com uma média anual de R\$ 3,26 bilhões (ou 4,93 bilhões, valor atualizado pelo mesmo índice), e são exemplos de sucesso da concessão de serviços à iniciativa privada. Os valores contemplam, principalmente, a melhoria e a recuperação da malha, a compra e a reforma de material rodante, além da aquisição de novas tecnologias, capacitação profissional, qualificação das operações, entre outros.

Para implementar a gestão de manutenção, as concessionárias ferroviárias buscam diferentes ferramentas imprescindíveis para o setor, como afirma Peres e Lima (2005), os

esforços aplicados dentro de uma proposta de modelagem permite o controle de custos de manutenção e garante à engenharia da manutenção a função responsável pelo gerenciamento de indicadores e resultados esperados no plano estratégico corporativo.

Antes de aprofundar nos conceitos embasadores do custo de manutenção, existem teorias econômicas que precisam ser ponderadas, como a rentabilidade e o risco. SILCON (2021) desenvolveu um estudo comparativo sobre a rentabilidade das ferrovias e rodovias, e pondera que qualquer atividade está sujeita ao risco, algumas mais do que outras. Nos projetos de ferrovias, a diferença de retorno (e de risco) entre as atividades resulta de muitos fatores, existem atividades complementares que são fundamentais para as decisões de investimento, como exemplos, a cadeia complexa de operações de logística, armazenagem, transbordo, instalações portuárias integração com outros modais etc., complementares à ferrovia propriamente dita que define a rentabilidade e risco do sistema.

Atividades e investimentos com maior risco devem apresentar retorno esperado mais elevado – incorporando um prêmio - para atrair interesse e recursos. É uma exigência ex-ante do mercado, e muitas evidências empíricas suportam esta regra.

Na pesquisa comparativa de risco e rentabilidade, SILCON levantou uma amostra de 2.174 empresas de 35 setores com informações contábeis para o período 2013-2018. A taxa de retorno privado foi medida pela relação entre lucro operacional e o ativo operacional, e o risco, simplesmente pelo desvio padrão do período. A relação entre retorno e risco segue o esperado: as ferrovias têm o retorno médio mais baixo e igualmente risco menor. O retorno e o risco refletem os critérios da ótica privada e os resultados do estoque de projetos implantados no passado no desempenho contábil atual das empresas do setor. Novos projetos devem apresentar rentabilidade privada mais elevada do que as médias dos estoques de capital em operação. O retorno esperado de um novo projeto – com melhor tecnologia, logística etc. – deve superar a taxa média de retorno, exceto nos casos de deseconomias de escala, restrição na oferta de fatores etc.

Nos anos seguintes, mais exigências foram implantadas, nem sempre atendidas e, menos ainda, verificadas. Em 2019, o tema foi resgatado pelo governo federal para reimplantar a ACB principalmente nos projetos de infraestrutura. A avaliação dos projetos sob a ótica da economia como um todo exige uma metodologia específica (denominada ACB – análise custo-benefício). E nesta ótica, existem questões pouco exploradas no setor de ferrovias, e quando abordadas, de forma incompleta e inadequada.

A ACB incorpora todos os custos e benefícios do projeto ao longo da sua vida no processo de decisão, inclusive as externalidades e outros itens não considerados nas decisões

do setor privado. Com a visão sistêmica, a metodologia permite avaliar qualquer projeto sobre o ponto de vista dos seus benefícios líquidos para a economia como um todo, complementando a análise normalmente utilizada pelo setor privado, interessada apenas na rentabilidade para o investidor. Ou seja, os projetos são avaliados considerando não a rentabilidade financeira privada, mas a sua contribuição para a sociedade.

A grande relevância da manutenção dentro da estratégia das empresas se dá por conta do aumento da complexidade e a grande inserção da automação nos mais diversos sistemas de produção, a fim de manter máquinas e equipamentos em níveis de performance desejados, reduzindo as paradas por manutenção e conseqüentemente os custos envolvidos (JUNIOR et al, 2015).

Segundo Correa (2021) a gestão de custos de manutenção é muito importante para a operação de toda empresa. Os seus gastos devem ser administrados com o intuito de impedir investimentos desnecessários, além de auxiliar na redução de orçamento de manutenção, sem perder a eficiência do trabalho. E reforça, “um bom gestor que cuida da administração dos gastos consegue fazer um planejamento mais eficiente e investir o capital de forma mais assertiva”.

O custo de manutenção, segundo Kardec e Nascif (2009) define e baliza as ações que foram decididas, bem como, as estratégias de manutenção adotadas para manter a disponibilidade operacional, a eficiência dos equipamentos, e podem ser avaliados em nível empresarial através de dois indicadores, como:

- o custo da manutenção em relação ao faturamento bruto da empresa (%) e;
- o custo da Manutenção em relação ao patrimônio (valor estimado dos ativos) (%).

Para uma correta previsão do custo por manutenção, Júnior et al (2015); Cavalcante e Almeida (2005), considera que o modelo ideal para os custos de manutenção corretiva, preventiva e preditiva esteja em função do tempo. E dentro deste arcabouço estratégico devem ser considerados os seguintes custos gerados, classificados em:

- Custos Diretos: estão relacionados com as atividades que fazem parte do dia a dia da operação, como peças de reposição; ferramentas; mão de obra; insumos e serviços.
- Custos Indiretos: é possível adicionar a depreciação de equipamento, com o tempo os ativos perdem o valor naturalmente, seja pelo seu uso ou pelo desgaste dos componentes. Então, é necessário fazer manutenções para que eles voltem a funcionar corretamente.

- Custos Induzidos: surge pelos impactos que são gerados na cadeia produtiva, como um problema no setor de manutenção. Por exemplo, hora extra em virtude de um problema que surgiu em um determinado equipamento.

Retterer e Kowalski (1984) e SAE G-11 (1990) classificaram os custos fundamentais associados com a manutenção mais especificamente em quatro áreas: custos diretos, custos de produção perdidos, custos de degradação e custos de espera, respectivamente:

- custos diretos estão associados à manutenção do equipamento operacional e incluem os custos de inspeção periódica e manutenção preventiva, custo de reparo, custo de revisão e custo de manutenção.
- custos de produção perdidos estão associados à perda de produção devido a avaria do equipamento primário e indisponibilidade do equipamento de reserva. Degradação os custos estão associados à deterioração da vida útil do equipamento devido a menor manutenção.
- os custos de espera estão associados à operação e manutenção equipamento de espera. O equipamento de reserva é usado quando as instalações primárias estão em manutenção ou inoperante.

Muitos fatores influenciam os custos de manutenção, incluindo a condição do ativo (ou seja, idade, tipo e condição), conhecimento e experiência do operador, política da empresa, tipo de serviço, habilidades do pessoal de manutenção, ambiente operacional, especificação de equipamentos e controles regulatórios.

Segundo o Manual de Custos Referenciais Ferroviários –ANTT (2019) os custos horários de manutenção são obtidos por meio da seguinte expressão:

Equação (2):

$$Mh = Va \times Kn \times HTA$$

onde: Mh representa o custo de manutenção horária (R\$/h);

Va representa o valor de aquisição do veículo (R\$);

k representa o coeficiente de manutenção;

n representa a vida útil (anos);

HTA representa o total de horas trabalhadas por ano.

Neste manual é possível verificar que os coeficientes de manutenção dos equipamentos são fornecidos pelos fabricantes e estimados em função de levantamentos e análises da série histórica dos custos.

No coeficiente de manutenção (k) estão incluídos os seguintes itens:

- Manutenção corretiva;
- Manutenção preventiva;
- Reparos;
- Substituição de peças e componentes (lâminas, caçambas, garras, escarificadores, etc.);
- Custo do veículo lubrificador;
- Perda de produção relativa a horas paralisadas para a manutenção;
- Mão de obra especializada para a manutenção (encarregado de oficina, mecânicos, eletricitas, soldadores e ajudantes).

Dhillon (2002) descreve no livro *Engineering maintenance: A Modern Approach* que a fase de manutenção é um elemento importante do ciclo de vida do equipamento que devem ser mantidos satisfatoriamente para um desempenho eficaz, e que o custo de manutenção desses equipamentos muitas vezes varia de 2 a 20 vezes o custo de aquisição. O autor ainda aborda a definição do custo de manutenção como custos que incluem oportunidades perdidas em tempo de atividade, taxa, rendimento e qualidade devido a equipamentos inoperantes ou operando insatisfatoriamente em além dos custos envolvidos com a degradação da segurança das pessoas relacionada aos equipamentos, propriedade e o meio ambiente.

Percebe-se a escassez da abordagem da oportunidade perdida ou negligenciada na análise de custo de manutenção nas literaturas existentes. O cenário das externalidades nesta atividade pode ser positivo e negativo, ainda que do ponto de vista da empresa, reflete no contexto global da economia local e na logística de infraestrutura de transporte. Deixa-se de lado, a prática do conceito de que o custo de manutenção é simplesmente descrito como as despesas de mão de obra e materiais necessários para manter os equipamentos/itens em satisfação estado operacional.

2.4 - ANÁLISE CUSTO-BENEFÍCIO E CUSTO DE OPORTUNIDADE

A relação entre o custo e o lucro as vezes é incompreendida na visão dos profissionais ligados às atividades empresariais, à medida que o montante significativo se sobrepõe aos custos gerados em determinadas atividades, acabando por camuflar oportunidades de investimento em outras áreas que equilibrem a conformidade do sistema, isto quer dizer, o melhor custo-benefício raramente é equacionado.

Segundo Talaszka, 2018; Medeiros, 2005 o estudo do comportamento dos custos é relevante pelo fato de que a base de sustentação de muitas decisões gerenciais está no

conhecimento de como os custos podem variar em função do nível de atividade. A partir de uma análise criteriosa é possível o gestor identificar com precisão o cenário e a realidade de empresa.

De um modo geral, os custos relacionados com a produção de um serviço de transporte público podem ser distinguidos em custos de investimento (também conhecido como custo de capital), necessários para a realização ou aquisição de componentes (para ferrovias: infraestrutura, material rodante, instalações etc.) para a ação planejada e custos operacionais (também conhecidos como custos de gestão), incluindo aqueles para a operação e manutenção do serviço de transporte.

Dentre os estudos que incluem tanto os custos de investimento como de operação nas ferrovias, cabe mencionar o artigo de Baumgartner (2001). Além disso, um estudo sobre a reativação de uma linha férrea no centro da Itália (Santinelli, 2007) inclui uma indicação básica de investimento e custo operacional para ela. Ott (2001) comparou os custos de infraestrutura rodoviária e ferroviária. Mancuso e Reverberi (2003) estudaram custos operacionais e organização de mercado em serviços ferroviários italianos.

Segundo Afonso (2016) é importante compreender que para avaliação de projetos e serviços, tanto a análise financeira como a análise econômica ou análise custo-benefício, auxiliam na definição e seleção dos projetos mais efetivos para o bem-estar de um país. Enquanto a análise financeira de projetos tem como objetivo avaliar a qualidade do projeto do ponto de vista do investidor, por exemplo, uma autoridade portuária, a análise custo-benefício visa a avaliação do projeto na perspectiva da sociedade em geral onde ele irá ser executado, ou seja, medir os seus efeitos sobre a sociedade como um todo.

Destarte, é importante diferenciar o custo efetividade do custo-benefício, haja vista que ambos são almejados pela gestão das concessionárias no mundo para a melhor tomada de decisão de negócios, onde trazem uma melhor avaliação do investimento seja do ponto de vista monetário e ou qualitativo.

O conceito da análise do custo eficácia ou custo efetividade é a sua utilização como ferramenta de avaliação, em que os resultados produzidos por um projeto ou investimento, embora possa ser expresso em termos monetários, não devem ser avaliados apenas pelos seus valores monetários, fatores qualitativos também devem ser considerados. A "relação custo-eficácia" pode ser calculada (STREPHONSAYS, 2017), segue:

Equação (3):

$$ICE = \frac{CI}{RI}$$

ICE = Índice de Custo – Efetividade

CI = Custo do Investimento

RI = Resultado do Investimento

Por outro lado, a PontoTel, empresa do grupo Sankhya na área de tecnologia (2021) explica que a relação custo e benefício pode ser entendida como a rentabilidade de um projeto, compra ou serviço, e pode estar ligada a redução de custos, ou a garantia de qualidade e funcionamento. E, quando o risco envolvido na ação não valer a pena, o benefício não valerá o custo envolvido. Para Strephonsays (2021) a análise de custo-benefício é um compromisso entre as adições de custos e benefícios para implementar uma decisão de negócios.

Assim, os critérios de tomada de decisão serão prosseguir com o investimento se os benefícios superarem os custos. Ainda sobre essa análise, o autor explica que tanto os custos quanto os benefícios são expressos em valores monetários, e não deve ser subestimado os custos ou superestimar os benefícios, além de considerar todos os custos diretos e indiretos, custos fixos e variáveis, devem ser considerados. Além disso, devem ser considerados o custo de oportunidade do projeto (benefício potencial conjurando fundos em um investimento alternativo), o custo de não executar o projeto e os custos potenciais de falha do projeto.

O retorno envolvido no custo-benefício não deve estar ligado somente a questões financeiras, pode ser visto em qualidade de serviço, lucratividade, mais oportunidades, crescimento da marca, projeção de projetos, dentre outras questões. Para tanto, algumas considerações acerca de custos e benefícios devem ser respondidas, como mostra o Quadro 2.3:

Quadro 2.3 - Considerações entre o custo e benefício

CUSTOS	BENEFÍCIOS
<ul style="list-style-type: none"> • quanto o processo diminuirá a produtividade; • quais serão os custos operacionais; • quantas horas serão gastas na implementação, treinamento e execução; • quais ferramentas precisarão ser contratadas; • quais os efeitos disso no cotidiano dos colaboradores; • dentre outros fatores, que quando respondidos simbolizam os custos. 	<ul style="list-style-type: none"> • qual será a receita gerada; • haverá aumento de produtividade com o tempo; • quais serviços serão melhorados; • os efeitos positivos na gestão e na vida dos funcionários; • quanto a fidelização dos consumidores será impactada; • dentre outros benefícios que serão notados, refletindo o custo-benefício da estratégia adotada.

A diferença entre análise de custo-eficácia ou efetividade (ACE) e análise de custo-benefício (ACB) depende principalmente se o foco é dado ao valor do resultado (na análise de custo-eficácia ou o valor monetário análise custo benefício) de um projeto. (Ver Quadro 2.4).

Quadro 2.4 – Diferença entre custo-benefício e custo efetividade

CUSTO EFETIVIDADE	CUSTO-BENEFÍCIO
Conceito	
É uma forma de análise econômica que compara os custos relativos e os resultados (efeitos) de um projeto.	Atribui um valor monetário à medida do efeito de um projeto.
Natureza da Avaliação	
É uma técnica mista (quantitativa e qualitativa) de avaliação de projetos	É uma técnica de avaliação quantitativa de projetos.
Uso	
É adequada para organizações relacionadas a serviços, especialmente para aquelas no setor de saúde.	É adequada para avaliar projetos altamente técnicos e industriais, uma vez que os valores monetários podem ser facilmente atribuídos a esses projetos.
Custo de Oportunidade	
Geralmente não considera os custos de oportunidade.	O custo de oportunidade deve ser considerado.

O custo de um projeto de transporte em termos econômicos é o valor dos recursos que devem ser consumidos para realizar o projeto. O que deve ser estimado é o valor total dos custos de implementação e quaisquer custos operacionais adicionais. Segundo Lang (2007) apud Abreu e Stephan (1982) os custos e benefícios representam efeitos sobre as pessoas ou entidades em momentos distintos de forma a fornecer uma estimativa da variação do bem-estar da coletividade resultante do projeto.

Lang (2007) também traz algumas literaturas sobre o custo efetividade, no geral, esta avaliação pode ser empregada para comparar diferentes projetos de infraestrutura que abordam o mesmo problema de maneira diferente, sendo este projeto expresso: por unidade de um padrão conseguido por unidade monetária gasta; por unidade do padrão conseguindo em diferentes programas com iguais custos; pelo custo por unidade desse indicador conseguido; ou ainda pelo custo de programas que atingem um mesmo padrão desse indicado analisado.

Vale destacar que o padrão analisado se refere ao critério utilizado para medir a efetividade, como por exemplo, no caso das ferrovias, para segurança, produção de transporte, tarifa do frete, entre outros critérios.

2.4.1 - Custo de oportunidade e a análise de decisão (trade-off)

No Caderno de Estudos nº 02, FIPECAFI, Pereira et al. 1990 aborda a origem da expressão “custo de oportunidade” por Frederick Von Wieser, para definir o valor de um fator de produção em qualquer uso que lhe fosse dado, sendo tal custo de oportunidade “a renda líquida gerada pelo fator (de produção) em seu melhor uso alternativo”. Pereira 2022 publicou no site Dicionário Financeiro um artigo que traz uma abordagem teórica do custo de oportunidade, onde se mensura o custo daquilo que se deixa de fazer quando é preciso fazer uma escolha de qualquer tipo. Este custo se diferencia de um custo real, também conhecido como um custo contábil, que acontece de maneira direta e quantitativa.

Em algumas literaturas econômicas, o custo de oportunidade é chamado de custo alternativo, o qual se refere ao custo econômico da alternativa que deixou de ser escolhida. Para a ciência econômica o custo de oportunidade é explicado pelo conflito de escolha de um agente econômico em contexto de escassez. Este conflito de escolha é conhecido como trade-off, é a decisão de longo prazo que a empresa deve adotar, que se traduz em uma situação de escolha quando se ganha uma coisa e se perde outra, sendo aquilo que se perde, o custo de oportunidade. Por outro lado, essa escolha se deve ao fato de que não será possível ter, em simultâneo, os objetos da escolha, isto é, realizar escolhas que possibilitem o melhor benefício ou de menor custo associado (PEREIRA, 2022).

As literaturas existentes sobre os conceitos e definições do custo de oportunidade, trazem interpretações contábeis quantitativas e qualitativas. Leone (1982): “benefício que se deixa de ganhar”; Martins (1987): “quanto à empresa sacrificou em termos de remuneração”. Gray & Johnston (1977): “lucro que se poderia ter conseguido”. Sizer (1974) ao dizer que custo de oportunidade concerne a melhor alternativa da qual se desistiu, isto é, a própria alternativa é o atributo, o que de certa forma, torna o conceito compacto e deixa a cargo do usuário caracterizar o que lhe é conveniente. Morse & Roth (1986): “o custo de oportunidade de uso de um recurso em um modo específico é o recebimento líquido de caixa esperado que poderia ser obtido se o recurso fosse usado na outra ação alternativa mais desejável”; Glautier & Underdown (1986) “o custo de oportunidade pode ser medido como o valor da próxima melhor alternativa abandonada, ou o recebimento líquido de caixa permitido como resultado de preferir uma alternativa ao invés da melhor seguinte.

Yip (2014) traz uma abordagem da revisão crítica do custo de oportunidade onde o conceito foi defendido como o principal custo de decisão por economistas e contadores, principalmente estudiosos da Escola de Londres na década de trinta. Em sua análise, o autor

identifica que os gerentes chamam apenas o modelo de contabilidade de custos de oportunidade no cálculo dos possíveis ganhos de diferentes alternativas de decisão mediante duas condições satisfeitas, a primeira é que não se encontre dificuldade em utilizar o modelo de contabilidade de custos de oportunidade; e a segunda condição é que o modelo forneça uma classificação prioritária das alternativas de decisão desejadas pelos gerentes que estão maximizando seus próprios benefícios de decisão.

“A adoção do modelo de custo de oportunidade, em um processo de decisão deve ser considerado como análise, em vez de uma análise estática do ponto do tempo e um modelo de processamento de valor deve ser estabelecido para identificar a aplicação realista do conceito de custo em um processo de tomada de decisão” (YIP, 2014).

Observa-se, pois, que o conceito de custo de oportunidade tanto em economia, como em contabilidade e em finanças tem como fundamento à questão da escolha entre alternativas de utilização de recursos, todavia, ele – custo de oportunidade – só aparece claramente após as alternativas terem sido elencadas e mensuradas.

Desta forma, o cálculo do custo de oportunidade, quando é possível mensurar, é calculado a partir do benefício que se teria com a opção que não foi a escolhida. Este cálculo considera o benefício da alternativa que se teve de abandonar, seja por uma atividade, como também ao adquirir um bem, tornando-se um custo para a que foi realmente escolhida.

É possível calcular qualitativamente de uma maneira mais elaborada seguindo uma comparação entre uma atividade A e uma outra B, um conflito entre duas escolhas. Para isso será considerado o benefício atribuído a cada uma, do ponto de vista da outra.

Entende-se que o desenvolvimento econômico e social impulsiona a infraestrutura ferroviária a aumentar o nível de qualidade, que por sua vez, requer tomadas de decisões estratégicas de manutenção cada vez mais atreladas ao custo-benefício potencial. Como afirma Xavier (2008) quanto mais se produz, mais premente se torna à necessidade de manutenção e menor as oportunidades em executá-la.

Diante deste cenário, a manutenção de ativos ferroviários, em especial para a correção de falhas provenientes de tensões residuais nos trilhos das ferrovias na Amazônia Brasileira, tem utilizado estratégias de manutenção de emergência, apresentando uma eficiência muito baixa, com tempo adicional desperdiçado e custos maiores são incorridos para que isso aconteça, porém, não mensurados segundo seus gestores. Este hiato na gestão de manutenção é explicado pelo desconhecimento do real impacto que o custo geral no processo, como perdas de receita e de custos de oportunidades nas atividades, como explica Talaszka, (2018) apud Medeiros (2005):” o estudo do comportamento dos custos é relevante, haja vista que a base de

sustentação de muitas decisões gerenciais está no conhecimento de como os custos podem variar em função do nível de atividade”.

Na busca por trabalhos anteriores que estudaram o custo de oportunidade como ferramenta de decisão, a pesquisa e análise são relativamente baixas ou quando estudadas individualmente. Pereira (2018) faz uma abordagem econômica do custo de oportunidade, onde o conflito de escolha é conhecido como trade-off. Castello Branco (2008) considera o custo marginal como o custo de oportunidade, ou seja, o valor dos recursos adicionais empregados, se estes vierem a ser utilizados numa atividade alternativa. Van Jaarsveld et al., (2013) determinam os custos de oportunidade logística do uso do transporte rodoviário dentro de uma mineradora.

Kussano e Batalha (2012) consideram o custo de oportunidade como variável em um modelo de estrutura de custos logísticos do escoamento da safra de soja. Budai-Balke (2009) considera a otimização de manutenção clássica, onde o objetivo é encontrar um trade-off entre manutenção preventiva e corretiva. Sousa et al., (2018) apresentam uma metodologia útil como ferramenta de gestão de ativos, onde aplicaram um modelo linear que disponibiliza uma gama de soluções para análise de *trade-off* pelo decisor. Por fim, Van Essen et al (2004) explicam que em geral, a avaliação dos custos de oportunidade no setor ferroviário e aéreo continua a ser uma abordagem escassa e pouco pesquisada.

2.5 - MECANISMO DE ANÁLISE PARA UMA TOMADA DECISÃO TRADE-OFF

2.5.1 - Machine Learning

O aprendizado da máquina já é uma realidade para o gestor ferroviário, quanto ferramenta de I.A implementadora nas tomadas de decisões preditivas de suas atividades operacionais. Nakhaee et al., (2019) afirma que nos últimos anos, as pesquisas adotaram o uso de algoritmos de aprendizado de máquina (ML) para diagnóstico e prognóstico de defeitos ferroviários, a fim de ajudar a indústria ferroviária a realizar respostas oportunas às falhas. Bukhsha et al., (2018) traz uma abordagem do uso do aprendizado de máquina como solução preditiva auxiliar na tomada de decisão de manutenção de ativos ferroviários. Seguindo este raciocínio, a predição sugerida nesta seção tem como foco o uso do *Machine Learning*, um subconjunto de IA (inteligência artificial) na predição comportamental dos custos de manutenção preventiva e corretiva para subsidiar a análise trade off para o gerenciamento estratégico na tomada de decisões de manutenção.

Segundo o *System Analysis Program Development - SAP* (2022) no *machine learning*, os algoritmos são treinados para identificar padrões e correlações em grandes conjuntos de dados, tomar melhores decisões e fazer previsões com base nessa análise.

Para a implementação do algoritmo de aprendizado de máquina foi utilizada a linguagem Python de programação, na versão 3.9.7. O algoritmo utilizado foi o *Gradient Boosted Decision Trees* (GBDT), como mostra a Figura 2.8, o qual utiliza métodos de conjunto tanto para classificação, quanto para regressão, implementado por meio das ferramentas do *Scikit-learn*, por ter código aberto e distribuição com licença BSD.

Por utilizar métodos de conjunto, os dados de entrada são ajustados por uma combinação de vários estimadores individuais e, por ser um método de *boosting*, estes estimadores são construídos em sequência para gerar um modelo mais preciso. Para a implementação de regressão, o GBDT é implementado por meio do *Gradient Boosting Regressor* (GBR), o qual gera bons resultados evitando sobre ajuste, quando o modelo se ajusta perfeitamente aos dados fornecidos, porém é ineficiente na predição de dados diferentes (WOLPERT, 2022).

Todavia, quando se trata de implementação dos mecanismos de análise para uma tomada de decisão, o uso de Inteligência Artificial como auxílio já é uma realidade nas empresas. Uma pesquisa elaborada pela empresa Gartner Resecher Circle, realizada em 2017, aponta que o uso de Inteligência Artificial nos últimos quatro anos, foi de 270%. Para 2022, esse número está previsto crescer ainda mais, chegando a 80% dos avanços relacionados ao *Marchine Learnig* e à Inteligência Artificial.

A utilização do aprendizado de máquina disponibiliza informações mais claras e cria um terreno fértil para o apontamento de ideias e tomada de decisão. Como apontam Wang e Mamo (2020) “o desafio está em desenvolver um modelo preditivo preciso para um sistema de gerenciamento de bateria”.

Para garantir o bom desempenho dos sistemas de micro irrigação, García Nieto et al. (2018) determinam as perdas de pressão produzidas a partir da abordagem de aprendizado de máquina não paramétrica conhecida como árvore de regressão impulsionada por gradiente (GBRT). Esse método é usado para tarefas de classificação, mas existe a possibilidade de aplicá-lo para uma tarefa de regressão (Baturynska e Martinsen, 2020). Estas pesquisas reforçam a importância em desenvolver estudos do custo de oportunidade no campo da gestão de manutenção ferroviária. Tomando o conhecimento da correlação entre as variáveis econômicas e operacionais em uma série temporal, e buscando instrumentos que permitam a identificação do custo de oportunidade nas previsões e gerenciá-los.

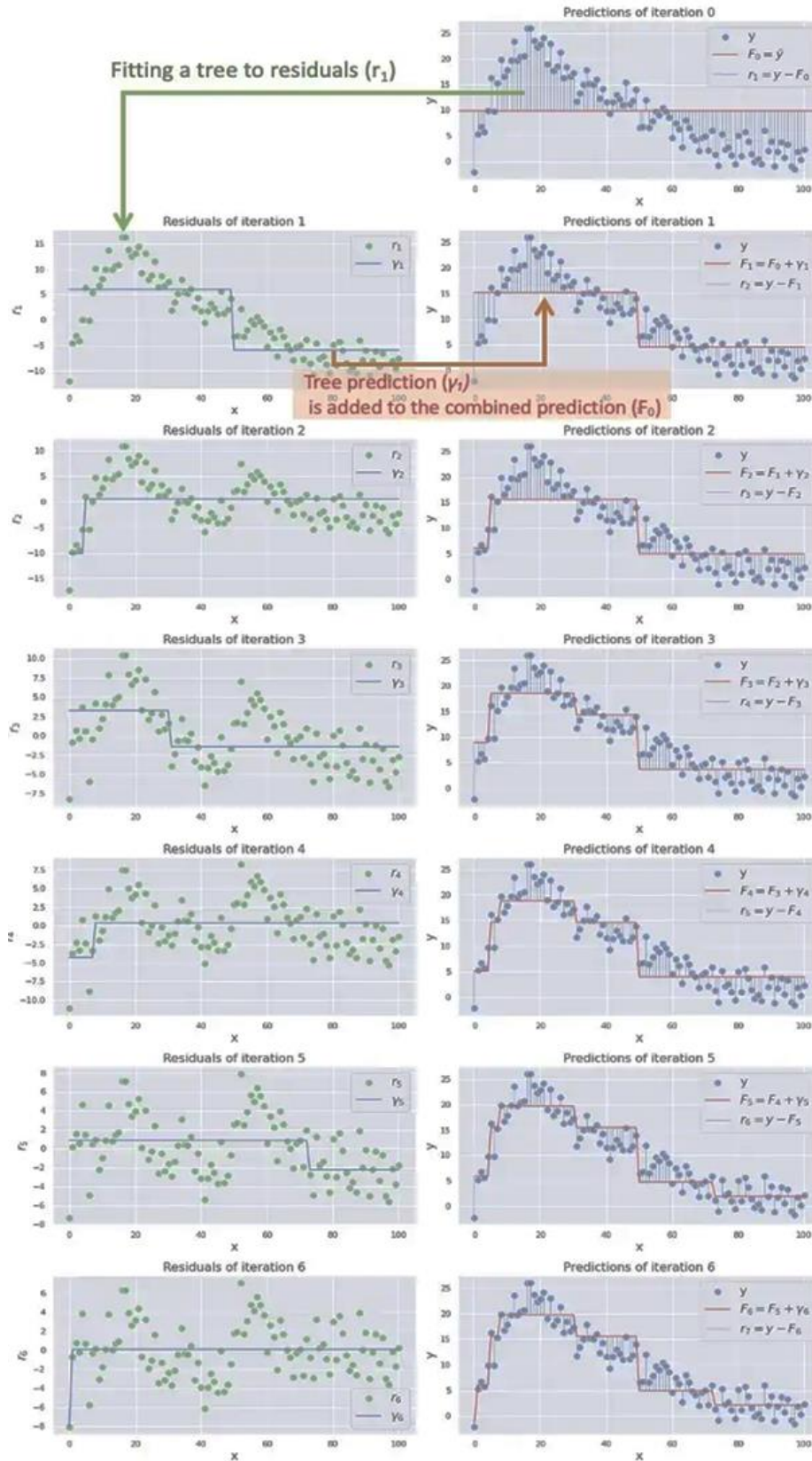


Figura 2.8 – Processo de otimização - Gradient *Boosted Decision Trees* (GBDT)

Apesar de conceitos econômicos tradicionais e das aplicabilidades do custo de oportunidade serem norteadores para esta pesquisa, um estudo simples foi encontrado, onde a predição de dados operacionais e financeiros de manutenção servem para a análise do custo de oportunidade trade-off em estratégias de manutenção ferroviária. Portanto, é necessária uma análise detalhada das compensações do custo de oportunidade para determinar a tomada de decisão (trade-off) mais adequada.

Para promover a relação do uso de aprendizado de máquina e a análise trade-off, a busca pelos artigos científicos utilizados para a construção deste referencial teórico, se deu através da base de dados Scielo, ResecherGate e Scholar Academic, onde foram selecionados artigos e tese com as palavras chaves: “*opportunity cost (trade off) for railway maintenance*” e “*machine learning approach to predicting rail asset maintenance*” que atendiam os objetivos do estudo. Segue o Quadro 2.5, onde constam os 15 artigos principais embaixadores deste estudo.

Quadro 2.5 – Principais artigos embaixadores para o estudo de aprendizado de máquina e análise trade-off

Nome do Artigo	Ano	Autores
Contabilidade e Administração	1977	Gray, J. and Johnston, K.S.
Custo de oportunidade: conceitos e contabilização	1990	Anísio Cândido Pereira et al.
A critical review of the opportunity cost concept	1999	David S. O. Yip
A importância econômica das ferrovias para o Brasil	2002	Márcio Rogério Silveira
Cost Analysis of Maintenance and Non-maintenance Policies for Productive Equipment	2003	W.R.Marcorin; R.C.Lima
Cost-effectiveness of railway infrastructure renewal main	2006	G. A. Grimes; C.P. L. Barkan
Operations Research Models for Scheduling Railway Infrastructure Maintenance	2009	Gabriela Budai-Balke
Logistics Opportunity Costs: a mining case study	2013	Leani Van Jaarsveld et al.
Logistics opportunity costs: a mining case study	2013	Gert J. Heyns; Peter J. Kibourn
Railway infrastructure maintenance – a survey of planning problems and conducted research	2015	Tomas Lidén
A modeling framework for railway infrastructure reliability analysis	2017	C. J. Fourie; N. T. Zhuwaki
Ensemble Methods in Machine Learning: What are They and Why use They?	2017	Evan Lutins
Optimal scheduling of track maintenance activities for railway networks	2018	Zhou Su; Bart De Schutter
A machine learning approach for maintenance prediction of railway asset	2018	Zaharah A. Bukhsha et al.
The recent applications of machine learning in rail track maintenance: a survey	2019	Muhammad C. Nakhaee et al.
A Quantitative Ex Post Facto Study for one-to-one mobile technology	2019	Sohil Sharma
Gradient-Boosted Decision Trees for Multiple	2019	Z.Zhang; Cheolkon Jung
Prediction of geometry deviations in additive manufactured parts	2020	I. Baturynska; K.Martinsen
Trade – off entre custo e tempo em projetos	2021	R.Thielmann; K. Queiroz

CAPÍTULO 3

MATERIAIS E MÉTODOS

3.1 - MATERIAIS

3.1.1 - Objeto de Estudo

Historicamente, os grandes projetos de exploração mineral na região amazônica, no estado do Pará, privilegiaram a produtividade atrelada às escalas geográficas da economia mundial. Neste quesito, os empreendimentos na região caminham para uma economia espacial dos processos produtivos e logísticos, enfatizando as relações intermodais desses processos de produção e circulação do capital, especialmente priorizando o ganho de eficiência no transporte de cargas com as ferrovias.

A extensão ferroviária na região norte do país, segundo a Agência Nacional de Transportes Terrestres (2021) é de aproximadamente 2372 Km, o que equivale a cerca de 11% da malha brasileira, correspondente às ferrovias concessionadas: Norte-Sul Tramo Norte, 744,5 Km; Rumo Malha Norte S.A., 735,3 Km e a VALE - Estrada de Ferro Carajás, 892 Km.

O objeto deste estudo refere-se aos custos de manutenção de uma Estrada de Ferro com significativo volume de transporte de minérios para exportação. Além de ligar a maior mina do Brasil, na região de Carajás – Estado do Pará até o porto da madeira no Estado do Maranhão (VALE, 2018) (Fig. 1).

Por seus trilhos, são transportados 171,2 milhões de toneladas quilometro-útil (TKU) de minério de ferro (CNT, 2022) e 350 mil passageiros por ano. Circulam cerca de 35 composições simultaneamente, entre os quais um dos maiores trens de carga em operação regular do mundo, com 330 vagões e 3,3 quilômetros de extensão (Vale, 2022).



Figura 3.1 - Rede Ferroviária na Amazônia Oriental.

3.1.2 - Delimitação do Estudo

Os indicadores econômico e financeiro e custos de manutenção referidos à Ferrovia estudada correspondem as demonstrações financeiras *Carve Out* presentes na base de dados da Agência Nacional de Transporte Terrestre – ANTT. Os dados operacionais foram obtidos no Anuário de Pesquisa da Confederação Nacional de Transportes – CNT.

Os custos com manutenção de áreas industriais e de ativos relevantes, incluindo os trilhos, são registrados no ativo imobilizado e depreciados. Os ativos imobilizados são depreciados com base nas vidas úteis dos bens definidas na Resolução nº 4540 de 19 de dezembro de 2014 emitida pela Agência Nacional de Transportes Terrestres - ANTT, correspondente a uma vida útil de 10 anos a 20 anos.

A presente gestão classifica os ativos financeiros com base no seu modelo de negócios para o gerenciamento dos ativos e nas características dos fluxos de caixa contratuais desses ativos. No caso dos serviços de manutenção, a receita de carga geral e de passageiros é reconhecida quando a concessionária transfere para seus clientes todos os riscos e benefícios significativos referentes aos serviços prestados.

Para a receita de transporte próprio, a concessionária apropria à ferrovia uma receita com base em preço de transferência para o minério próprio transportado. Adicionalmente a ferrovia estudada dispõe de recursos financeiros (disponibilidades) junto a concessionária, os quais são utilizados para financiar suas operações e realizar os investimentos necessários.

A concessionária não fornece uma tabela de manutenção para a via permanente, como faz para a manutenção de vagões e locomotivas. A aferição das manutenções preventiva e corretiva foram formuladas com base nos indicadores econômico-financeiros.

O custo com manutenção corretiva deve ser menos de 2% de sua manutenção total, visto que é altamente ineficiente e cara (normalmente pelo menos 3 a 5 vezes mais cara que uma manutenção preventiva bem planejada) (HUPJE, 2020). Lembrando que as manutenções preventiva e corretiva, quando necessário, acontecem simultaneamente.

As orientações neste trabalho sobre a conformidade e mantabilidade seguem a Norma NBR 5462, para que haja um entendimento da perda de lucro ou lucro cessante nas estratégias de manutenção utilizadas para o período estudado. Gray & Johnston (1977) concluiu que a partir da aferição do impacto que o lucro cessante tem sob os custos, é possível identificar a decisão que foi tomada.

O estudo do custo de oportunidade no referido trabalho será de caráter qualitativo para a predição a curto prazo seguindo a segunda condição de Yip (2014) e Sizer (1974) onde inclui a qualificação para o atributo escolhido como caracterizador do custo de oportunidade, uma qualificação do tipo “melhor”, “piores”, etc., ao dizer que custo de oportunidade concerne a melhor alternativa da qual se desistiu, tornando o conceito compacto e deixa a cargo do usuário caracterizar o que lhe é conveniente. Para esboçar a delimitação do estudo, segue o fluxograma demonstrativo na Figura 3.2.

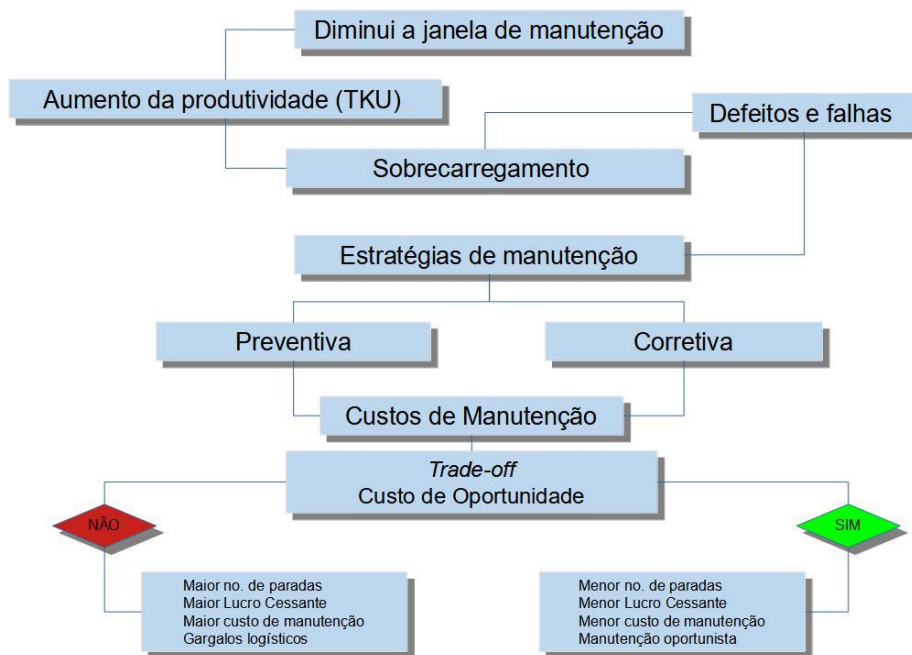


Figura 3.2 - Fluxograma demonstrativo da delimitação do estudo

3.2 - MÉTODOS

3.2.1 - Análise do banco de dados: variáveis (input and output)

Realizou-se a apropriação de custos, despesas e características operacionais para a manutenção de uma ferrovia na Amazônia Oriental, que foram agrupados em três parâmetros de informações em um banco de dados (Tabela 3.1). Os dados selecionados correspondem a média dos meses para o cada horizonte de estudo entre 2012 e 2021, são eles:

- Dados operacionais e de produtividade – refletem o “esforço” para a produção de um transporte por concessionária, coletados junto a CNT – 2022.
- Dados econômico-financeiro (apropriação dos custos e despesas calculadas e estimadas) – refletem os valores despendidos em cada natureza de gasto, coletados no Anuário da Estrada de Ferro Carajás 2012 – 2021.

Tabela 3.1 - Dados operacionais e de produtividade, econômico-financeiros e custos de manutenção.

Ano	Total de Carga Transportada	Total de Carga por Quilômetro Útil	Número de Paradas para Manutenção	Custo de Pessoal de Serviços (R\$)	Lucro Cessante (R\$)	Custo de Manutenção Preventiva (R\$)	Custo de Manutenção Corretiva (R\$)
2012	116.428.388	102.157.360.107	100	1.898.208,00	180.813.000,00	1.158.190,80	181.971.190,80
2013	115.005.956	101.010.500.150	104	5.072.112,00	344.994.000,00	3.397.488,10	348.391.488,10
2014	118.454.346	101.177.266.706	107	3.002.076,00	333.338.000,00	1.880.127,90	335.218.127,90
2015	139.695.191	118.593.600.192	112	6.386.914,80	736.476.000,00	2.580.460,82	739.056.460,80
2016	155.252.299	136.268.213.635	127	3.432.000,00	2.224.160.000,00	1.913.866,00	2.226.073.886,00
2017	175.332.542	155.577.535.773	122	2.703.652,80	1.237.678.000,00	4.789.908,23	1.242.467.908,00
2018	205.406.565	184.375.642.877	130	3.506.220,00	3.061.104.000,00	7.232.352,37	3.068.336.352,00
2019	193.571.861	174.485.955.732	117	5.138.880,00	1.151.817.000,00	5.158.468,80	1.156.975.469,00
2020	195.737.931	177.376.810.365	125	7.434.240,00	1.887.353.000,00	3.703.111,83	1.891.056.112,00
2021	190.300.204	172.084.973.708	109	8.321.580,00	419.861.000,00	6.769.555,20	426.630.555,20

Verificou-se que nos anos de 2012 a 2015 o custo com manutenção preventiva esboçou um cenário onde o número de paradas para manutenção não foi expressivo para manutenções corretivas. A partir do ano de 2016 houve um expressivo número de paradas que inferiram em custos maiores com manutenção corretiva, chegando em 2018 no vértice maior de número de paradas para manutenções corretivas e preventivas. No ano de 2021, o custo com manutenção preventiva retomou uma posição mais frequente, enquanto o custo com manutenção corretiva teve uma queda abaixo da média dos anos anteriores. Notou-se que o número de paradas acompanhou reduziu significativamente ao ano anterior (Gráfico. 3.1).

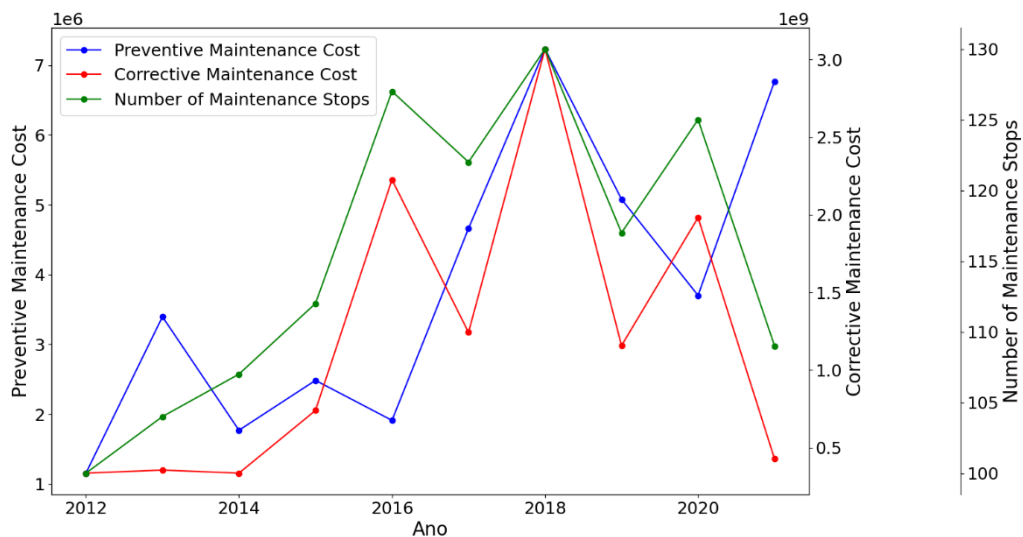


Gráfico 3.1 - Número de paradas não linear para manutenção de trilhos (corretiva e preventiva) no período de 2012 a 2021.

Na base de dados identificou-se os custos de serviço e material e o custo de mão de obra, ambos não foram considerados pois já estão inseridos nos custos de manutenção preventiva e corretiva.

No tratamento das respectivas variáveis operacionais e de produtividade e econômico-financeiras, a velocidade comercial e extensão ferroviária foram consideradas constantes para a série anual. O investimento na via permanente não foi considerado para o modelo por dois motivos: o primeiro, que o investimento tipo OPEX, referente as despesas operacionais, são previstos obrigatoriamente em segurança e o aumento de capacidade, pactuados junto a Agência Nacional de Transportes Terrestres (ANTT), não favorecendo a análise quando se trata de investimento em manutenção especificadamente, e o segundo por não se tratar de um estudo econômico quantitativo, a decisão da aplicação anual do investimento em determinada atividade é de caráter subjetivo.

O lucro cessante foi considerado como variável dependente por impactar substancialmente no custo de manutenção corretiva, além de sofrer influência de outras variáveis independentes, como por exemplo, o expressivo número de paradas para manutenção em 2018.

Neste estudo, os dados foram separados em três partes: indicadores operacionais e de produtividade, dados econômico-financeiros e custo de manutenção corretiva e preventiva. Entre eles, 3 variáveis dependentes output (y): Lucro cessante, custo de manutenção preventiva e corretiva; e 5 independentes input (x): total de carga transportada, tonelada por quilômetro útil, número de paradas, custo pessoal de serviço e ano.

3.2.2 - Modelagem e Método *Gradient Boosting Regression*

O procedimento inicial é a depuração do banco de dados, na classificação das variáveis para dar entrada no método e, conseqüentemente, a correlação dessas variáveis que vão gerar uma predição para a análise trade-off, seguem as etapas (Fig.3.3):

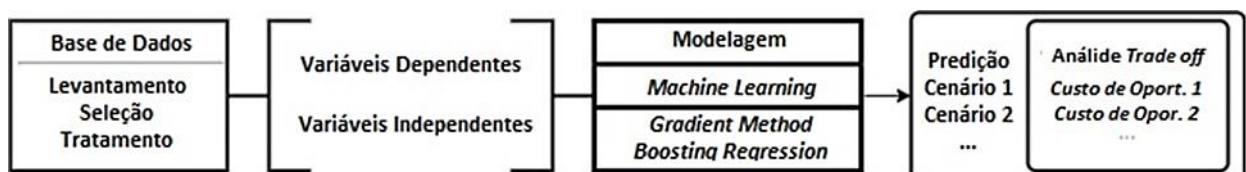


Figura 3.3 - Fluxograma da Metodologia de Pesquisa.

A metodologia adotada propõe o estudo do *trade-off* custo de oportunidade a partir da modelagem de dados referentes a indicadores operacionais e de produtividade, dados econômico-financeiros e custo de manutenção corretiva e preventiva para 892 quilômetros de extensão de uma ferrovia que transporta minério no Amazônia brasileira, que se baseia em dados de 2012 a 2021.

A abordagem metodológica segue os seguintes requisitos:

- Os dados ferroviários devem ser coletados para o período (2012-2021);
- As variáveis devem ser tratadas e classificadas como dependentes e independentes;
- Será determinado o coeficiente de determinação R_2 dos grupos de treino e teste para as variáveis dependentes;
- A modelagem deve mostrar o grau de correlação e importância entre as variáveis dependentes e independentes;
- O algoritmo *Gradient Boosting Regression Tree* deve apresentar cenários de previsão de curto prazo (2022);
- A análise de *trade-off* será estabelecida a partir de um nível de conflito, utilizando nomenclatura binomial (ótimo, bom, razoável e ruim) para comparar os benefícios encontrados entre as alternativas oferecidas para cada cenário.
- O estudo de custo de oportunidade levará em consideração os melhores benefícios para previsão de curto prazo.

3.2.3 - Aprendizado de máquina

O conjunto de dados é dividido em dois grupos, um para treinamento, que serve para criar o modelo para ajuste dos dados pelo GBRT, e outro para teste, que avalia a capacidade preditiva do modelo obtido. A avaliação do sucesso do ajuste pode ser feita através do coeficiente de determinação (r_2), que deve ser igual a 1 para o grupo de teste ou o mais próximo possível. Seguem as etapas de modelagem:

1. Defina os valores de entrada: $x_1, x_2, x_3 \dots x_n$ (variáveis independentes).
2. Divida os valores de entrada em grupos de treinamento e teste.
3. Definir os parâmetros ideais para que o grupo de teste seja previsto com erro mínimo do algoritmo treinado pelo grupo de treinamento (Execução de Machine Learning (n... n +1... n+2 vezes) até encontrar os melhores resultados)
4. Avalie o ajuste usando o coeficiente de determinação (r_2), que deve ser o mesmo para o grupo de teste ou o mais próximo possível.
5. O código prevê os valores de saída: $y_1, y_2, y_3 \dots y_n$ (variáveis dependentes)

6. O número de estimativas determina o número de árvores de decisão utilizadas e o número mínimo de dados para divisão em folhas e a profundidade máxima dessas árvores.

Para o uso do GBR é necessário definir as variáveis independentes e uma variável dependente da Estrada de Ferro no período entre 2012 e 2021 (Anexo 1 - Tutorial). As independentes foram: o ano, o total de carga transportada, o total de carga por quilometro-útil, o número de paradas para manutenção e o custo do pessoal de serviços. As variáveis dependentes foram: o lucro cessante, o custo de manutenção preventiva e o custo de manutenção corretiva. Como são três variáveis, o algoritmo precisou ser utilizado separadamente para cada uma (Fig. 3.4).

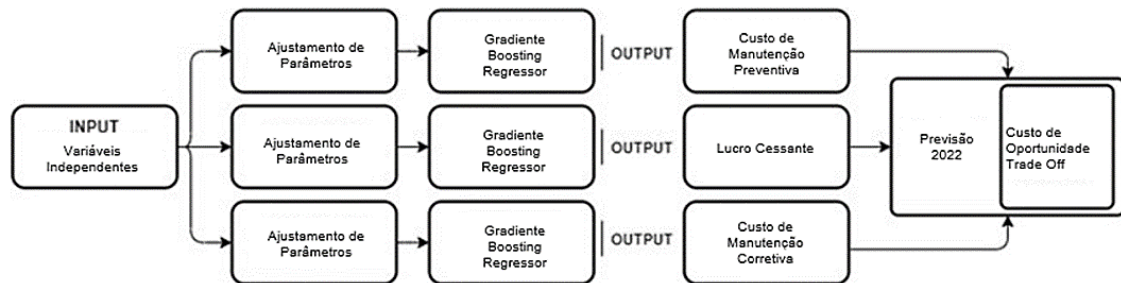


Figura 3.4 - Diagrama do modelo *Gradient Boosting Regression*.

Nos três casos, o conjunto de dados é dividido em dois grupos, um de treino, que serve para a criação do modelo de ajuste dos dados pelo GBR, e um de teste, o qual avalia a capacidade preditiva do modelo obtido. A avaliação do sucesso do ajuste pode ser feita por meio do coeficiente de determinação (r^2), no qual deve ser igual à 1 para o grupo de teste e o mais próximo possível para o grupo de teste.

Para garantir esses resultados, é preciso definir os parâmetros utilizados pelo algoritmo, o que é feito de forma automatizada por meio da execução de um código à parte, o qual busca o valor de r^2 mais próximo dos modelos resultantes de várias execuções do GBR. Cada uma delas é feita com leves alterações aleatórias nos parâmetros utilizados, esse fato evita que o código fique “preso” em mínimos locais para r^2 e encontre valores melhores (Fig. 3.5). Este código foi utilizado para determinar a porcentagem da divisão dos dados para os grupos de treino e teste, em que a taxa de aprendizado determina o peso de cada nova iteração no modelo final, o número de estimadores determina a quantidade de árvores de decisão utilizadas e o número mínimo de dados para divisão, além do número mínimo de dados nas folhas e a profundidade máxima destas árvores.

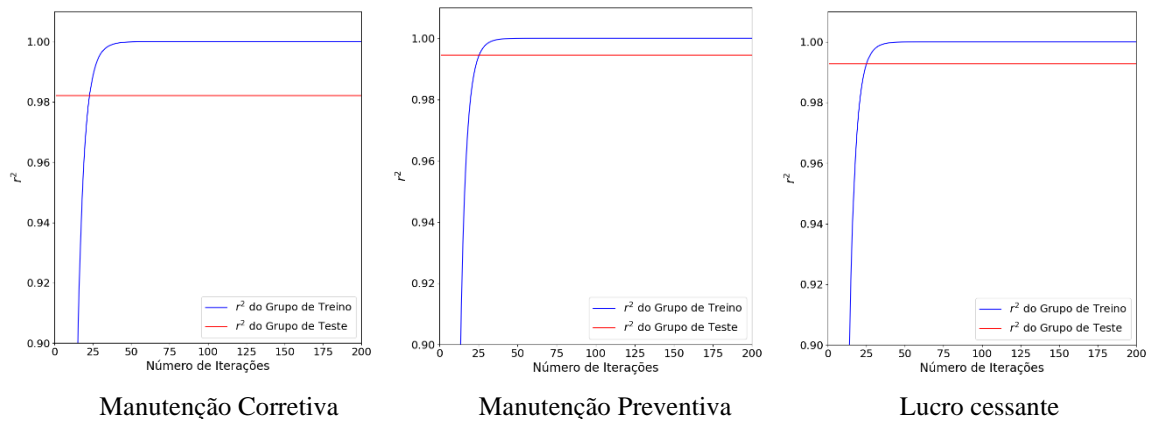


Figura 3.5 - Dados de treino e teste: Coeficiente de Determinação dos grupos de treino e de teste para o custo de manutenção corretiva (a), preventiva (b) e lucro cessante (c).

O coeficiente de determinação para o grupo de variáveis de treino foi igual a 1, e as variáveis do grupo de teste foram de 0,99, aproximadamente 1, com exceção da variável de teste do custo de manutenção corretiva apontando o equivalente a 0,98; ainda assim é considerada satisfatória.

Após esse procedimento, o GBR foi implementado para as três variáveis dependentes, o que permitiu gerar um modelo capaz de prever dados futuros a partir do fornecimento das variáveis independentes.

CAPÍTULO 4

RESULTADOS E DISCUSSÕES

4.1 - APLICAÇÃO DO APRENDIZADO DE MÁQUINA

4.1.1 - A escolha do Modelo

A proposta da análise feita neste trabalho consiste em obter previsões para as variáveis escolhidas como dependentes, de modo que é necessário usar algum método de regressão. Basicamente, este tipo de método é uma técnica de modelagem preditiva que busca determinar a relação entre variáveis dependentes e independentes, o que possibilita previsões para valores fora do conjunto de dados iniciais.

Dentre os métodos de regressão, o mais simples é o linear, que basicamente ajusta uma reta aos dados. Porém, ele é inadequado para dados que não apresentem um comportamento linear. Uma alternativa simples é a regressão polinomial, que gera modelos com erros menores do que a linear para casos não-lineares, porém fica restrita a conjuntos de dados com comportamento semelhante a funções polinomiais. Outro problema é que esses métodos não são aplicáveis diretamente a sistemas com várias variáveis independentes.

Como alternativa, podem ser utilizadas as árvores de decisão (Decision Trees – DT), as quais usam sequências de decisões que determinam a estimativa final para as variáveis dependentes. Por exemplo, se a primeira variável tiver valor menor que 10, então o algoritmo verifica se o valor da segunda variável tem valor maior que 5, se tiver o algoritmo estima a variável dependente como 3. Ou se a primeira tem valor menor que 10, mas o valor da segunda é menor que 5, então a estimativa é 2. Basicamente, o algoritmo determina as condições que o levam a fazer a estimativa final de forma mais precisa possível.

Um problema com esse método é que ele tende a apresentar *Overfitting*, basicamente quando o ajuste é perfeito para os dados de entrada, mas as previsões são ruins. Este método também tem problemas para extrapolação, o que significa que ele não tem boas previsões para dados fora do conjunto fornecido. Uma alternativa é a sua implementação como parte de métodos de conjunto (*Ensemble Methods*), os quais combinam vários estimadores, basicamente usam vários métodos mais simples, para obter um estimador mais completo e com maior capacidade de generalização.

Esta abordagem minimiza os dois problemas citados, overfitting e extrapolação, e existem dois métodos de conjunto principais os Averaging Methods (AM) e os Boosting Methods (BM). No primeiro, os estimadores são independentes e suas estimativas são usadas em médias ponderadas para gerar a estimativa final para a variável dependente. No segundo, os estimadores são utilizados em sequência, com cada um melhorando o resultado do anterior, para gerar a estimativa final.

Entre os métodos de conjunto que usam as árvores de decisão, como AM há o algoritmo de floresta aleatória (Random Forest - RF) e como BM há o Gradient Boosted Decision Trees (GBDT). No geral, o GBDT tem maior precisão por minimizar o erro do ajuste dos dados a cada nova árvore, porém ele tende a apresentar overfitting quando há muito ruído nos dados. Isso acontece quando há muitos dados sem correlação forte entre si, o que não é o caso dos dados do trabalho. Como o objetivo é a regressão de dados, o algoritmo do GBDT usado é o Gradient Boosting Regressor (GBR). Caso o objetivo fosse a classificação, seria utilizado o Gradient Boosting Classifier (GBC), sendo ambas implementações do GBDT.

Em resumo, a escolha do GBR foi feita pelas características dos dados disponíveis, que apresentam um comportamento complexo, difícil de ser ajustado por métodos mais simples de machine learning, como lineares e polinomiais. Como as DT são úteis para esses casos, a implementação poderia ser por RF ou GBDT. Este foi escolhido por ser mais preciso.

4.1.2 – Características da variáveis

As variáveis são discretas, pois tem valores definidos específicos (dez valores, no caso). Uma variável contínua, por outro lado, pode assumir qualquer valor.

As variáveis são obtidas de dados estimados a partir da literatura, segundo o que me mandaste, então não são randômicas, no sentido de ter uma correlação entre elas originada de dados reais reportados nas tuas fontes. Não necessariamente são determinísticas, pois os valores dependem de muitos fatores adversos, quando se leva em consideração a estrutura operacional de uma empresa. Por causa disso, não é possível afirmar com certeza se dado um conjunto de valores, os resultados serão exatos, como seria com variáveis determinísticas. Quanto a serem valores médios ou a mediana, depende da origem dos dados. Pelo que entendi, são dados anuais da operação de uma empresa, então são só os valores totais e não uma média, por exemplo.

Foram classificados os três grupos de variáveis:

- indicadores operacionais e de produtividade,

- dados econômico-financeiros e,
- custo de manutenção corretiva e preventiva.

Entre eles, 3 variáveis dependentes output (y): Lucro cessante, custo de manutenção preventiva e corretiva; e 5 independentes input (x): total de carga transportada, tonelada por quilômetro útil, número de paradas, custo pessoal de serviço e ano.

A variável Ano (year) foi utilizada em conjunto com as outras variáveis de entrada e o próprio código mostrou a variação de todas com o ano, como é esperado. Após isso, ela foi usada como valor do eixo x com os outros valores no eixo y, pois essa é uma abordagem comum para avaliar as variáveis dependentes e as independentes, já que é mais fácil entender um gráfico em função do tempo.

A variável de resposta y pode vir de qualquer conjunto de dados de entrada, porém sempre será necessário treinar novamente o modelo. Desta forma o modelo fica condicionado aos dados utilizados. Neste processo, é necessário identificar novamente os parâmetros ótimos para que o resultado seja o mais preciso possível.

Para o caso deste trabalho, apesar das variáveis serem “nomeadas” seus valores são todos numéricos, então temos algo do tipo [custo em 2020 = x reais] e nunca algo do tipo [custo em 2020 = caro], ou seja, os valores nunca são expressos por palavras. Na prática, o código ignora o nome das variáveis e considera apenas a organização das tabelas de valores para definir cada uma através das posições dos dados. Em resumo, não há um tratamento dos dados além de sua organização no formato observado nos arquivos “dados.txt” e “2022.txt” utilizados no código.

4.1.3 – Funções de aprendizado aplicadas as *splines*

Durante a execução do *Gradient Boosting Regressor* pelo pacote *scikit learn*, as funções utilizadas foram por partes, aproximadas cada vez mais dos dados através da redução do erro de uma função inicial, a qual é basicamente o valor médio da variável de saída, como mostrado na iteração 0. Neste primeiro caso, qualquer estimativa para a variável dependente seria sua média, que é a função de partida. Nas iterações seguintes, a função do modelo se torna cada vez mais próxima dos valores dos dados, levando a previsões progressivamente melhores, as quais são garantidas quando o algoritmo busca minimizar os erros resultantes.

Apesar deste tipo de aproximação ter o problema de dar o mesmo resultado para variáveis de entrada próximas, quando elas estão na mesma reta da função, quando há muitas iterações isso é minimizado, também levando em consideração as demais especificidades do *Gradient Boosting Regressor*, como o *learning rate*, a função de perda etc.

A utilização de *splines* pode levar a melhores ajustes de dados, porém essa não é uma opção simples de se implementar com o pacote *scikit learn*, com o GBR já construído com o uso de funções por partes.

4.1.4 – Nível de Entropia e Ganho de Informação

O nível de entropia está associado à imprevisibilidade do sistema. Por exemplo, em problemas de classificação a entropia é grande quando os dados de entrada estão muito misturados, tornando difícil a sua separação e, por consequência, sua classificação por árvores de decisão. No geral, este conceito é relevante neste tipo de problema e tem menor importância em problemas de regressão, sobretudo com poucos valores, pois os dados de entrada já são bem específicos e não estão sendo divididos em grupos, visando a menor impureza.

O ganho de informação é definido como 1 menos a entropia (links acima), de modo que durante a divisão dos dados nas árvores de decisão, quanto mais são reduzidas as impurezas, isto é, quanto mais dados iguais são agrupados sem dados diferentes, mais a entropia diminui e maior é o ganho de informação. Novamente, esse é um conceito relevante para problemas de classificação.

4.1.5 - O Método do Gradient: considerações gerais

Basicamente, o GBR busca minimizar o erro da previsão dos valores do grupo de teste com o algoritmo treinado com o grupo de treino. Neste ponto, cada conjunto de parâmetros do algoritmo leva a erros diferentes. Para minimizar o erro é executado inicialmente o código para ajustar esses parâmetros, o que basicamente significa encontrar os parâmetros que levem a erros menores. Para isso, um conjunto de parâmetros inicial é testado, o próximo varia um pouco os parâmetros e depois o código avalia se o erro aumenta. Caso aumente, o código descarta os parâmetros e testa o próximo conjunto com pouca variação. Quando o erro diminui, esse conjunto passa a ser o de referência e os próximos parâmetros são obtidos variando este, até um apresentar um erro menor. Nesse contexto, mínimos locais são conjuntos de parâmetros cujo erro não é o menor possível, porém o código não consegue obter erros menores com poucas variações dos seus parâmetros.

Para contornar esse problema, o código usado adiciona uma variação aleatória nesse processo, a qual aleatoriamente faz o código encontrar erros menores ao apresentar parâmetros bem distintos do que os que levam à mínimos locais. Este problema é contornado através da execução do código muitas vezes até encontrar valores aceitáveis. Além disso, o erro considerado foi o r^2 o que nos permite dizer que valores próximos a 1 são ideais. Com isso,

mesmo que erros sejam pequenos (próximos a 1), podendo ser apenas mínimos locais, desde que sejam erros aceitáveis, já são suficientes, de modo que os resultados consideram essa margem de erro como o valor de r^2 .

O Gradient Boosting Regressor é um método de conjunto (Ensemble Method) de regressão baseado no Boosting Method. Como apontado nas documentações do pacote *scikit learn*, este tipo de método é de aprendizado supervisionado. Basicamente, neste método são utilizadas variáveis “rotuladas”, isto é, identificadas na execução do código. Por exemplo, as variáveis utilizadas neste trabalho foram todas nomeadas, como ano, número de paradas etc.

No caso do aprendizado não supervisionado, as variáveis não têm identificação e o método serve para identificar correlações mais “escondidas”, sem a supervisão humana (nomeando as variáveis, basicamente).

4.1.6 - Restrições na execução e a Função de perda

Não foram aplicadas restrições na execução do GBR por dois motivos: o conjunto de dados de entrada foi bastante simples e já possuía indícios de correlação com as variáveis de saída; e o pacote utilizado em *python* não dá suporte direto à utilização de restrições (Constraints). Basicamente, elas serviriam para acelerar o código, excluindo resultados irrelevantes como, por exemplo, estimativas negativas para os custos analisados.

A função de perda (*Loss Function*) utilizada foi o erro quadrático médio (Mean Square Error), a qual é o padrão no pacote utilizado para implementar o GBR e não foi definida diretamente no código. Esta função de perda considera o erro quadrático entre o valor real e o valor previsto pelo algoritmo, de modo que ela penaliza estimativas que diferem muito do valor correto ao elevar o erro ao quadrado. Isso atrapalha no ajuste de dados com comportamento bem definido, pois sempre haverá dados fora da curva ajustada, os quais apresentam alto erro quadrático. Porém, para este trabalho, esta função de perda ajuda a garantir um melhor ajuste dos dados de treino, para que sejam feitas previsões melhores dos dados de teste e, sobretudo, das previsões fora dos dados fornecidos, como feito para 2022, as quais são avaliadas pelo r^2 .

Não foi possível aplicar a metodologia de Gradient Boosted em uma dataset já convalidado, a fim de avaliar o seu algoritmo, porque não tem-se um conjunto de dados (dataset) já convalidado, uma vez que nosso próprio conjunto de dados foi estimado justamente por não termos acesso a dados mais completos, mesmo na literatura. De certa forma, a aplicação do algoritmo nos grupos de treino e de teste serve para isso. O grupo de treino gera o modelo e o de teste avalia a sua precisão por meio do r^2 .

4.1.7 – Estimação do erro e Acurácia

Uma medida precisa apresentar vários valores próximo, mas não necessariamente exatos. Uma medida com alta acurácia apresenta valores mais próximos dos exatos, mas não necessariamente precisos.

Por exemplo, se o valor exato é cinco e as estimativas são 8,1; 7,9; 8,0 e 8,2; o modelo que previu esses valores é preciso, porém tem baixa acurácia, pois os valores estão próximos entre si, mas longe do exato. Se o valor exato é cinco e as estimativas são 4; 6; 3 e 7; então a medida tem acurácia, pois está em torno do valor exato, porém não tem precisão porque os dados estão longe entre si.

Cada novo estimador modifica o resultado do anterior e a taxa de aprendizado é o peso desta modificação. Basicamente, ela varia de 0 a 1 (sem contar o zero) e altera a “intensidade” desta modificação. Por exemplo, se o novo estimador tem como resultado “adicionar 5 no valor anterior” e a taxa de aprendizado é 0.1, então será adicionado 0.5 ao valor do estimador anterior. Isso torna necessário mais estimadores, porém melhora a precisão do resultado.

Dito isso, podemos considerar que o r^2 avalia ambos, pois ele tende a 1 (melhor valor) quando os valores previstos para o grupo de teste são próximos aos valores reais, isto é, quando a previsão tem boa acurácia, o que indica que o modelo construído pode representar os dados originais com boa precisão. Neste sentido, se os valores previstos pelo modelo diferem muito dos valores reais, mesmo que os valores tenham boa precisão, isto é, estejam próximos uns dos outros, eles apresentarão baixa acurácia e, por consequência, apresentarão um valor baixo de r^2 . Inversamente, quando os valores de r^2 estão próximos de 1, como neste trabalho, os valores previstos estão com alta acurácia e devem estar, também, com alta precisão, pois isto também reduz o valor de r^2 . Na prática, o modelo gerado através do Gradient Boosting Regressor, ajusta perfeitamente os dados de treino (com $r^2 = 1$) e obtém alta precisão e acurácia, indicados pelo r^2 maior que 0,9 para os dados de teste. Isto indica que o modelo deve ser capaz de fazer previsões razoáveis ligeiramente fora de seus dados fornecidos, como foi possível observar para as previsões do ano de 2022, as quais não foram absurdamente altas ou baixas, se comparadas com a tendência dos valores observados nos gráficos.

4.1.8 - Correlação de Pearson para as variáveis

Foi determinada a relevância de cada variável independente para o valor final das variáveis dependentes, de acordo com os modelos obtidos. Este ponto foi validado pelos resultados obtidos por uma matriz de correlação de todas as variáveis, feita a partir dos dados

fornecidos ao algoritmo que identifica o coeficiente de correlação de Pearson de cada variável com relação às demais, cujo valor 1 indica proporcionalidade direta, 0 indica correlação inexistente e -1 proporcionalidade inversa.

Os coeficientes de correlação encontrados estão relacionados as seguintes variáveis: CMC (Custo de Manutenção Corretiva), PMC (Custo de Manutenção Preventiva), OP/L (Lucro Cessante), P.S (Pessoal de serviço), NMS (Número de paradas), TKU (Tonelada por quilômetro útil), TU (Total de carga transportada) e YEAR (ano) (Figura 4.1).

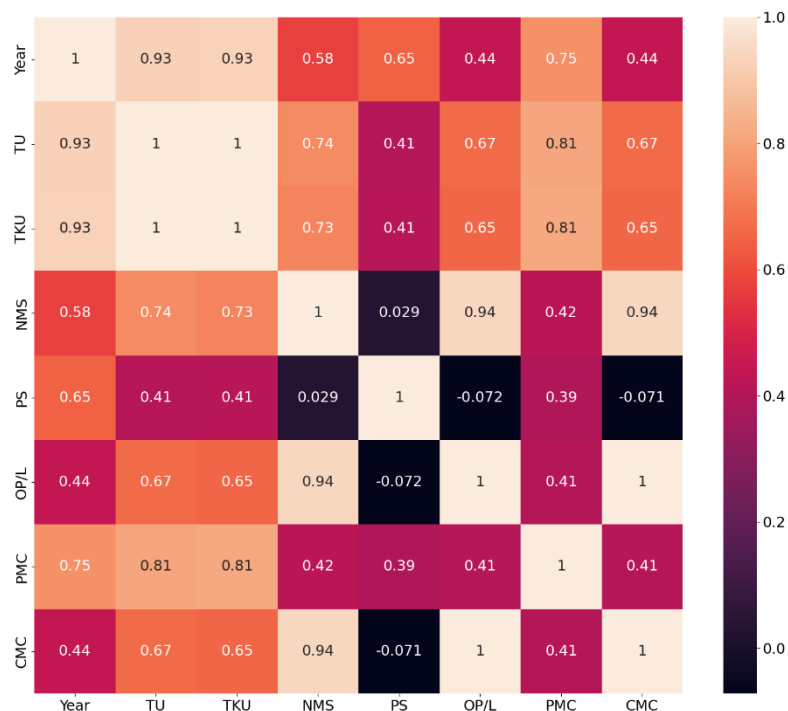


Figura 4.1 - Variáveis analisadas na matriz de correlação

As correlações de Pearson estão divididas em 2 variáveis de previsão para os custos de manutenção, entre elas estão:

1. Custo de Manutenção Preventiva: TKU = 0,81.
2. Custo de Manutenção Corretiva: Lucro cessante = 1; Número de paradas = 0,94. O lucro cessante teve baixa correlação referente ao Custo de Manutenção Preventiva, 0,41, assim como o número de paradas, 0,42.

A matriz de correlação apontou que o número de paradas foi crucial para impactar no lucro cessante e, conseqüentemente, para o custo de manutenção corretiva. Vale lembrar que o lucro cessante é um custo indireto que compõe o custo de manutenção corretiva e tem relação direta com os custos de risco, Nielsen et al.; (2013) identificou que existe um custo de risco

devido à manutenção inadequada levando a casos extremos, descarrilamentos, perda de bens, perda de receitas e pagamentos de indenizações.

A análise feita com a matriz de correlação utilizou todas as variáveis ao mesmo tempo. Os algoritmos do *machine learning* foram feitos separadamente, com o mesmo código e mudando apenas os parâmetros e cada variável dependente. Estamos considerando que as variáveis dependentes são independentes entre si, variando apenas com as variáveis independentes (input), cada uma com seu comportamento em particular (e, logo, com padrões individuais).

O modelo *Gradiente Boosting Regression* determinou valores de importância para cada variável e a Figura 4.2 mostra a proporcionalidade desses valores. As relações não foram linearizadas. Nesta análise, o valor de cada variável dependente depende do conjunto de variáveis de entrada (independentes), de modo que cada uma delas contribui para o valor final da previsão da variável dependente. Sendo assim, a importância de uma variável independente mostra basicamente quantos por cento do valor final é determinado por ela. Por exemplo, se há uma variável de entrada que não tem nenhuma relação com a variável dependente, a sua importância será nula, pois o valor da previsão não é alterado pelo seu valor. Outra forma de entender é se uma variável tem muita importância, sua variação muda muito o valor final da previsão, como foi o caso do número de paradas para manutenção para o lucro cessante.

O modelo *Gradient Boosting Regression* determinou valores de importância para cada variável e a Figura 4.2 mostra a proporcionalidade desses valores:

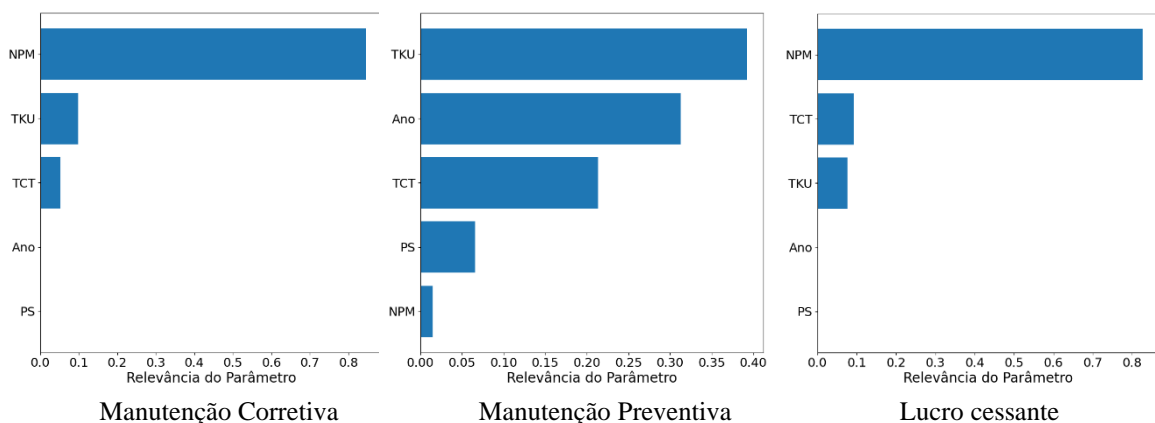


Figura 4.2 - Grau de importância das variáveis para a previsão dos custos de manutenção corretiva, preventiva e lucro cessante.

O parâmetro percentual de 89% indica que o número de paradas tem forte impacto no custo de manutenção corretiva, o que justifica a correlação do número de paradas com o lucro

cessante que compõe este custo, como indicado na Figura 5. Enquanto que o custo de manutenção preventiva sofre influência da tonelagem por quilometro-útil por ano em 39%.

O lucro cessante foi significativo frente ao número de paradas, fator este que já fora identificado no ano de 2018, devido a maior produção mineral sobrecarregando a infraestrutura, o que gerou mais paradas e, conseqüentemente, maior custo de manutenção neste ano. Segundo a Norma NBR 5462, os altos custo e tempo; o baixo investimento; a interrupção do processo ou a redução da performance interfere diretamente no lucro cessante, isto é, o mau funcionamento do equipamento impediu o lucro esperado para esse período. Este caso está relacionado ao crescimento na movimentação de cargas desde o ano de 2015 (Tabela 3.1).

Em 2012, a carência em investimentos voltados para a manutenção da via permanente era uma realidade até a política de investimentos começar a dar retorno, os problemas logísticos e operacionais já eram factíveis. Como aponta a Pesquisa do Instituto de Pesquisa Econômica Aplicada - IPEA em 2018, a greve dos caminhoneiros frente a crise econômica e política no Brasil, neste mesmo ano, foi o alerta para tentar atrair a iniciativa privada para o setor ferroviário, porém precisaria de um alto investimento público.

O algoritmo de *Gradiente Boosting Regression* apresenta 8 previsões para o conjunto de dados operacionais e econômicos subdivididas em grupos (Tab.4.1).

Tabela 4.1: Previsão a curto prazo para os Custos de Manutenção Preventiva e Corretiva.

Ano	Total de Carga Transportada	Total de Carga por Quilômetro Útil	Número de Paradas para Manutenção	Custo de Pessoal de Serviços (RS)	Lucro Cessante (RS)	Custo de Manutenção Preventiva (RS)	Custo de Manutenção Corretiva (RS)	Grupo
2022	200.000.000	180.000.000.000	110	6.000.000,00	1.477.911.593,40	4.320.014,90	1.430.291.825,30	a-I-1
2022	200.000.000	180.000.000.000	130	6.000.000,00	3.061.104.000,00	4.202.734,30	3.068.336.352,08	a-II-1
2022	190.000.000	172.000.000.000	110	6.000.000,00	419.863.611,73	6.278.657,06	427.776.351,09	b-I-1
2022	190.000.000	172.000.000.000	130	6.000.000,00	2.298.452.426,09	6.141.626,23	2.371.531.461,37	b-II-1
2022	200.000.000	180.000.000.000	110	8.000.000,00	1.477.909.354,01	4.810.913,03	1.429.169.585,93	a-I-2
2022	200.000.000	180.000.000.000	130	8.000.000,00	3.061.104.000,00	4.599.421,99	3.068.336.352,08	a-II-2
2022	190.000.000	172.000.000.000	110	8.000.000,00	419.861.000,00	6.769.555,20	426.630.555,28	b-I-2
2022	190.000.000	172.000.000.000	130	8.000.000,00	2.298.452.426,09	6.538.313,92	2.371.531.460,99	b-II-2

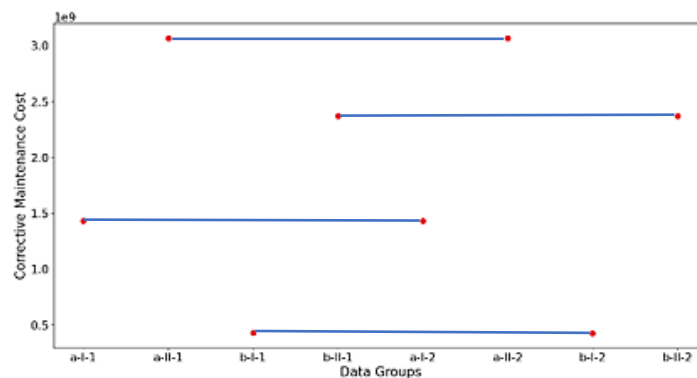
Cada Grupo foi definido por 3 valores (letra, número romano e número arábico), onde ficaram estabelecidos os parâmetros de menor a maior valor:

- a e b: maior e menor TCT e TKU, respectivamente (a: 200.000.000 Total de Carga Transportada e 180.000.000.000 Total de Carga por Quilômetro-Útil) (b: 190.000.000 Total de Carga Transportada e 172.000.000.000 Total de Carga por Quilômetro-Útil)
- I e II: menor e maior valor de número de paradas (I: 110 Número de Paradas para Manutenção) (II: 130 Número de Paradas para Manutenção)

- 1 e 2: menor e maior valor do custo de pessoal e serviço (1: 6.000.000,00 Custo de Pessoal de Serviços) (2: 8.000.000,00 Custo de Pessoal de Serviços).

A partir da definição dos grupos é possível criar quatro cenários, cada um com 2 grupos que apresentam parâmetros comuns entre si para as manutenções corretiva e preventiva.

Na Figura 4.3, a previsão para 2022 o Custo de Manutenção Corretiva está distribuída em oito grupos, os quais apresentam um conjunto de parâmetros essenciais para análise, são eles: Lucro Cessante, TU (TCT) e TKU, NPS (NPM) e PS (CPS). Seguindo a relevância dos parâmetros entre si (Fig 4.2.) criou-se 4 cenários para o Custo de Manutenção Corretiva com valores decrescentes.



CENÁRIO 1	[a - II - 1] [a - II - 2]	C.M.C: 3.068.336.352,08 LC: 3.061.104.000,00	TCT: 200.000.000 TKU: 180.000.000.000	NPM 130	CPS1 6.000.000,00	CPS2 8.000.000,00
CENÁRIO 2	[b - II - 1] [b - II - 2]	C.M.C: 2.371.531.461,37 LC: 2.298.452.426,09	TCT: 190.000.000 TKU: 172.000.000.000	NPM 130	CPS1 6.000.000,00	CPS2 8.000.000,00
CENÁRIO 3	[a - I - 1] [a - I - 2]	*C.M.C: 1.430.291.825,30 LC: 1.477.911.593,40	TCT: 200.000.000 TKU: 180.000.000.000	NPM 110	CPS1 6.000.000,00	CPS2 8.000.000,00
CENÁRIO 4	[b - I - 1] [b - I - 2]	*C.M.C: 427.776.351,09 LC: 419.861.000,00	TCT: 190.000.000 TKU: 172.000.000.000	NPM 110	CPS1 6.000.000,00	CPS2 8.000.000,00

Figura 4.3 - Grupos de previsão e cenários construídos para a Manutenção Corretiva.

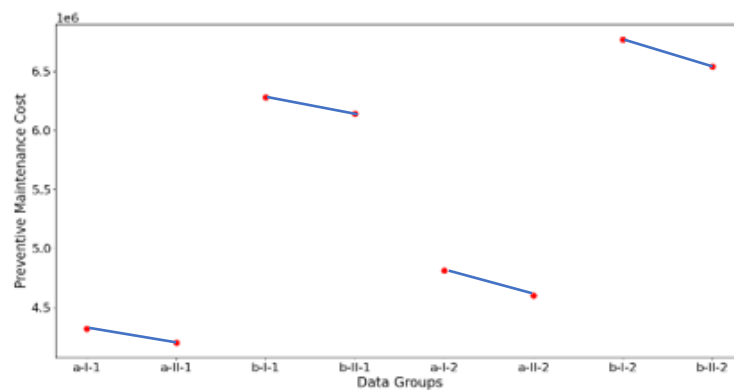
Constata-se que para maiores números de paradas, os cenários 1 e 2 apresentaram baixa variação no lucro cessante em relação ao custo de manutenção corretiva, entretanto o TCT e TKU podem variar de maior para o menor.

Observa-se que no Cenário 3 para um volume maior de TCT e TKU, o número de paradas se mantêm ao do cenário 4 e o custo pessoal de serviços segue o mesmo raciocínio do anterior. No entanto, o Lucro cessante é 3% maior que o custo de manutenção corretiva declarado.

O Cenário 4 apresentou o menor lucro cessante, conseqüentemente terá um custo de manutenção menor, porém apresentou movimentações relativamente baixas de TCT e TKU

em relação aos cenários 1 e 3, em contrapartida o número de paradas foi menor ao esperado. Por sua vez, o custo pessoal de serviços não tem uma variação expressiva frente a variação do custo de manutenção corretiva*.

O parâmetro de relevância para o custo de manutenção preventiva contribuiu para a criação dos 4 cenários, que foram agrupados em 2 grupos. Considerou-se o custo de manutenção preventiva, o total de carga transportada, o total de carga útil por quilômetro, o número de paradas e o custo com pessoal e serviço Fig.4.4.



CENÁRIO 1	[b - I - 2]	C.M.P: 6.769.555,20	TCT:190.000.000	NPM:110	CPS2: 8.000.000,00
	[b - II - 2]	C.M.P: 6.538.313,92	TKU: 172.000.000.000	130	
CENÁRIO 2	[b - I - 1]	C.M.P: 6.278.657,06	TCT:190.000.000	NPM:110	CPS1: 6.000.000,00
	[b - II - 1]	C.M.P: 6.141.626,23	TKU: 172.000.000.000	130	
CENÁRIO 3	[a - I - 2]	C.M.P: 4.819.913,03	TCT: 200.000.000	NPM: 110	CPS2: 8.000.000,00
	[a - II - 2]	C.M.P: 4.599.421,99	TKU: 180.000.000.000	130	
CENÁRIO 4	[a - I - 1]	C.M.P: 4.320.014,90	TCT: 200.000.000	NPM:110	CPS1: 6.000.000,00
	[a - II - 1]	C.M.P: 4.202.734,39	TKU: 180.000.000.000	130	

Figura 4.4 - Oito grupos de previsões e seus cenários para o Custo de Manutenção Preventiva

Os cenários 1 e 2 têm em comum entre si o TCT(TU), TKU e o Custo Pessoal e Serviços (CPS ou P.S), o mesmo acontece para os cenários 3 e 4. Nos parâmetros I e II presentes nos 4 cenários, observou-se que o número de paradas é inversamente proporcional ao Custo de Manutenção Preventiva.

Quando o custo de manutenção preventiva aumenta, o número de paradas diminui por cenário. Isto quer dizer que o NPM não está totalmente impactando nos custos gerados nesta estratégia de manutenção. Enquanto que, em um cenário qualquer, o número de paradas pode variar entre 110 e 130, para o CPS maior ou menor, isto é, o número de paradas não será determinante para a contratação de pessoal e serviço.

Os cenários que apresentam o menor custo de manutenção preventiva são os cenários 3 e 4, com valores relativamente próximos para os parâmetros TCT e TKU, o número de paradas podem variar de 110 a 130, porém com um custo pessoal de serviço também que varia por cenário.

O método de *Boosting Regression* foi capaz de correlacionar as variáveis independentes com os Custos de Manutenção Corretiva e Preventiva, bem como estimar suas projeções futuras. A partir desta modelagem do cenário (2012 – 2021), realizou-se um estudo do custo de oportunidade, obtendo-se a previsão dos custos de manutenção a curto prazo (2022).

O nível de conflito de classes ruim, razoável, bom e ótimo não foram definidos pelo Decision Tree, para este caso seria necessário construir posteriormente um novo código usando o Gradient Boosting Classifier.

4.1.9 – Análise Trade-Off ou Tomada de Decisão

A análise trade-off ou tomada de decisão foi estabelecida a partir de um nível de conflito (ruim, razoável, bom, ótimo) para comparar os benefícios encontrados entre as alternativas previstas em cada cenário, onde o estudo do custo de oportunidade é de caráter qualitativo para a previsão a curto prazo, sugerido pela condição de Yip (2014) e Sizer (1974) (Tab.4.2)(Tab 4.3).

De uma maneira mais elaborada, é possível comparar entre uma atividade e outra, seguindo um conflito entre duas escolhas ou mais. Para isso, foi considerado o benefício atribuído a cada uma, do ponto de vista da outra, como sugere Pereira et al. (1990).

Considerou-se os seguintes benefícios para a estratégia de manutenção corretiva: o menor custo de manutenção corretiva; o menor lucro cessante; o menor número de paradas e a maior movimentação de carga; e o custo pessoal e serviço é indiferente segundo a matriz de correlação (Figura 4.1).

Tabela 4.2 - Nível de conflito para estabelecer comparações para a Estratégia de Manutenção Corretiva

Estratégia MC	RUIM	RAZOÁVEL	BOM	ÓTIMO
	ALTERNATIVA 1	ALTERNATIVA 2	ALTERNATIVA 3	ALTERNATIVA 4
CMC	Maior em relação A2/A3/A4	Menor em relação A1	Menor em relação A1/A2	Menor em relação A1/A2/A3
L.C	Maior em relação A2/A3/A4	Menor em relação A1	Menor em relação A1/A2	Menor em relação A1/A2/A3
N. P.M	Maior em relação A3/ A4	Maior em relação A3/ A4	Menor em relação A1/A2	Menor em relação A1/A2
TCT & TKU	Maior em relação A2/A4	Menor em relação A1/A3	Maior em relação A2/A4	Menor em relação A1/A3

A alternativa 1 não atendeu os 3 benefícios (custo de manutenção corretiva, lucro cessante e número de paradas), parâmetros que têm grande relevância para o custo de manutenção corretiva, atendendo apenas como benefício ótimo para a movimentação de carga. A alternativa 2 superou os benefícios da alternativa 1 e nenhum em relação as demais alternativas.

Para a alternativa 3 observam-se 4 benefícios, sendo dois bons (custo de manutenção corretiva e lucro cessante) e dois ótimos (número de paradas e movimentação de carga em relação às alternativas (A1 e A2) e A4, respectivamente. No entanto, na alternativa 4 constatou-se os benefícios (C.M.C), lucro cessante e número de paradas) como ótimos e uma boa movimentação de carga.

A partir da análise dos cenários para o custo de manutenção preventiva, Figura 4.4, constatou-se a forte relação com a movimentação de carga TCT e TKU e o número de paradas variou por cenário, esta última desconsiderou-se segundo o parâmetro de relevância (Figura 4.2).

O custo de oportunidade para a estratégia de manutenção preventiva considera como benefícios: maior total de carga transportada, maior transporte de carga útil por quilômetro e menor custo pessoal e serviço. O número de paradas apesar de variar por cenário não tem correlação expressiva com o custo de manutenção preventiva (Fig. 4.1). O nível de conflito para a estratégia de manutenção preventiva é ponderado na Tabela 4.3.

Tabela 4.3 - Nível de conflito para estabelecer comparações para a Estratégia de Manutenção Preventiva.

Estratégia MP	RUIM	RAZOÁVEL	BOM	ÓTIMO
	ALTERNATIVA 1	ALTERNATIVA 2	ALTERNATIVA 3	ALTERNATIVA 4
CMP	Maior em relação A2/A3/A4	Maior em relação A3/A4	Maior em relação A4	Menor em relação A1/A2/A3
TCT & TKU	Menor em relação A3/A4	Menor em relação A3/A4	Maior em relação A1/A2	Maior em relação A1/A2
CPS	Maior em relação A2/A4	Menor em relação A1/A3	Maior em relação A2/A4	Menor em relação A1/A3

A estratégia de manutenção preventiva apresentou poucos pontos de conflito, onde os custos de manutenção foram de ruim a ótimo da alternativa 1 à alternativa 4. As alternativas 1 e 2 apresentaram o pior nível de conflito entre as demais. Por sua vez, as alternativas 3 e 4 detêm da mesma movimentação de carga, na qual a comparação é nitidamente favorável à Alternativa 4 que contempla os benefícios desejados.

Na análise de decisão *trade-off*, os custos de oportunidades identificados para as diferentes estratégias de manutenção em 2022, estão relacionados à diferentes pontos de vista gerenciais, o que sugeriu o melhor custo benefício da alternativa, mesmo quando encontrou-se duas alternativas possíveis para as estratégias de manutenção (Tabela 4.4) como abordado por Pereira et al. 1990 apud Backer & Jacobsen (1984) que também supôs duas alternativas aceitáveis mas não possíveis de serem realizadas ao mesmo tempo.

Tabela 4.4 - Seleção do Custo de Oportunidade para as Estratégias de Manutenção.

Estratégia MC	ALTERNATIVA 3	ALTERNATIVA 4
CMC	Menor em relação A1/A2	Menor em relação A1/A2/A3
L.C	Menor em relação A1/A2	Menor em relação A1/A2/A3
N. P	Menor em relação A1/A2	Menor em relação A1/A2
TCT & TKU	Maior em relação A2/A4	Menor em relação A1/A3
Estratégia MP	ALTERNATIVA 3	ALTERNATIVA 4
CMP	Maior em relação A4	Menor em relação A1/A2/A3
TCT & TKU	Maior em relação A1/A2	Maior em relação A1/A2
C.P.S	Maior em relação A2/A4	Menor em relação A1/A3

Notou-se que o custo de oportunidade para a estratégia de manutenção corretiva é relativamente equilibrado na Alternativa 3, traduzindo-se uma predição equilibrada, uma vez que se terá uma maior carga transportada, com menor número de paradas e, conseqüentemente, o lucro cessante não irá variar tanto ao ponto de impactar nos altos custos de manutenção corretiva. Para tanto, verifica-se a necessidade de implementações tecnológicas no controle de manutenção.

Do contrário, a alternativa 4 apresentou uma predição para 2022 idealista, onde existem menores (custos, lucro cessante e número de paradas) para menores movimentações de carga (TCT e TKU), porém não positivo do ponto de vista produtivo, onde o volume transportado é fator ímpar para dinamizar receitas e lucros para as concessionárias. A decisão para esta estratégia estaria condicionada a formulações de investimentos em ampliações, assim evitaria a sobrecarga nos trilhos e conseqüentemente maior paradas não programadas.

Observou-se o inverso com o custo de oportunidade na estratégia de manutenção preventiva, a alternativa 3 é realista para as condições atuais, maior a movimentação de carga e um bom planejamento para o custo de manutenção preventivo, proporcional a um custo com pessoal e serviços, contando com um custo de manutenção preventiva maior.

Analisou-se o custo de oportunidade na alternativa 4 como idealista e pretensiosa, uma vez que manter um alto padrão produtivo alinhado a um planejamento de manutenções e menor custo com pessoal e de serviços, serão necessários estudos econômicos para investir em tecnologia operacional e de manutenção. Para que isso ocorra, Bukhsh et al., (2018) explica que pelo menos teoricamente, um tomador de decisão deve considerar os estados anteriores de um ativo e orçamento limitados ao responder às questões de decisão.

Quanto ao atendimento à Norma NBR 5462, observou-se que o Lucro cessante ou perda de lucro esteve presente em 1.85% do custo de manutenção corretiva para a Alternativa 4, sendo confirmada a melhor decisão tomada para a estratégia de manutenção corretiva (Gray e Johnston, 1977).

Para um cenário em que as estratégias de manutenção corretiva e preventiva ocorrem simultaneamente, o menor custo com manutenção corretiva da Alternativa 4, R\$ 427.776.351.09 correspondeu 98,8% do custo com manutenção total, não atendendo a análise de Hupje (2020), apenas confirmando que continuará pelo menos 3 a 5 vezes mais cara que uma manutenção preventiva da Alternativa 3.

De posse da teoria do custo de oportunidade trade-off, segundo o conceito de Yip (2014), e Budai-Balke (2009) “o objetivo é encontrar um *trade-off* entre manutenção preventiva

e corretiva”, o custo de oportunidade será o equilíbrio entre estes dois cenários, quando planejados simultaneamente (Tab. 4.5).

Tabela 4.5 - Tomada de Decisão do Custo de Oportunidade para as Estratégias de Manutenção, simultaneamente.

Estratégia MC	ALTERNATIVA 3	ALTERNATIVA 4
CMC		Menor em relação A1/A2/A3
L.C		Menor em relação A1/A2/A3
N P	Menor em relação A1/A2	Menor em relação A1/A2
TCT & TKU	Maior em relação A2/A4	-
Estratégia MP	ALTERNATIVA 3	ALTERNATIVA 4
CMP	Maior em relação A4	-
TCT & TKU	Maior em relação A1/A2	Maior em relação A1/A2
CPS	Maior em relação A2/A4	-

Desta forma, identificou-se um custo de oportunidade para atender, segundo os parâmetros de relevância, as manutenções preventiva e corretiva para o cenário do ano 2022, sendo eles:

- Maior TCT e TKU (maior benefício produtivo)
- Menor Número de paradas para manutenção (menos custo com mão de obra e lucro cessante)
- Maior Custo de Pessoal e Serviços (mais custo em contratação de engenheiros e serviços de tecnologia, porém maior o benefício para as operações)
- Maior Custo de Manutenção Preventiva (gasta-se mais com C.P.S e serviços, porém previne-se de custos adicionais para a manutenção de emergência e o benefício é na baixa depreciação do ativo ferroviário.
- Menor Custo de Manutenção Corretiva (maior benefício logístico, diminui-se gargalos logísticos e despesas contratuais extras, e garantem-se a segurança, a maneabilidade e confiabilidade.

CAPÍTULO 5

CONCLUSÕES E RECOMENDAÇÕES

O estudo analisou os indicadores econômico-financeiros, operacionais e os custos de manutenção em uma estrada de ferro na região amazônica nos últimos 10 anos. Esta cooperação viabiliza as decisões gerenciais a partir de como o custo de oportunidade tem impacto sobre as estratégias de manutenção.

A correlação de Pearson entre o custo de manutenção corretiva e o lucro cessante e o número de paradas é 1,00 e 0,94, respectivamente. Enquanto que o custo de manutenção preventiva em relação ao TCT e TKU é de 0,81. Para a análise dos cenários para 2022, os custos de manutenção corretiva, preventiva e o lucro cessante pelo algoritmo de *Gradiente Boosting Regression*, apresentaram 8 (oito) grupos de predição, com estimativas de Tonelada de Carga Transportada variando de 190.000.000 a 200.000.000, total de carga por quilometro-útil de 172.000.000.000 a 180.000.000.000; número de paradas para manutenção variando de 110 a 130 e uma variação de R\$ 6.000.000,00 à R\$ 8.000.000,00 com Custo de Pessoal de Serviços.

Estes grupos apresentaram um conjunto de parâmetros cruzados entre si, o qual resultou 4 cenários para os custo de manutenção preventiva e corretiva, o lucro cessante foi absorvido por este último custo. A partir do cruzamento destes dados, o menor custo de manutenção preventiva foi para os cenário 3 e 4, com valores relativamente próximos para os parâmetros TCT e TKU, o número de paradas podem variar de 110 a 130, porém com um custo pessoal de serviço variante.

Para o custo de manutenção corretiva, o cenário 4 apresentou o menor lucro cessante, consequentemente terá um custo de manutenção menor, porém apresentou movimentações relativamente baixas de TCT e TKU em relação aos cenários 1 e 3, em contrapartida o número de paradas foi menor ao esperado. Por sua vez, o custo pessoal de serviços pode variar devido a baixa variação do custo de manutenção corretiva.

O método de *Boosting Regression* desempenhou bem o processamento das variáveis em relação aos Custo de Manutenção Corretiva e Preventiva e o Lucro Cessante para fazer projeções futuras. A análise do custo de oportunidade foi elaborada a partir da modelagem do cenário (2012 - 2021), onde obteve a predição dos custos de manutenção a curto prazo (2022).

Esta análise de decisão (*trade-off*) foi estabelecida a partir de um nível de conflito (ruim, razoável, bom, ótimo) onde comparou os benefícios encontrados entre as alternativas previstas em cada cenário.

Os custos de oportunidades identificados para as diferentes estratégias de manutenção em 2022, apresentou diferentes alternativas que sugeriu decidir a escolha do melhor custo-benefício da alternativa, mesmo quando encontra-se duas alternativas possíveis para as estratégias de manutenção, assim como também propõe duas alternativas aceitáveis, mas não possíveis de serem realizadas ao mesmo tempo.

Atendendo o objetivo deste estudo, o custo de oportunidade foi encontrado entre o equilíbrio das Alternativas 3 e 4 para as estratégias de manutenção preventiva e corretiva, resultado de correlações entre variáveis e parâmetros de relevância encontrados a partir do *Machine Learning - Gradient Boosting Regression*. O equilíbrio entre estes dois cenários, sugerindo por Lamberts (2009) quando planejados simultaneamente, tem-se um custo de oportunidade para suas respectivas manutenções, para o cenário do ano 2022, sendo eles: o maior TCT e TKU (maior benefício produtivo); menor número de paradas para manutenção (menos custo com mão de obra e lucro cessante); maior custo com pessoal e serviços (mais custo em contratação de engenheiros e serviços de tecnologia, porém maior o benefício para as operações); maior custo com manutenção preventiva (gasta-se mais com C.P.S e serviços, porém previne-se de custos adicionais para a manutenção de emergência e o benefício é na baixa depreciação do ativo ferroviário; e menor custo com manutenção corretiva (maior benefício logístico, diminui-se gargalos logísticos e despesas contratuais extras, e garantem-se a segurança, a manutenibilidade e confiabilidade).

Os resultados obtidos com este estudo evidenciam que o estudo do custo de oportunidade aliado ao aprendizado de máquina serve como um instrumento de auxílio às empresas na busca por uma maior compreensão do melhor custo benefício nas suas decisões sobre as estratégias de manutenção, contribuindo para a melhoria da gestão dos ativos.

No desenvolver deste estudo foi observada a ausência de dados específicos sobre os custos de manutenção de trilhos, assim como, literaturas que abordassem o custo de oportunidade na manutenção ferroviária. Como sugestões para continuidade do presente estudo, têm-se o aprofundamento do custo de oportunidade na logística ferroviária de mineração, para grandes multinacionais; e estudar o custo de oportunidade a partir da análise de sobrevivência, um estudo estatístico, onde é aferido o custo de oportunidade negligenciado desde a primeira decisão do gestor até ao final de uma determinada atividade.

REFERÊNCIAS BIBLIOGRÁFICAS

ABNT. ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **Norma NBR 5462.: Confiabilidade e Manutenibilidade.** Rio de Janeiro: ABNT, 1994.

ANTF. ASSOCIAÇÃO NACIONAL DOS TRANSPORTES FERROVIÁRIOS. Disponível em: www.antf.org.br/informacoes-gerais/ Acesso em: 08 mar. Brasília, 2022.

ANTT. AGÊNCIA NACIONAL DE TRANSPORTE TERRESTRE. **Anuário da Demonstrações financeiras Carve Out da Estrada de Ferro Carajás 2012 – 2021.** Brasília. 2022.

ANTT. AGÊNCIA NACIONAL DE TRANSPORTE TERRESTRE. **Manual de Custos Referenciais Ferroviários. Volume 1. Metodologia e Conceitos.** Brasília, 2021.

ASSIS, R. M. **Estudo de fratura em trilho com origem no patim na ferrovia do aço – MRS Logística.** Instituto Militar de Engenharia. 2018.

BATURYNSKA, I.; MARTINSEN, K. **Prediction of geometry deviations in additive manufactured parts: comparison of linear regression with machine learning algorithms.** Journal of Intelligent Manufacturing. 2020.

BRAGA. E. M. **Análise de Tensões Residuais em Trilhos Ferroviários.** Projeto de Pesquisa. Parceria VALE & Universidade Federal do Pará. Laboratório de Macânica. 2018.

BUDAI-BALKE, G. 2009. **Operations Research Models for Scheduling Railway Infrastructure Maintenance.** Tesis. 2013. Erasmus Universiteit Rotterdam. Netherlands.

BUKSHA, Z. A.; SAEED A.; STIPANOVICA, I. **A machine learning approach for maintenance prediction of railway assets a Department of Construction Management and Engineering.** Proceedings of 7th Transport Research Arena TRA 2018, April 16-19, 2018, Vienna, Austria, University of Twente, Enschede, Netherland.

CASTELLO BRANCO, J. E. S. **A Segregação da Infraestrutura como elemento reestruturador do sistema ferroviário brasileiro.** Tese de doutorado. UFRJ/COPPE/Programa de Engenharia de Transportes, 2008.

CNT. CONFEDERAÇÃO NACIONAL DE TRANSPORTES. **Anuário de Pesquisa da Confederação Nacional de Transportes. Painel do Transporte Ferroviário.** Disponível em: <https://www.cnt.org.br/painel-cnt-transporte-ferroviario>. Acesso: Maio/2022.

FOURIE, J; ZHUWAKI; N. T. **A modelling framework for railway infrastructure reliability analysis**South African. Journal of Industrial Engineering C. Published 2017.

GARCÍA NIETO, P. J.; GARCÍA-GONZALO, E.; ARBAT G.; DURAN-ROS, M.; CARTAGENA, F. R.; PUIG-BARGUÉS, J. **Pressure drop modelling in sand filters in micro-irrigation using gradient boosted regression trees.** Biosystems Engineering. Volume 171, Pages 41-51. July 2018

GARTNER - GARTNER RESECHER CIRCLE. **Gartner Says the Majority of Technology Products and Services Will Be Built by Professionals Outside of IT by 2024.** Pesquisa de TI & Negócios. STAMFORD, Conn. June 14, 2021.

GLAUTIER, M. W. E.; UNDERDOWN, B. **Accounting: Theory and Praticce. London.** Pitrnan Publishing Limited. 3th edition. 1986.

GRAY, J.; JOHNSTON, K. S. **Contabilidade e Administração.** São Paulo. McGrawHill, 1977.

GRIMES, G. A.; BARKAN, C. P. L. **Cost-effectiveness of railway infrastructure renewal main.** Journal of Transportation Engineering. Published: 2006.

HEYNS, G. J.; KIBOURN, P. J. **Logistics opportunity costs: a mining case study.** Journal of Transport and Suplly Chain Management. Published 2013

HUPJE, E. **Principles of a modern preventive maintenance program.** Available:<https://roadtoreliability.com/reliability-centered-maintenance-principles/>. Access: may/2022.

IPEA. INSTITUTO DE PESQUISA ECONÔMICA APLICADA. **O problema dos custos de transação em parcerias público-privadas em infraestrutura.** Texto para discussão. Brasília, 2018

KUSSANO, M. R.; BATALHA, M. O. **Custos logísticos agroindustriais: avaliação do escoamento da soja em grão do Mato Grosso para o mercado externo.** Grupo de Estudos e Pesquisas Agroindustriais – GEPAI. Departamento de Engenharia de Produção, Universidade Federal de São Carlos – UFSCar. 2012.

LEONE, G.S.G. **Custos, Panejamento, Implantação e Controle.** Atlas 1982. São Paulo.

LIDEN, T. 2015. **Railway infrastructure maintenance-a survey of planning problems and conducted research.** Transportation Research Procedia 10 (2015) 574 – 583. Delft, The Netherlands.

MORSE, W.J.; ROTH, H.P. **Cost Accounting: Processing, Evaluation and Using Cost** Dateúg Co-t Qtt.a. Addison VEsley Publishing. 3 th edition, 1986. p.36.

MARCORIN, W.R.; LIMA, C. R. C. **Cost Analysis of Maintenance and Non-maintenance Policies for Productive Equipment.** Revista de Ciência & Tecnologia v. 11. 2003.

MARTINS, E. **Contabilidade e Custos.** Atlas 3^a Edição. São Paulo.

MEDEIROS, O. R.; COSTA, P. S.; SILVA, C. A. T. **Testes empíricos sobre o comportamento assimétrico dos custos nas empresas brasileiras.** Revista Contabilidade e Finanças, São Paulo: FEA/USP, v. 16, n. 38, mai./ago. 2005.

MOURA, F. **Parecer sobre ACB da modernização da Linha do Norte Estudo realizado para a Agência para o Desenvolvimento e Coesão.** Novembro de 2017 Relatório FUNDEC PS nº 52/2017 Relatório CERIS EP nº 52/2017.

NAKHAEE, M. C.; HIEMSTRA, D.; STOELINGA, M; VAN NOORT, M. **The recent applications of machine learning in rail track maintenance: a survey.** In Book: (pp.91-105) & Conference paper. Part of the Lecture Notes in Computer Science book series (LNPSE, volume 11495) First Online: 24 April 2019.

NIELSEN, D.; CHATTOPADHYAY, G.; RAMAN, D. **Life Cycle Cost Estimation for Railway Bridge Maintenance.** Centre for Railway Engineering, CQUniversity Austrália. 2CRC for Rail Innovation, 2013.

ONLT. OBSERVATÓRIO NACIONAL DE TRANSPORTE E LOGÍSTICA DA EPL. **Boletim de Logística. Setor Ferroviário.** Disponível em: <https://ontl.epl.gov.br/paineis-analiticos/> Acesso: maio/2021.

PEREIRA, A. C.; SOUZA, B. F.; REDAELLI, D. R.; IMONIANA, J. O. **Custo de Oportunidade: Conceitos e Contabilização Caderno de Estudos nº 02.** São Paulo, FIPECAFI – abril/1990.

PEREIRA, L. **O que é o Custo de Oportunidade na Economia? Artigo Publicado na Revista Eletrônica - Dicionário Financeiro.** 2018. Disponível em: www.dicionariofinanceiro.com. Acesso em: maio/2022.

RUS, G.; CAMPOS, J.; PADILLA, A. O; SOCORRO, M. P. **Cost-benefit analysis of railway projects: high-speed lines and suburban lines.** Project: Economic evaluation of transport projects and policies: Methodology and Applications. September 2020.

SCHNEIDER, E. L.; REGULY, A. **Análise da vida útil de trilhos com defeitos transversais desgastados em serviço. Revista Tecnologia e Tendências.** Volume 6, jan – jun 2014.

SICILIANO, G.; BARONTINI, F; ISLAM, S. M. Z.; ZUNDER, T. H.; MAHLER, S.; GROSSONI, I. 2016. **Adapted cost-benefit analysis methodology for innovative railway services.** European Transport Research Review 8, Article number: 23, 2016.

SILCON – ESTUDOS ECONÔMICOS. **Notas para reflexão sobre os investimentos em ferrovias no Brasil.** Série relatórios SILCON RS 092 – março de 2021.

SILVEIRA, M. R. **A importância econômica das ferrovias para o Brasil. ANTP - Associação Nacional de Transporte e Pesquisa.** Brasília, 2002.

SIZER, J. **Noções Básicas de Contabilidade Gerencial.** Saraiva. 1ª Edição. São Paulo, 1974.

SOUSA, N.; ALÇADA-ALMEIDA, L.; COUTINHO-RODRIGUES, J. **Planeamento multiobjetivo em projetos de renovação da ferrovia.** 8º congresso luso-brasileiro para o planeamento urbano, regional, integrado e sustentável. Outubro, 2018.

SOUZA, R. D. **Análise da Gestão da Manutenção: estudo de caso MRS Logística. Monografia (graduação). Curso de Engenharia de Produção.** Universidade Federal de Juiz de Fora (MG), 2008.

SYSTEM ANALYSIS PROGRAM DEVELOPMENT – SAP. **What is machine learning? | Definition, types, and examples|SAP Insights.** Disponível www.sap.com/insights/machine-learning, 2022.

TALASZKA, A. A. T. **Comportamento dos custos em empresas do segmento de transportes ferroviários listados na BM&FBOVESPA.** Monografia de especialização em gestão financeira. Universidade Tecnológica Federal do Paraná. Curitiba – Paraná, 2018.

THIELMANN, R.; QUEIROZ, K.O. L. **Trade – off entre custo e tempo em projetos.** Revista Valore. Universidade Federal Fluminense, Volta Redonda, Brasil, 2021.

VALE - EMPRESA PRIVADA DE MINERAÇÃO. **Conheça mais sobre a história de Carajás, a maior mina de minério de ferro do mundo.** 2018. Disponível em: <http://www.vale.com/hotsite/PT/Paginas/conheca-mais-sobre-historia-carajas-maior-mina-minerio-ferro-mundo.aspx>. Acesso: maio 2022.

VALE - EMPRESA PRIVADA DE MINERAÇÃO. **Informações sobre a estrada de ferro carajás. Dados históricos e operacionais.** Disponível em: <http://www.vale.com/brasil/PT/initiatives/innovation/carajas-railway/Paginas/default.aspx>. Acesso: maio 2022.

VAN ESSEN, C. E.; BART, B.H. BOON, C. E.; EELCO, L. C.; DEN BOER, C. E.; JASPER, J.; FABER, C. E.; KOEN, K.T.H.; VERVOORT, E.; ROCHEZ, C. **Marginal costs of infrastructure use - towards a simplified approach Final report.** Bibliographical data: H.P. (Huib), CE, 2004.

VAN JAARVELD, L.; HEYNS, G. J.; KILBOURN, P. J. **Logistics opportunity costs: A mining case study.** Journal of Transport and Supply Chain Management. 2013.

WANG, F. K.; MAMO, T. **Gradient boosted regression model for the degradation analysis of prismatic cells.** Computers & Industrial Engineering. Volume 144. Department of Industrial

Management, National Taiwan University of Science and Technology, Taipei 10607, Taiwan. June 2020.

WOLPERT, D. H. “**Stacked generalization.**” *Neural networks 5.2*: 241-259 (<https://scikit-learn.org/stable/modules/ensemble.html>). Access: may/2022.

WORSLEY, T. “**Comparing Road and Rail Investment in Cost-Benefit Analysis**”, *International Transport Forum Discussion Papers*, No. 2020/29, OECD Publishing, Paris

YIP, D. S. O. **A Critical Review of the Opportunity Cost Concep.** The London School of Economics and Political Science. Theses. 2014 Access from Institution:<http://etheses.lse.ac.uk/3053/>.

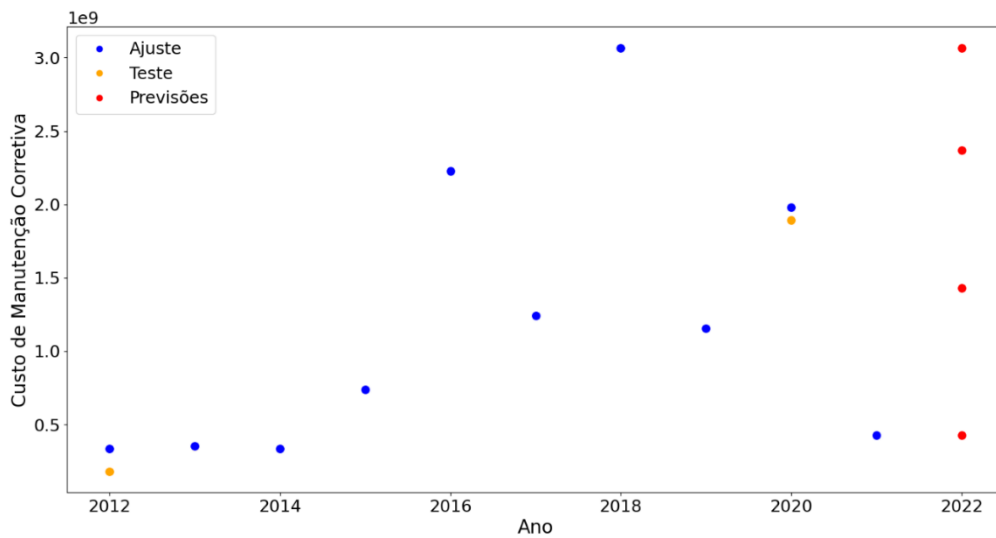
XAVIER, R. M. R. S. **Melhor aproveitamento do intervalo concedido.** Monografia apresentada ao Curso de Especialização em Transporte Ferroviário de Carga do Instituto Militar de Engenharia. 2008.

ZHOU SU; SCHUTTER. **Optimal sheduling of track maintenance activies for railway networks.** IFAC Papers (International Federation of Automatic Control). Delft Center for Systems and Control, Delft University of Technology, the Netherlands, 2018.

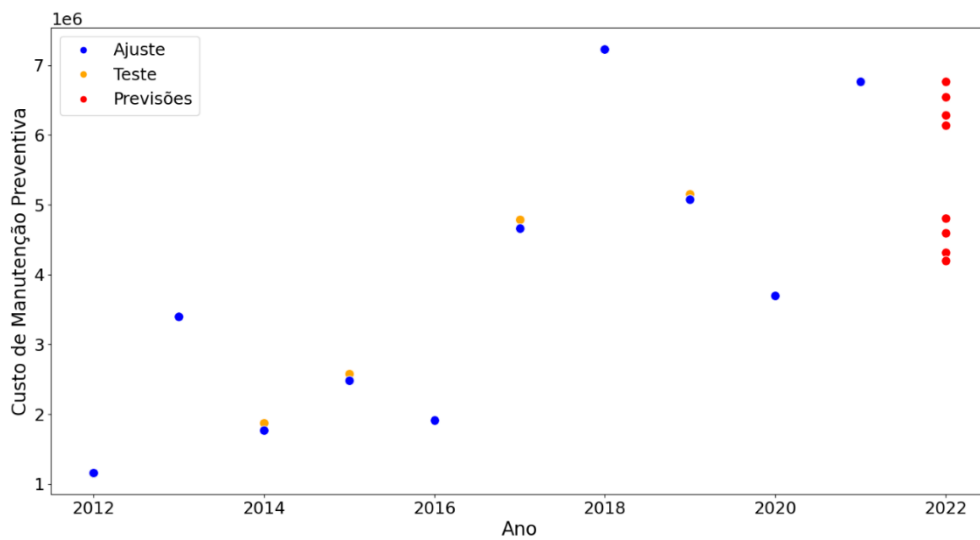
APÊNDICE

APÊNDICE A

GRÁFICOS COMPARATIVOS DOS RESULTADOS OBTIDOS NO APRENDIZADO DE MÁQUINA PARA A MANUTENÇÃO CORRETIVA E PREVENTIVA

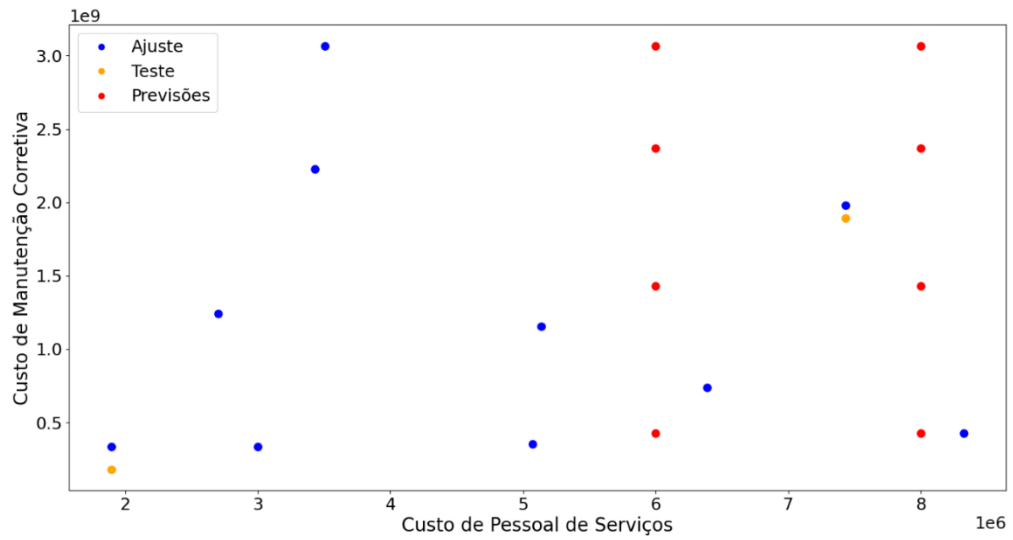


(a)

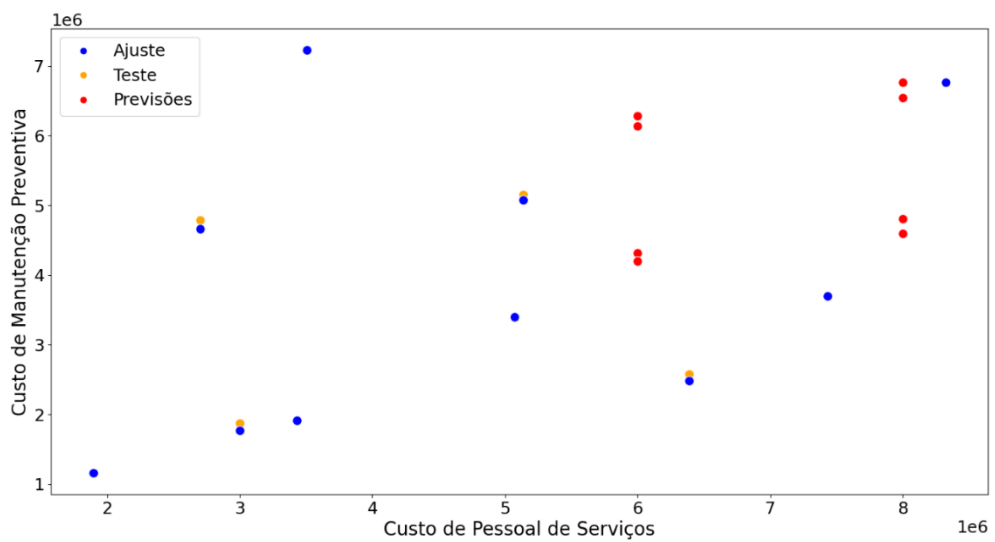


(b)

Figura A1 – Comparativos dos resultados obtidos no aprendizado de máquina para o ajuste, teste e previsões em função do ano. (a) Manutenção corretiva e (b) Manutenção preventiva.

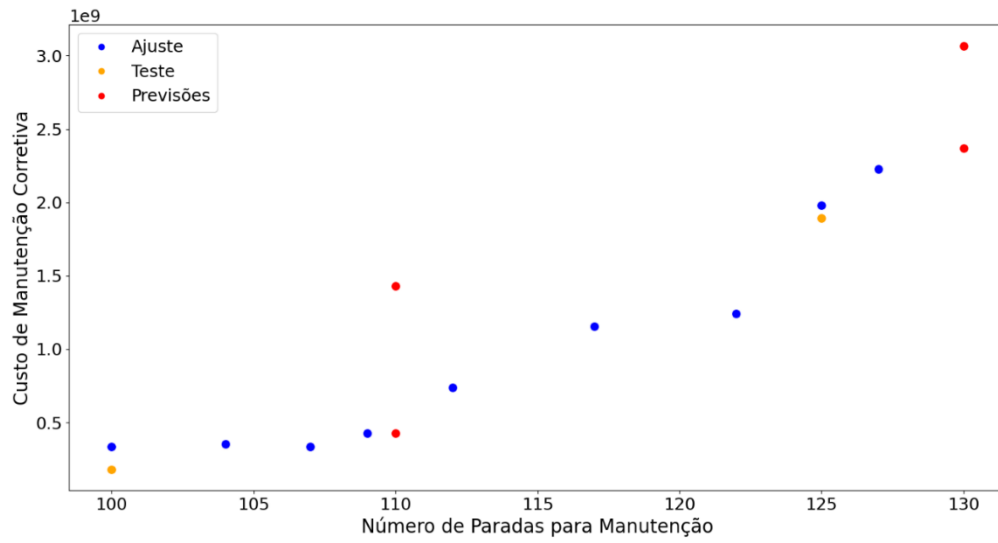


(a)

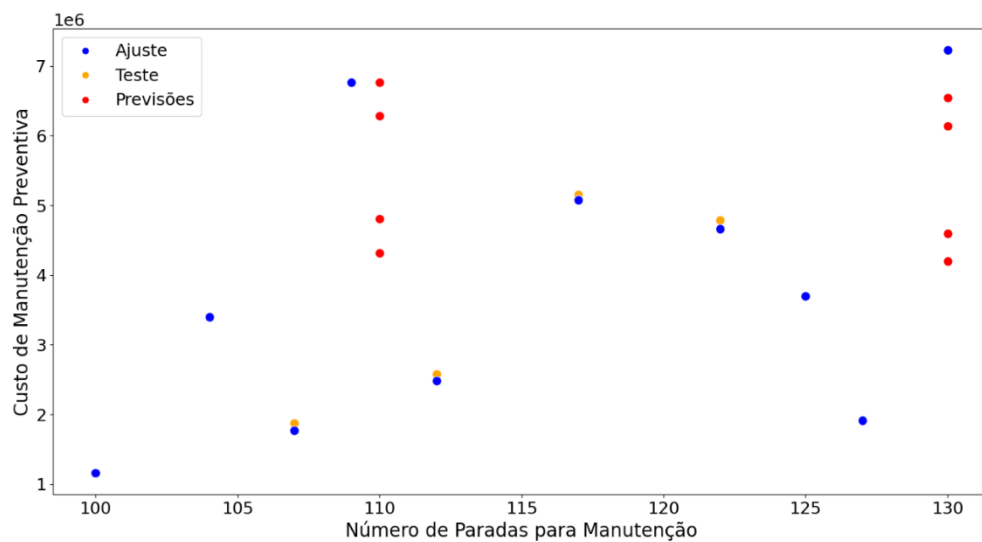


(b)

Figura A2 – Comparativos dos resultados obtidos no aprendizado de máquina para o ajuste, teste e previsões em função do Custo de Pessoal de Serviço. (a) Manutenção corretiva e (b) Manutenção preventiva.

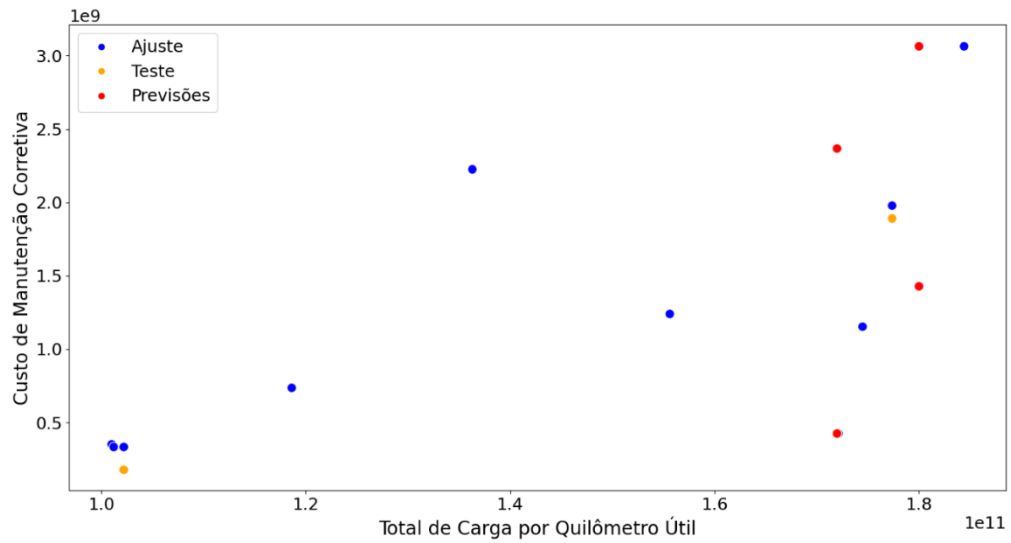


(a)

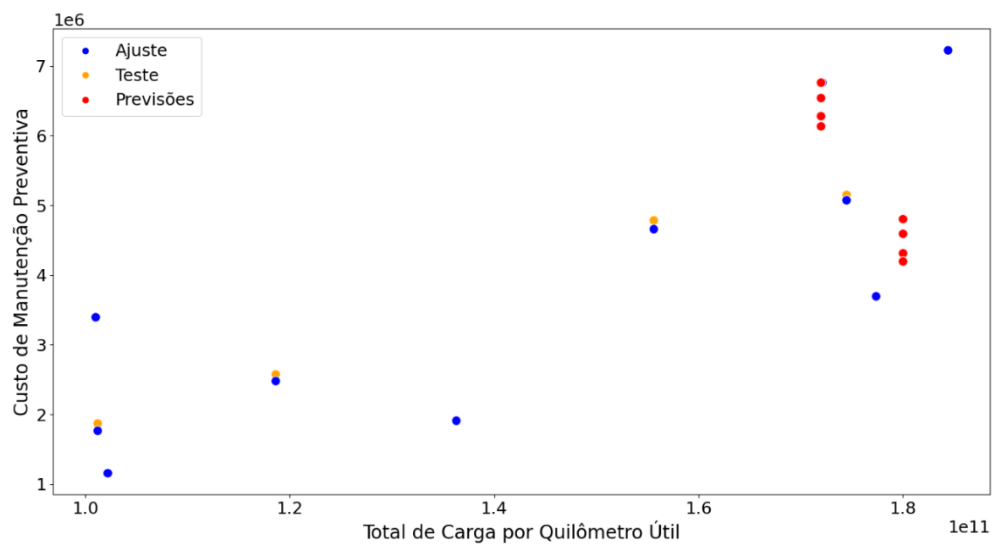


(b)

Figura A3 – Comparativos dos resultados obtidos no aprendizado de máquina para o ajuste, teste e previsões em função do Número de Paradas. (a) Manutenção corretiva e (b) Manutenção preventiva.

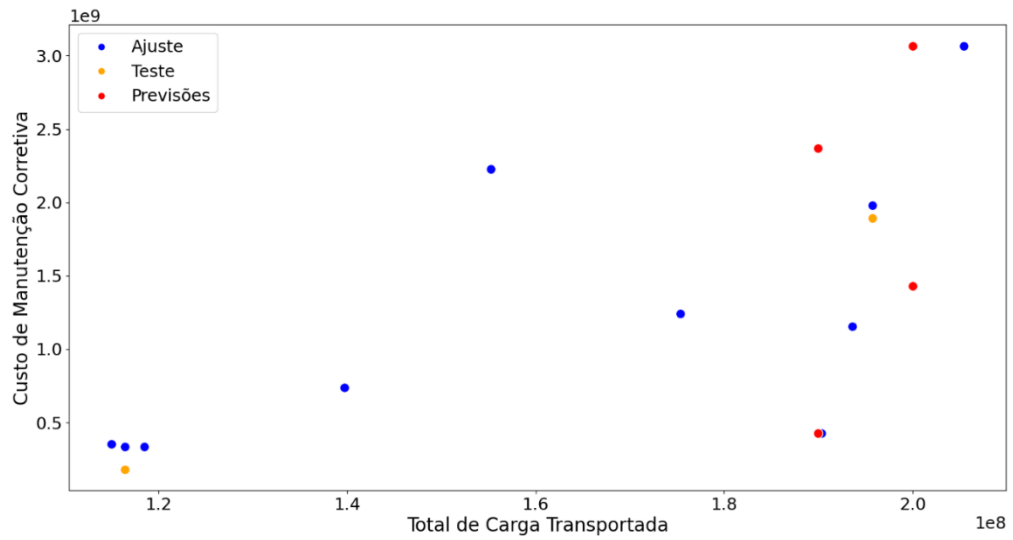


(a)

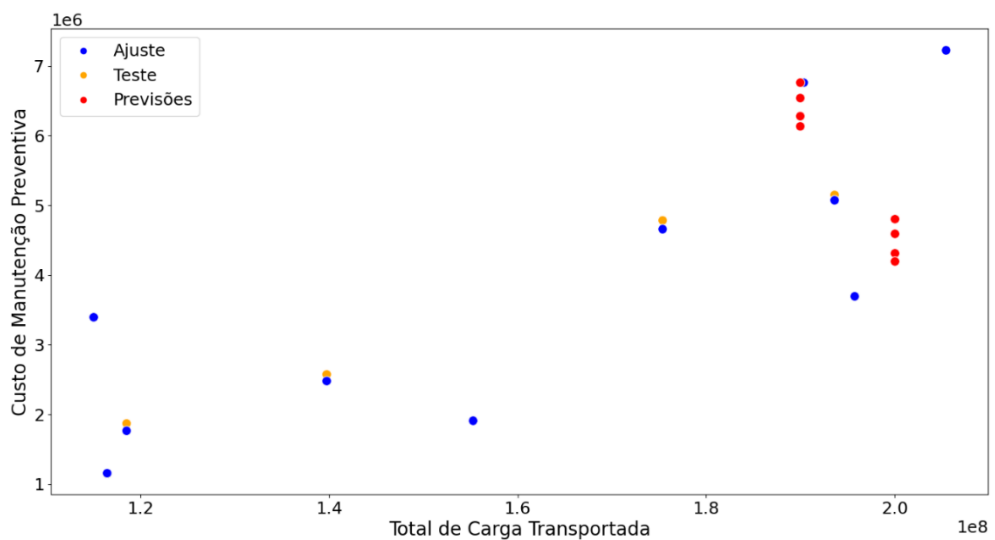


(b)

Figura A4 – Comparativos dos resultados obtidos no aprendizado de máquina para o ajuste, teste e previsões em função da Tonelada Quilometro Útil. (a) Manutenção corretiva e (b) Manutenção preventiva.



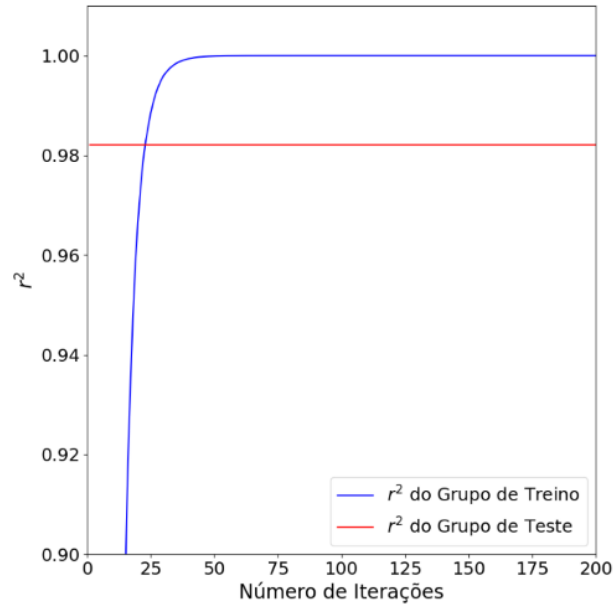
(a)



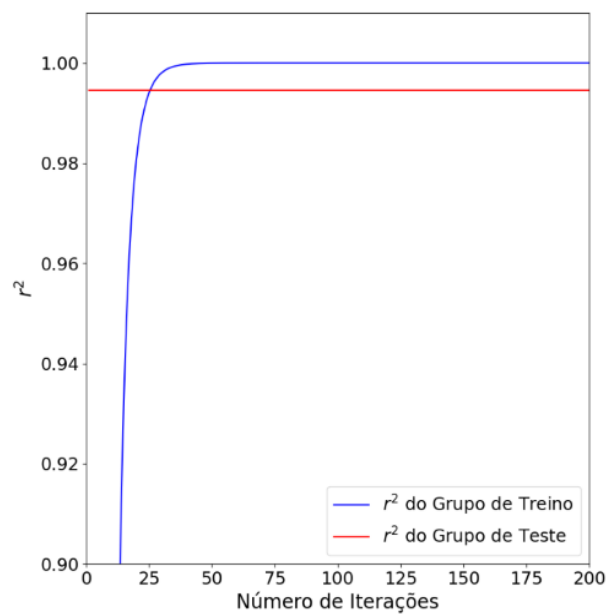
(b)

Figura A5 – Comparativos dos resultados obtidos no aprendizado de máquina para o ajuste, teste e previsões em função do Total de Carga Transportada. (a) Manutenção corretiva e (b) Manutenção preventiva.

APÊNDICE B

RESULTADOS DO NÚMERO DE INTERAÇÕES DE ERRO DE TREINO E TESTE
PARA A MANUTENÇÃO CORRETIVA E PREVENTIVA

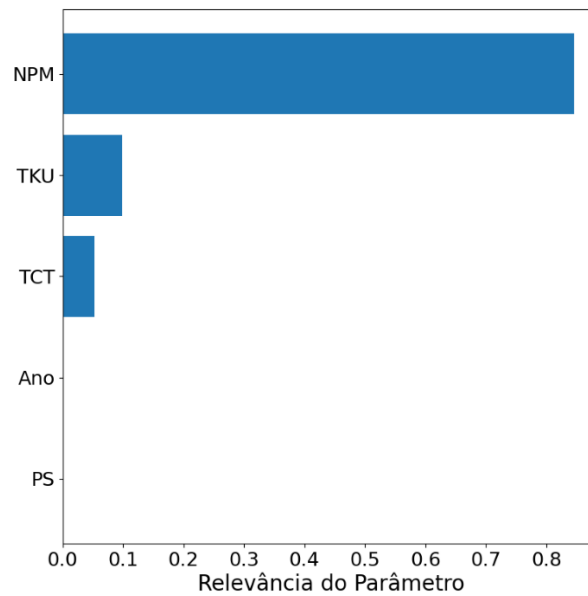
(a)



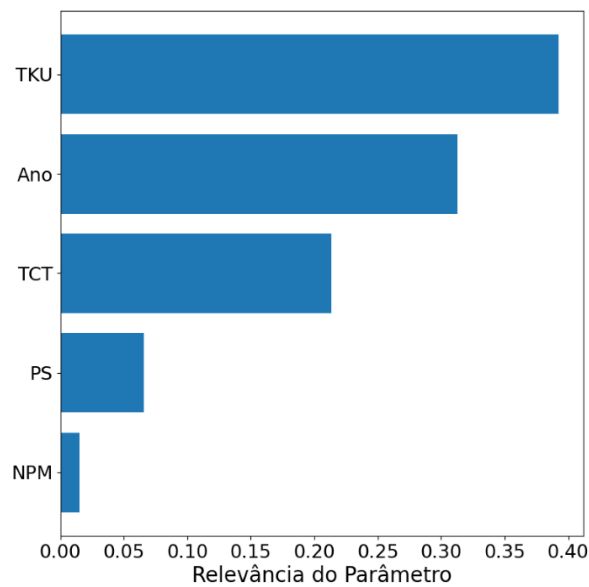
(b)

Figura B – Resultados da interação no aprendizado de máquina para o erro do grupo de treino e de teste. (a) Manutenção corretiva e (b) Manutenção preventiva.

APÊNDICE C

GRÁFICOS COMPARATIVOS DA RELEVÂNCIA DO PARÂMETRO VARIÁVEL
VERSUS O CUSTO DE MANUTENÇÃO CORRETIVA E PREVENTIVA

(a)

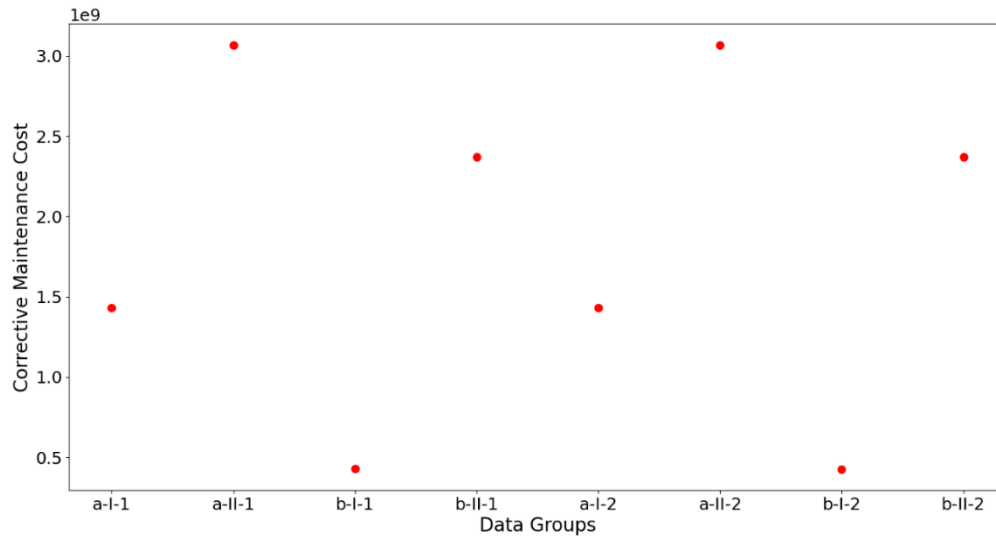


(b)

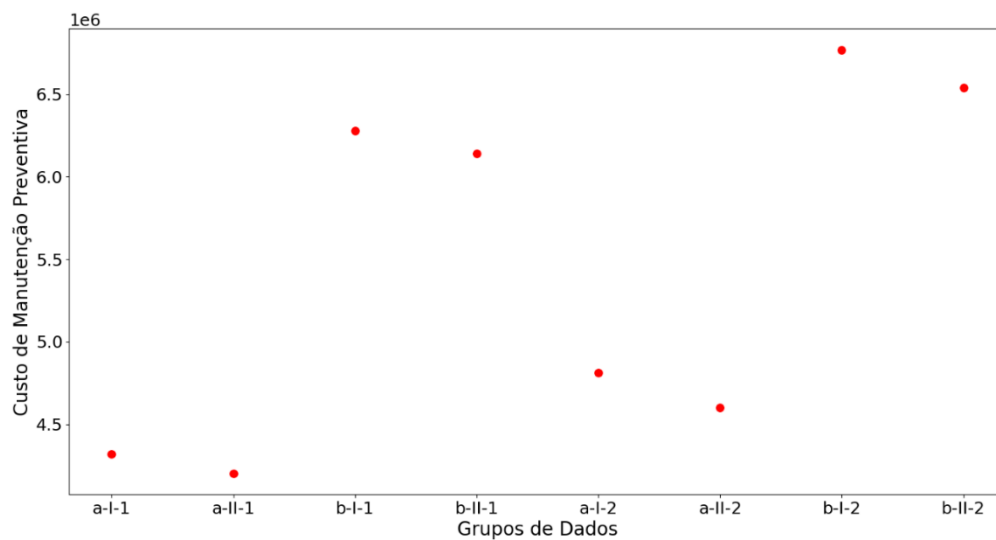
Figura c – Gráficos comparativos da relevância do parâmetro variável versus o custo de manutenção (a) corretiva e (b) preventiva

APÊNDICE D

RESULTADO DA PREVISÃO DO CUSTO DE MANUTENÇÃO CORRETIVA E PREVENTIVA PARA O ANO DE 2022 SEGUNDO A CARACTERIZAÇÃO DOS OITO GRUPOS DE DADOS



(a)



(b)

Figura D – Gráficos da previsão do custo de manutenção corretiva e preventiva para o ano de 2022, segundo a caracterização dos oito grupos de dados (a) manutenção corretiva e (b) manutenção preventiva.

ANEXO

DETALHAMENTO DO APRENDIZADO DE MÁQUINA NO TRATAMENTO DAS VARIÁVEIS: GRÁFICOS, REFERÊNCIAS DE PARÂMETRO, MATRIZ DE CORRELAÇÃO

Importação das Bibliotecas Necessárias

```

# Bibliotecas de manipulação de fontes de dados e matemática
import pandas as pd
import numpy as np

# Bibliotecas de geração de gráficos
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib.pyplot import figure
from matplotlib.lines import Line2D

# Definições para os gráficos
SMALL_SIZE = 18
MEDIUM_SIZE = 20
BIGGER_SIZE = 22

plt.rc('font', size=SMALL_SIZE)          # Tamanho dos textos
plt.rc('axes', titlesize=SMALL_SIZE)     # Tamanho dos números dos
eixos
plt.rc('axes', labelsize=MEDIUM_SIZE)   # Tamanho dos textos dos eixos
plt.rc('xtick', labelsize=SMALL_SIZE)   # Tamanho das marcas do eixo x
plt.rc('ytick', labelsize=SMALL_SIZE)   # Tamanho das marcas do eixo y
plt.rc('legend', fontsize=SMALL_SIZE)   # Tamanho da fonte das
legendas
plt.rc('figure', titlesize=BIGGER_SIZE)  # Tamanho da fonte do título

# Bibliotecas de Machine Learning (ML)
from sklearn.model_selection import train_test_split
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.metrics import r2_score

##### Carregamento dos Dados
#####

do = pd.read_csv('dados.txt')           # Importa os dados do
arquivo dados.txt

X = do[['Ano', 'PS', 'TU', 'TKU', 'NPM']] # Define as variáveis
independentes (input)
ylc = do[['LC']]                         # Define a variável dependente
para o Lucro Cessante (input)
ycmp = do[['CMP']]                       # Define a variável dependente
para o Custo de Manutenção Preventiva (input)
ycmc = do[['CMC']]                       # Define a variável dependente
para o Custo de Manutenção Corretiva (input)
# Os quatro são inputs, porque
mesmo o y sendo resultado de X, ele é fornecido para o ML

```

```

d2022 = pd.read_csv('2022.txt') #
Importa os dados do arquivo 2022.txt, contendo dois valores de entrada
para cada variável independente para o ano de 2022 para previsão

X_2022 = d2022[['Ano', 'PS', 'TU', 'TKU', 'NPM', 'Grupo']]
# Define as variáveis independentes de 2022

### Lucro Cessante ### Divisão em Conjuntos de Treino e de Teste ###

X_trainlc, X_testlc, y_trainlc, y_testlc = train_test_split(
    # O train_test_split divide X e y em um conjunto de treino e
um de teste, na proporção definida pelo test_size
    X.to_numpy(), ylc.to_numpy().ravel(), test_size=0.2, random_state=0
)

##### Definição dos Parâmetros do Gradient
Boosting Regressor (GBR) #####

paramslc = {
    # Junto com o test_size e o
random_state do train_test_split, estes são os parâmetros determinados
através do outro código
    "criterion": "squared_error", # Tipo de erro, dentre as
opções do GBR, para seu ajuste
    "learning_rate": 0.1, # Peso (no mesmo sentido da
média ponderada) de cada novo ajuste
    "n_estimators": 1000, # Número de iterações
    "min_samples_split": 3, # Número mínimo de valores nos
nós das Árvores para que haja uma divisão
    "min_samples_leaf": 1, # Número mínimo de amostras em
uma folha
    "max_depth": 50, # Número Máximo de divisões de
uma Árvore
    "random_state": 100,
}

##### Execução do GBR #####

gbrlc = GradientBoostingRegressor(**paramslc) # Definição do
comando do GBR como 'gbrlc'
gbrlc.fit(X_trainlc, y_trainlc) # Treinamento do
algoritmo. Uso do GBR para o ajuste (fit) dos dados de treino. Após
esta linha, o termo 'gbr' está treinado com base nos dados fornecidos.

y_predict = gbrlc.predict(X_testlc) # Predição dos
valores de y do conjunto de teste, a partir dos dados de X desse
conjunto por meio do algoritmo treinado. Este é o primeiro output.

print(r2_score(y_trainlc, gbrlc.predict(X_trainlc))) # Mostra o r2
do grupo de treino no terminal. Idealmente é 1, pois o ajuste deve ser
perfeito.
print(r2_score(y_testlc, gbrlc.predict(X_testlc))) # Mostra o r2 do
grupo de teste no terminal. Idealmente deve estar acima de 0.95, para
garantir que o algoritmo faça boas previsões com dados que ele não usou
como treino.

##### Criação dos Gráficos #####

```

```

legenda = [Line2D([], [], marker='o', color='blue', linestyle='None'),
           # Algumas definições para fazer a legenda dos gráficos
           Line2D([], [], marker='o', color='orange', linestyle='None'),
           Line2D([], [], marker='o', color='red', linestyle='None')]

legenda2 = [Line2D([], [], marker='o', color='red', linestyle='None'),
            # Algumas definições para fazer a legenda dos gráficos
            Line2D([], [], marker='o', color='blue', linestyle='None')]

legenda3 = [Line2D([], [], marker='o', color='red', linestyle='None'),
            # Algumas definições para fazer a legenda dos gráficos
            Line2D([], [], marker='o', color='blue', linestyle='None'),
            Line2D([], [], marker='o', color='green', linestyle='None')]

ylc_2022 = gbrlc.predict(X_2022.to_numpy()[:, [0,1,2,3,4]])
           # Gera os valores do lucro cessante a partir da
previsão do GBR com os dados de 2022.

##### Lucro Cessante por Ano #####

plt.figure(figsize=(18, 9))
    # Define o tamanho da figura

sns.scatterplot(x=X_trainlc[:,0], y=y_trainlc,s=100,color='black')
    # Plota os dados de treino como pontos pretos
sns.scatterplot(x=X_testlc[:,0], y=y_testlc,s=100,color='orange')
    # Plota os dados de teste como pontos laranjas
sns.scatterplot(x=X.to_numpy()[:,0],y=gbrlc.predict(X),s=100,color='blue')
    # Plota os dados previstos, a partir de todos os dados de X, tanto
de treino quanto de teste, como pontos azuis
sns.scatterplot(x=X_2022.to_numpy()[:,0],y=ylc_2022,s=100,color='red')
    # Plota os dados previstos para 2022 como pontos vermelhos

plt.legend(legenda, ['Ajuste', 'Teste', 'Previsões'], loc='upper left')
    # Definidos em 'legenda' acima. Gera a legenda das figuras
para identificar os pontos

plt.ylabel('Lucro Cessante')
    # Nome do eixo y
plt.xlabel('Ano')
    # Nome do eixo x

plt.savefig("Lucro Cessante\lucroXano.png")
    # Salva o gráfico com nome 'lucroXano' no formato .png na pasta
Lucro Cessante (precisa existir)

##### Lucro Cessante pelo Total de Carga
Transportada #####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_trainlc[:,2], y=y_trainlc,s=100,color='black')
sns.scatterplot(x=X_testlc[:,2], y=y_testlc,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,2],y=gbrlc.predict(X),s=100,color='blue')
sns.scatterplot(x=X_2022.to_numpy()[:,2],y=ylc_2022,s=100,color='red')

```

```

plt.legend(legenda, ['Ajuste', 'Teste', 'Previsões'], loc='upper left')

plt.ylabel('Lucro Cessante')
plt.xlabel('Total de Carga Transportada')

plt.savefig("Lucro Cessante\lucroXtotaldecarga.png")

##### Lucro Cessante pelo Total de Toneladas por Quilometro útil
#####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_trainlc[:,3], y=y_trainlc,s=100,color='black')
sns.scatterplot(x=X_testlc[:,3], y=y_testlc,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,3],y=gbrlc.predict(X),s=100,color='blue')
sns.scatterplot(x=X_2022.to_numpy()[:,3],y=ylic_2022,s=100,color='red')

plt.legend(legenda, ['Ajuste', 'Teste', 'Previsões'], loc='upper left')

plt.ylabel('Lucro Cessante')
plt.xlabel('Total de Carga por Quilometro Útil')

plt.savefig("Lucro Cessante\lucroXtoneladaquilometroutil")

##### Lucro Cessante pelo Número de Paradas
para Manutenção #####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_trainlc[:,4], y=y_trainlc,s=100,color='black')
sns.scatterplot(x=X_testlc[:,4], y=y_testlc,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,4],y=gbrlc.predict(X),s=100,color='blue')
sns.scatterplot(x=X_2022.to_numpy()[:,4],y=ylic_2022,s=100,color='red')

plt.legend(legenda, ['Ajuste', 'Teste', 'Previsões'], loc='upper left')

plt.ylabel('Lucro Cessante')
plt.xlabel('Número de Paradas para Manutenção')

plt.savefig("Lucro Cessante\lucroXnumerodeparadas.png")

##### Lucro Cessante pelo Custo do Pessoal de Serviços
#####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_trainlc[:,1], y=y_trainlc,s=100,color='black')
sns.scatterplot(x=X_testlc[:,1], y=y_testlc,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,1],y=gbrlc.predict(X),s=100,color='blue')
sns.scatterplot(x=X_2022.to_numpy()[:,1],y=ylic_2022,s=100,color='red')

plt.legend(legenda, ['Ajuste', 'Teste', 'Previsões'], loc='upper
left')

```

```

plt.ylabel('Lucro Cessante')
plt.xlabel('Custo de Pessoal de Serviços')

plt.savefig("Lucro Cessante\lucroXcustopessoalserv.png")

##### Lucro Cessante por Grupo de Dados de 2022 #####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_2022.to_numpy()[ :,5],y=y1c_2022,s=100,color='red')

plt.ylabel('Lucro Cessante')
plt.xlabel('Grupos de Dados')

plt.savefig("Lucro Cessante\grupo2022.png")

##### r2 dos Conjuntos de Treino e de Teste #####

train2_score = np.zeros((paramslc["n_estimators"],), dtype=np.float64)
for i, y_trainlc in enumerate(gbrlc.staged_predict(X_trainlc)):
    train2_score[i] = r2_score(y_trainlc, gbrlc.predict(X_trainlc))

test_score = np.zeros((paramslc["n_estimators"],), dtype=np.float64)
for i, y_trainlc in enumerate(gbrlc.staged_predict(X_trainlc)):
    test_score[i] = r2_score(y_testlc, gbrlc.predict(X_testlc))

fig = plt.figure(figsize=(9, 9))
plt.subplot(1, 1, 1)
plt.plot(
    np.arange(paramslc["n_estimators"]) + 1, train2_score, "b-",
    label="$r^2$ do Grupo de Treino",
)
plt.plot(
    np.arange(paramslc["n_estimators"]) + 1, test_score, "r-",
    label="$r^2$ do Grupo de Teste"
)
plt.legend(loc="lower right")
plt.xlabel("Número de Iterações")
plt.ylabel("$r^2$")
plt.xlim(0, 200)
plt.ylim(0.9, 1.01)
fig.tight_layout()

plt.savefig("Lucro Cessante\error2.png")

##### Relevância das Variáveis Independentes #####
feature_importance = gbrlc.feature_importances_
# Comando para a relevância das variáveis, da própria
biblioteca do GBR
sorted_idx = np.argsort(feature_importance)
pos = np.arange(sorted_idx.shape[0]) + 0.5

plt.figure(figsize=(9, 9))

plt.barh(pos, feature_importance[sorted_idx], align="center")
plt.yticks(pos, np.array(['Ano', 'PS', 'TCT', 'TKU', 'NPM'])[sorted_idx])

plt.xlabel('Relevância do Parâmetro')

```

```

plt.savefig("Lucro Cessante\importancia.png")

##### Operating Profit/Loss #####
##### Operating Profit/Loss per Year #####

plt.figure(figsize=(18, 9))
    # Define o tamanho da figura

sns.scatterplot(x=X_trainlc[:,0], y=y_trainlc,s=100,color='black')
    # Plota os dados de treino como pontos pretos
sns.scatterplot(x=X_testlc[:,0], y=y_testlc,s=100,color='orange')
    # Plota os dados de teste como pontos laranjas
sns.scatterplot(x=X.to_numpy()[:,0],y=gbrlc.predict(X),s=100,color='blue')
    # Plota os dados previstos, a partir de todos os dados de X, tanto
de treino quanto de teste, como pontos azuis
sns.scatterplot(x=X_2022.to_numpy()[:,0],y=y_2022,s=100,color='red')
    # Plota os dados previstos para 2022 como pontos vermelhos

plt.legend(legenda, ['Fit', 'Test', 'Predictions'], loc='upper left')
    # Definidos em 'legenda' acima. Gera a legenda das figuras
para identificar os pontos

plt.ylabel('Operating Profit/Loss')
    # Nome do eixo y
plt.xlabel('Year')
    # Nome do eixo x

plt.savefig("Lucro Cessante\lucroXanoen.png")
    # Salva o gráfico com nome 'lucroXano' no formato .png na pasta
Lucro Cessante (precisa existir)

##### Operating Profit/Loss per Total Load
Carried #####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_trainlc[:,2], y=y_trainlc,s=100,color='black')
sns.scatterplot(x=X_testlc[:,2], y=y_testlc,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,2],y=gbrlc.predict(X),s=100,color='blue')
sns.scatterplot(x=X_2022.to_numpy()[:,2],y=y_2022,s=100,color='red')

plt.legend(legenda, ['Fit', 'Test', 'Predictions'], loc='upper left')

plt.ylabel('Operating Profit/Loss')
plt.xlabel('Total Load Carried')

plt.savefig("Lucro Cessante\lucroXtotaldecargaen.png")

##### Operating Profit/Loss per Total Tons per
Useful Kilometer #####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_trainlc[:,3], y=y_trainlc,s=100,color='black')
sns.scatterplot(x=X_testlc[:,3], y=y_testlc,s=100,color='orange')

```

```

sns.scatterplot(x=X.to_numpy()[:,3],y=gbrlc.predict(X),s=100,color='blue')
sns.scatterplot(x=X_2022.to_numpy()[:,3],y=ylc_2022,s=100,color='red')

plt.legend(legenda, ['Fit', 'Test', 'Predictions'], loc='upper left')

plt.ylabel('Operating Profit/Loss')
plt.xlabel('Total Tons per Useful Kilometer')

plt.savefig("Lucro Cessante\lucroXtoneladaquilometroutilen.png")

##### Operating Profit/Loss per Number of
Maintenance Stops #####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_trainlc[:,4], y=y_trainlc,s=100,color='black')
sns.scatterplot(x=X_testlc[:,4], y=y_testlc,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,4],y=gbrlc.predict(X),s=100,color='blue')
sns.scatterplot(x=X_2022.to_numpy()[:,4],y=ylc_2022,s=100,color='red')

plt.legend(legenda, ['Fit', 'Test', 'Predictions'], loc='upper left')

plt.ylabel('Operating Profit/Loss')
plt.xlabel('Number of Maintenance Stops')

plt.savefig("Lucro Cessante\lucroXnumerodeparadasen.png")

##### Operating Profit/Loss per Service Personnel Cost #####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_trainlc[:,1], y=y_trainlc,s=100,color='black')
sns.scatterplot(x=X_testlc[:,1], y=y_testlc,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,1],y=gbrlc.predict(X),s=100,color='blue')
sns.scatterplot(x=X_2022.to_numpy()[:,1],y=ylc_2022,s=100,color='red')

plt.legend(legenda, ['Fit', 'Test', 'Predictions'], loc='upper left')

plt.ylabel('Operating Profit/Loss')
plt.xlabel('Service Personnel Cost')

plt.savefig("Lucro Cessante\lucroXcustopessoalserven.png")

##### Operating Profit/Loss per 2022 Data Groups #####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_2022.to_numpy()[:,5],y=ylc_2022,s=100,color='red')

plt.ylabel('Operating Profit/Loss')
plt.xlabel('Data Groups')

plt.savefig("Lucro Cessante\grupo2022en.png")

##### Train and Test Groups r2 #####

```



```

train2_score = np.zeros((paramslc["n_estimators"],), dtype=np.float64)
for i, y_trainlc in enumerate(gbrlc.staged_predict(X_trainlc)):
    train2_score[i] = r2_score(y_trainlc, gbrlc.predict(X_trainlc))

test_score = np.zeros((paramslc["n_estimators"],), dtype=np.float64)
for i, y_trainlc in enumerate(gbrlc.staged_predict(X_trainlc)):
    test_score[i] = r2_score(y_testlc, gbrlc.predict(X_testlc))

fig = plt.figure(figsize=(9, 9))
plt.subplot(1, 1, 1)
plt.plot(np.arange(paramslc["n_estimators"]) + 1, train2_score, "b-",
label="Train Group $r^2$",)
plt.plot(np.arange(paramslc["n_estimators"]) + 1, test_score, "r-",
label="Test Group $r^2$"
)
plt.legend(loc="lower right")
plt.xlabel("Number of Iterations")
plt.ylabel("$r^2$")
plt.xlim(0, 200)
plt.ylim(0.9, 1.01)
fig.tight_layout()

plt.savefig("Lucro Cessante\error2en.png")

##### Parameters Relevance #####

feature_importance = gbrlc.feature_importances_
# Comando para a relevância das
variáveis, da própria biblioteca do GBR
sorted_idx = np.argsort(feature_importance)
pos = np.arange(sorted_idx.shape[0]) + 0.5

plt.figure(figsize=(9, 9))

plt.barh(pos, feature_importance[sorted_idx], align="center")
plt.yticks(pos, np.array(['Ano', 'PS', 'TCT', 'TKU', 'NPM'])[sorted_idx])

plt.xlabel('Parameter Relevance')

plt.savefig("Lucro Cessante\importanciaen.png")

##### Custo de Manutenção Preventiva #####
##### Divisão em Conjuntos de Treino e de Teste #####

X_traincmp, X_testcmp, y_traincmp, y_testcmp = train_test_split(
# O train_test_split divide X e y em um conjunto
de treino e um de teste, na proporção definida pelo test_size
X.to_numpy(), ycmp.to_numpy().ravel(), test_size=0.4,
random_state=17
)

#### Definição dos Parâmetros do Gradient Boosting Regressor (GBR)
####

paramscmp = { # Junto com o test_size e o
random_state do train_test_split, estes são os parâmetros
determinados através do outro código

```

```

        "criterion": "squared_error",          # Tipo de erro, dentre as
opções do GBR, para seu ajuste
        "learning_rate": 0.1,                 # Peso (no mesmo sentido da
média ponderada) de cada novo ajuste
        "n_estimators": 659,                  # Número de iterações
        "min_samples_split": 2,              # Número mínimo de valores nos
nós das árvores para que haja uma divisão
        "min_samples_leaf": 1,               # Número mínimo de amostras em
uma folha
        "max_depth": 31,                     # Número máximo de divisões
de uma árvore
        "random_state": 12,
    }

##### Execução do GBR #####

gbrcmp = GradientBoostingRegressor(**paramscmp)          #
Definição do comando do GBR como 'gbrcmp'
gbrcmp.fit(X_traincmp, y_traincmp)                        #
Treinamento do algoritmo. Uso do GBR para o ajuste (fit) dos dados de
treino. Após esta linha, o termo 'gbr' está treinado com base nos
dados fornecidos.

y_predict = gbrcmp.predict(X_testcmp)                    # Predição dos
valores de y do conjunto de teste, a partir dos dados de X desse
conjunto por meio do algoritmo treinado. Este é o primeiro output.

print(r2_score(y_traincmp, gbrcmp.predict(X_traincmp))) # Mostra o r2
do grupo de treino no terminal. Idealmente é 1, pois o ajuste deve ser
perfeito.
print(r2_score(y_testcmp, gbrcmp.predict(X_testcmp)))   # Mostra o
r2 do grupo de teste no terminal. Idealmente deve estar acima de 0.95,
para garantir que o algoritmo faça boas previsões com dados que ele não
usou como treino.

##### Criação dos Gráficos #####

legenda = [Line2D([], [], marker='o', color='blue', linestyle='None'),
            # Algumas definições para fazer a legenda dos gráficos
            Line2D([], [], marker='o', color='orange', linestyle='None'),
            Line2D([], [], marker='o', color='red', linestyle='None')]

d2022 = pd.read_csv('2022.txt')
            # Importa os dados do arquivo 2022.txt, contendo dois valores de
entrada para cada variável independente para o ano de 2022 para
previsão

X_2022 = d2022[['Ano', 'PS', 'TU', 'TKU', 'NPM', 'Grupo']]
            # Define as variáveis independentes de 2022
ycmp_2022 = gbrcmp.predict(X_2022.to_numpy()[:, [0, 1, 2, 3, 4]])
            # Gera os valores do Custo de Manutenção Preventiva a partir da
previsão do GBR com os dados de 2022.

##### Custo de Manutenção Preventiva por Ano #####

plt.figure(figsize=(18, 9))
            # Define o tamanho da figura

```

```

sns.scatterplot(x=X_traincmp[:,0], y=y_traincmp,s=100,color='black')
    # Plota os dados de treino como pontos pretos
sns.scatterplot(x=X_testcmp[:,0], y=y_testcmp,s=100,color='orange')
    # Plota os dados de teste como pontos laranjas
sns.scatterplot(x=X.to_numpy()[:,0],y=gbrcmp.predict(X),s=100,color='bl
ue') # Plota os dados previstos, a partir de todos os dados de X, tanto
de treino quanto de teste, como pontos azuis
sns.scatterplot(x=X_2022.to_numpy()[:,0],y=ycmp_2022,s=100,color='red')
    # Plota os dados previstos para 2022 como pontos vermelhos

plt.legend(legenda, ['Ajuste', 'Teste', 'Previsões'], loc='upper left')
    # Definidos em 'legenda' acima. Gera a legenda das figuras
para identificar os pontos

plt.ylabel('Custo de Manutenção Preventiva')
    # Nome do eixo y
plt.xlabel('Ano')
    # Nome do eixo x

plt.savefig("Custo de Manutenção Preventiva\preventivaXano.png")
    # Salva o gráfico com nome 'lucroXano' no formato .png na
pasta Custo de Manutenção Preventiva (precisa existir)

### Custo de Manutenção Preventiva pelo Total de Carga Transportada ###

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_traincmp[:,2], y=y_traincmp,s=100,color='black')
sns.scatterplot(x=X_testcmp[:,2], y=y_testcmp,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,2],y=gbrcmp.predict(X),s=100,color='bl
ue')
sns.scatterplot(x=X_2022.to_numpy()[:,2],y=ycmp_2022,s=100,color='red')

plt.legend(legenda, ['Ajuste', 'Teste', 'Previsões'], loc='upper left')

plt.ylabel('Custo de Manutenção Preventiva')
plt.xlabel('Total de Carga Transportada')

plt.savefig("Custo de Manutenção
Preventiva\preventivaXtotaldecarga.png")

# Custo de Manutenção Preventiva pelo Total de Toneladas por Quilometro
útil ###

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_traincmp[:,3], y=y_traincmp,s=100,color='black')
sns.scatterplot(x=X_testcmp[:,3], y=y_testcmp,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,3],y=gbrcmp.predict(X),s=100,color='bl
ue')
sns.scatterplot(x=X_2022.to_numpy()[:,3],y=ycmp_2022,s=100,color='red')

plt.legend(legenda, ['Ajuste', 'Teste', 'Previsões'], loc='upper left')

plt.ylabel('Custo de Manutenção Preventiva')
plt.xlabel('Total de Carga por Quilometro útil')

```

```

plt.savefig("Custo de Manutenção
Preventiva\preventivaXtoneladaquilometrouil.png")

####Custo de Manutenção Preventiva pelo Número de Paradas para
Manutenção ####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_traincmp[:,4], y=y_traincmp,s=100,color='black')
sns.scatterplot(x=X_testcmp[:,4], y=y_testcmp,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,4],y=gbrcmp.predict(X),s=100,color='bl
ue')
sns.scatterplot(x=X_2022.to_numpy()[:,4],y=ycmp_2022,s=100,color='red')

plt.legend(legenda, ['Ajuste', 'Teste', 'previsões'], loc='upper left')

plt.ylabel('Custo de Manutenção Preventiva')
plt.xlabel('Numero de Paradas para Manutenção')

plt.savefig("Custo de Manutenção
Preventiva\preventivaXnumerodeparadas.png")

##### Custo de Manutenção Preventiva pelo Custo
do Pessoal de Serviços #####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_traincmp[:,1], y=y_traincmp,s=100,color='black')
sns.scatterplot(x=X_testcmp[:,1], y=y_testcmp,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,1],y=gbrcmp.predict(X),s=100,color='bl
ue')
sns.scatterplot(x=X_2022.to_numpy()[:,1],y=ycmp_2022,s=100,color='red')

plt.legend(legenda, ['Ajuste', 'Teste', 'previsões'], loc='upper left')

plt.ylabel('Custo de Manutenção Preventiva')
plt.xlabel('Custo de Pessoal de Serviços')

plt.savefig("Custo de Manutenção
Preventiva\preventivaXcustopessoalserv.png")

#### Custo de Manutenção Preventiva por Grupo de Dados de 2022 ####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_2022.to_numpy()[:,5],y=ycmp_2022,s=100,color='red')

plt.ylabel('Custo de Manutenção Preventiva')
plt.xlabel('Grupos de Dados')

plt.savefig("Custo de Manutenção Preventiva\grupo2022.png")

##### r2 dos Conjuntos de Treino e de Teste #####

train2_score = np.zeros((paramscmp["n_estimators"],), dtype=np.float64)
for i, y_traincmp in enumerate(gbrcmp.staged_predict(X_traincmp)):
    train2_score[i] = r2_score(y_traincmp, gbrcmp.predict(X_traincmp))

```

```

test_score = np.zeros((paramscmp["n_estimators"],), dtype=np.float64)
for i, y_traincmp in enumerate(gbrcmp.staged_predict(X_traincmp)):
    test_score[i] = r2_score(y_testcmp, gbrcmp.predict(X_testcmp))

fig = plt.figure(figsize=(9, 9))
plt.subplot(1, 1, 1)
plt.plot(
    np.arange(paramscmp["n_estimators"]) + 1, train2_score, "b-",
    label="$r^2$ do Grupo de Treino",
)
plt.plot(
    np.arange(paramscmp["n_estimators"]) + 1, test_score, "r-",
    label="$r^2$ do Grupo de Teste"
)
plt.legend(loc="lower right")
plt.xlabel("Número de Interações")
plt.ylabel("$r^2$")
plt.xlim(0, 200)
plt.ylim(0.9, 1.01)
fig.tight_layout()

plt.savefig("Custo de Manutenção Preventiva\error2.png")

### Relevância das Variáveis Independentes ###

feature_importance = gbrcmp.feature_importances_
# Comando para a relevância das
variáveis, da própria biblioteca do GBR
sorted_idx = np.argsort(feature_importance)
pos = np.arange(sorted_idx.shape[0]) + 0.5

plt.figure(figsize=(9, 9))

plt.barh(pos, feature_importance[sorted_idx], align="center")
plt.yticks(pos, np.array(['Ano', 'PS', 'TCT', 'TKU', 'NPM'])[sorted_idx])

plt.xlabel('Relevância do Parâmetro')

plt.savefig("Custo de Manutenção Preventiva\importancia.png")

##### Preventive Maintenance Cost #####
#####Preventive Maintenance Cost per Year #####

plt.figure(figsize=(18, 9))
# Define o tamanho da figura

sns.scatterplot(x=X_traincmp[:,0], y=y_traincmp,s=100,color='black')
# Plota os dados de treino como pontos pretos
sns.scatterplot(x=X_testcmp[:,0], y=y_testcmp,s=100,color='orange')
# Plota os dados de teste como pontos laranjas
sns.scatterplot(x=X.to_numpy()[:,0],y=gbrcmp.predict(X),s=100,color='blue')
# Plota os dados previstos, a partir de todos os dados de X,
tanto de treino quanto de teste, como pontos azuis
sns.scatterplot(x=X_2022.to_numpy()[:,0],y=ycmp_2022,s=100,color='red')
# Plota os dados previstos para 2022 como pontos
vermelhos

```

```

plt.legend(legenda, ['Fit', 'Test', 'Predictions'], loc='upper left')
        # Definidos em 'legenda' acima. Gera a legenda das
figuras para identificar os pontos

plt.ylabel('Preventive Maintenance Cost')
        # Nome do eixo y
plt.xlabel('Year')
        # Nome do eixo x

plt.savefig("Custo de Manutenção Preventiva\preventivaXanoen.png")
        # Salva o gráfico com nome 'lucroXano' no formato .png na
pasta Custo de Manutenção Preventiva (precisa existir)

##### Preventive Maintenance Cost per Total Load Carried #####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_traincmp[:,2], y=y_traincmp,s=100,color='black')
sns.scatterplot(x=X_testcmp[:,2], y=y_testcmp,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,2],y=gbrcmp.predict(X),s=100,color='bl
ue')
sns.scatterplot(x=X_2022.to_numpy()[:,2],y=ycmp_2022,s=100,color='red')

plt.legend(legenda, ['Fit', 'Test', 'Predictions'], loc='upper left')

plt.ylabel('Preventive Maintenance Cost')
plt.xlabel('Total Load Carried')

plt.savefig("Custo de Manutenção
Preventiva\preventivaXtotaldecargaen.png")

### Preventive Maintenance Cost per Total Tons per Useful Kilometer
###

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_traincmp[:,3], y=y_traincmp,s=100,color='black')
sns.scatterplot(x=X_testcmp[:,3], y=y_testcmp,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,3],y=gbrcmp.predict(X),s=100,color='bl
ue')
sns.scatterplot(x=X_2022.to_numpy()[:,3],y=ycmp_2022,s=100,color='red')

plt.legend(legenda, ['Fit', 'Test', 'Predictions'], loc='upper left')

plt.ylabel('Preventive Maintenance Cost')
plt.xlabel('Total Tons per Useful Kilometer')

plt.savefig("Custo de Manutenção
Preventiva\preventivaXtoneladaquilometroutilen.png")

#### Preventive Maintenance Cost per Number of Maintenance Stops ####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_traincmp[:,4], y=y_traincmp,s=100,color='black')
sns.scatterplot(x=X_testcmp[:,4], y=y_testcmp,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,4],y=gbrcmp.predict(X),s=100,color='bl
ue')

```

```

sns.scatterplot(x=X_2022.to_numpy()[:,4],y=ycmp_2022,s=100,color='red')

plt.legend(legenda, ['Fit', 'Test', 'Predictions'], loc='upper left')

plt.ylabel('Preventive Maintenance Cost')
plt.xlabel('Number of Maintenance Stops')

plt.savefig("Custo de Manutenção
Preventiva\preventivaXnumerodeparadasen.png")

#### Preventive Maintenance Cost per Service Personnel Cost ####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_traincmp[:,1], y=y_traincmp,s=100,color='black')
sns.scatterplot(x=X_testcmp[:,1], y=y_testcmp,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,1],y=gbrcmp.predict(X),s=100,color='blue')
sns.scatterplot(x=X_2022.to_numpy()[:,1],y=ycmp_2022,s=100,color='red')

plt.legend(legenda, ['Fit', 'Test', 'Predictions'], loc='upper left')

plt.ylabel('Preventive Maintenance Cost')
plt.xlabel('Service Personnel Cost')

plt.savefig("Custo de Manutenção
Preventiva\preventivaXcustopessoalservern.png")

##### Preventive Maintenance Cost per 2022 Data Groups #####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_2022.to_numpy()[:,5],y=ycmp_2022,s=100,color='red')

plt.ylabel('Preventive Maintenance Cost')
plt.xlabel('Data Groups')

plt.savefig("Custo de Manutenção Preventiva\grupo2022en.png")

##### Train and Test Groups r2 #####

train2_score = np.zeros((paramscmp["n_estimators"],), dtype=np.float64)
for i, y_traincmp in enumerate(gbrcmp.staged_predict(X_traincmp)):
    train2_score[i] = r2_score(y_traincmp, gbrcmp.predict(X_traincmp))

test_score = np.zeros((paramscmp["n_estimators"],), dtype=np.float64)
for i, y_traincmp in enumerate(gbrcmp.staged_predict(X_traincmp)):
    test_score[i] = r2_score(y_testcmp, gbrcmp.predict(X_testcmp))

fig = plt.figure(figsize=(9, 9))
plt.subplot(1, 1, 1)
plt.plot(
    np.arange(paramscmp["n_estimators"]) + 1, train2_score, "b-",
    label="Train Group  $r^2$ ",
)
plt.plot(
    np.arange(paramscmp["n_estimators"]) + 1, test_score, "r-",
    label="Test Group  $r^2$ "

```

```

)
plt.legend(loc="lower right")
plt.xlabel("Number of Iterations")
plt.ylabel("$r^2$")
plt.xlim(0, 200)
plt.ylim(0.9, 1.01)
fig.tight_layout()

plt.savefig("Custo de Manutenção Preventiva\error2en.png")

##### Parameters Relevance #####
feature_importance = gbrcmp.feature_importances_
                    # Comando para a relevância das variáveis, da própria
biblioteca do GBR
sorted_idx = np.argsort(feature_importance)
pos = np.arange(sorted_idx.shape[0]) + 0.5

plt.figure(figsize=(9, 9))

plt.barh(pos, feature_importance[sorted_idx], align="center")
plt.yticks(pos, np.array(['Ano', 'PS', 'TCT', 'TKU', 'NPM'])[sorted_idx])

plt.xlabel('Parameter Relevance')

plt.savefig("Custo de Manutenção Preventiva\importanciaen.png")

##### Custo de Manutenção Corretiva
##### Divisão em Conjuntos de Treino e de Teste

X_traincmc, X_testcmc, y_traincmc, y_testcmc = train_test_split(
    # O train_test_split divide X e y em um conjunto
de treino e um de teste, na proporção definida pelo test_size
    X.to_numpy(), ycmc.to_numpy().ravel(), test_size=0.3,
    random_state=90
)

##### Definição dos Parâmetros do Gradient Boosting Regressor (GBR)
#####
paramscmc = {
    # Junto com o test_size e o
random_state do train_test_split, estes são os parâmetros determinados
através do outro código
    "criterion": "squared_error",      # Tipo de erro, dentre as opções
do GBR, para seu ajuste
    "learning_rate": 0.09359098154363317, # Peso (no mesmo sentido
da media ponderada) de cada novo ajuste
    "n_estimators": 235,                # Número de interações
    "min_samples_split": 3,             # Número mínimo de valores nos
nós das Árvores para que haja uma divisão
    "min_samples_leaf": 1,             # Número mínimo de amostras em
uma folha
    "max_depth": 46,                   # Número máximo de divisões de
uma Árvore
    "random_state": 52,
}

##### Execução do GBR #####

```



```

gbrcmc = GradientBoostingRegressor(**paramscmc) #
Definição do comando do GBR como 'gbrcmc'
gbrcmc.fit(X_traincmc, y_traincmc) #
Treinamento do algoritmo. Uso do GBR para o ajuste (fit) dos dados de
treino. Após esta linha, o termo 'gbr' está treinado com base nos
dados fornecidos.

y_predict = gbrcmc.predict(X_testcmc) # Predição dos
valores de y do conjunto de teste, a partir dos dados de X desse
conjunto por meio do algoritmo treinado. Este é o primeiro output.

print(r2_score(y_traincmc, gbrcmc.predict(X_traincmc))) # Mostra o r2
do grupo de treino no terminal. Idealmente é 1, pois o ajuste deve ser
perfeito.
print(r2_score(y_testcmc, gbrcmc.predict(X_testcmc))) # Mostra o
r2 do grupo de teste no terminal. Idealmente deve estar acima de 0.95,
para garantir que o algoritmo faça boas previsões com dados que ele não
usou como treino.

##### Criação dos Gráficos #####

legenda = [Line2D([], [], marker='o', color='blue', linestyle='None'),
           # Algumas definições para fazer a legenda dos gráficos
           Line2D([], [], marker='o', color='orange', linestyle='None'),
           Line2D([], [], marker='o', color='red', linestyle='None')]

d2022 = pd.read_csv('2022.txt')
           # Importa os dados do arquivo 2022.txt, contendo dois valores de
entrada para cada variável independente para o ano de 2022 para
previsão

X_2022 = d2022[['Ano', 'PS', 'TU', 'TKU', 'NPM', 'Grupo']]
           # Define as variáveis independentes de 2022
ycmc_2022 = gbrcmc.predict(X_2022.to_numpy()[:, [0, 1, 2, 3, 4]])
           # Gera os valores do Custo de Manutenção Corretiva a partir da
previsão do GBR com os dados de 2022.

##### Custo de Manutenção Corretiva por Ano #####

plt.figure(figsize=(18, 9))
           # Define o tamanho da figura

sns.scatterplot(x=X_traincmc[:, 0], y=y_traincmc, s=100, color='black')
           # Plota os dados de treino como pontos pretos
sns.scatterplot(x=X_testcmc[:, 0], y=y_testcmc, s=100, color='orange')
           # Plota os dados de teste como pontos laranjas
sns.scatterplot(x=X.to_numpy()[:, 0], y=gbrcmc.predict(X), s=100, color='bl
ue') # Plota os dados previstos, a partir de todos os dados de X, tanto
de treino quanto de teste, como pontos azuis
sns.scatterplot(x=X_2022.to_numpy()[:, 0], y=ycmc_2022, s=100, color='red')
           # Plota os dados previstos para 2022 como pontos vermelhos

plt.legend(legenda, ['Ajuste', 'Teste', 'previsões'], loc='upper left')
           # Definidos em 'legenda' acima. Gera a legenda das figuras
para identificar os pontos

```

```

plt.ylabel('Custo de Manutenção Corretiva')
    # Nome do eixo y
plt.xlabel('Ano')
    # Nome do eixo x

plt.savefig("Custo de Manutenção Corretiva\corretivaXano.png")
    # Salva o gráfico com nome 'lucroXano' no formato .png na
    pasta Custo de Manutenção Corretiva (precisa existir)

## Custo de Manutenção Corretiva pelo Total de Carga Transportada ##

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_traincmc[:,2], y=y_traincmc,s=100,color='black')
sns.scatterplot(x=X_testcmc[:,2], y=y_testcmc,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,2],y=gbrcmc.predict(X),s=100,color='blue')
sns.scatterplot(x=X_2022.to_numpy()[:,2],y=ycmc_2022,s=100,color='red')

plt.legend(legenda, ['Ajuste', 'Teste', 'previsões'], loc='upper left')

plt.ylabel('Custo de Manutenção Corretiva')
plt.xlabel('Total de Carga Transportada')

plt.savefig("Custo de Manutenção Corretiva\corretivaXtotaldecarga.png")

##### Custo de Manutenção Corretiva pelo Total de Toneladas por
Quilometro Útil
plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_traincmc[:,3], y=y_traincmc,s=100,color='black')
sns.scatterplot(x=X_testcmc[:,3], y=y_testcmc,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,3],y=gbrcmc.predict(X),s=100,color='blue')
sns.scatterplot(x=X_2022.to_numpy()[:,3],y=ycmc_2022,s=100,color='red')

plt.legend(legenda, ['Ajuste', 'Teste', 'previsões'], loc='upper left')

plt.ylabel('Custo de Manutenção Corretiva')
plt.xlabel('Total de Carga por Quilometro útil')

plt.savefig("Custo de Manutenção
Corretiva\corretivaXtoneladaquilometroutil.png")

##### Custos de Manutenção Corretiva pelo Número de Paradas para
Manutenção #####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_traincmc[:,4], y=y_traincmc,s=100,color='black')
sns.scatterplot(x=X_testcmc[:,4], y=y_testcmc,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,4],y=gbrcmc.predict(X),s=100,color='blue')
sns.scatterplot(x=X_2022.to_numpy()[:,4],y=ycmc_2022,s=100,color='red')

plt.legend(legenda, ['Ajuste', 'Teste', 'previsões'], loc='upper left')

plt.ylabel('Custo de Manutenção Corretiva')

```

```

plt.xlabel('Número de Paradas para Manutenção')

plt.savefig("Custo de Manutenção
Corretiva\corretivaXnumerodeparadas.png")

##### Custo de Manutenção Corretiva pelo Custo do Pessoal de Serviços
#####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_traincmc[:,1], y=y_traincmc,s=100,color='black')
sns.scatterplot(x=X_testcmc[:,1], y=y_testcmc,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,1],y=gbrcmc.predict(X),s=100,color='blue')
sns.scatterplot(x=X_2022.to_numpy()[:,1],y=ycmc_2022,s=100,color='red')

plt.legend(legenda, ['Ajuste', 'Teste', 'previsões'], loc='upper left')

plt.ylabel('Custo de Manutenção Corretiva')
plt.xlabel('Custo de Pessoal de Serviços')

plt.savefig("Custo de Manutenção
Corretiva\corretivaXcustopessoalserv.png")

##### Custo de Manutenção Corretiva por Grupo de Dados de 2022 #####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_2022.to_numpy()[:,5],y=ycmc_2022,s=100,color='red')

plt.ylabel('Custo de Manutenção Corretiva')
plt.xlabel('Grupos de Dados')

plt.savefig("Custo de Manutenção Corretiva\grupo2022.png")

##### r2 dos Conjuntos de Treino e de Teste #####

train2_score = np.zeros((paramscmc["n_estimators"],), dtype=np.float64)
for i, y_traincmc in enumerate(gbrcmc.staged_predict(X_traincmc)):
    train2_score[i] = r2_score(y_traincmc, gbrcmc.predict(X_traincmc))

test_score = np.zeros((paramscmc["n_estimators"],), dtype=np.float64)
for i, y_traincmc in enumerate(gbrcmc.staged_predict(X_traincmc)):
    test_score[i] = r2_score(y_testcmc, gbrcmc.predict(X_testcmc))

fig = plt.figure(figsize=(9, 9))
plt.subplot(1, 1, 1)
plt.plot(
    np.arange(paramscmc["n_estimators"]) + 1, train2_score, "b-",
    label="$r^2$ do Grupo de Treino",
)
plt.plot(
    np.arange(paramscmc["n_estimators"]) + 1, test_score, "r-",
    label="$r^2$ do Grupo de Teste"
)
plt.legend(loc="lower right")
plt.xlabel("Número de Iterações")
plt.ylabel("$r^2$")

```

```

plt.xlim(0, 200)
plt.ylim(0.9, 1.01)
fig.tight_layout()

plt.savefig("Custo de Manutenção Corretiva\error2.png")

##### Relevância das Variáveis Independentes #####

feature_importance = gbrcmc.feature_importances_
                        # Comando para a relevância das
variáveis, da própria biblioteca do GBR
sorted_idx = np.argsort(feature_importance)
pos = np.arange(sorted_idx.shape[0]) + 0.5

plt.figure(figsize=(9, 9))

plt.barh(pos, feature_importance[sorted_idx], align="center")
plt.yticks(pos, np.array(['Ano', 'PS', 'TCT', 'TKU', 'NPM'])[sorted_idx])

plt.xlabel('Relevância do Parâmetro')

plt.savefig("Custo de Manutenção Corretiva\importancia.png")

##### Corrective Maintenance Cost
##### Corrective Maintenance Cost per Year

plt.figure(figsize=(18, 9))
    # Define o tamanho da figura

sns.scatterplot(x=X_traincmc[:,0], y=y_traincmc,s=100,color='black')
    # Plota os dados de treino como pontos pretos
sns.scatterplot(x=X_testcmc[:,0], y=y_testcmc,s=100,color='orange')
    # Plota os dados de teste como pontos laranjas
sns.scatterplot(x=X.to_numpy()[:,0],y=gbrcmc.predict(X),s=100,color='blue') # Plota os dados previstos, a partir de todos os dados de X, tanto
de treino quanto de teste, como pontos azuis
sns.scatterplot(x=X_2022.to_numpy()[:,0],y=ycmc_2022,s=100,color='red')
    # Plota os dados previstos para 2022 como pontos vermelhos

plt.legend(legenda, ['Fit', 'Test', 'Predictions'], loc='upper left')
    # Definidos em 'legenda' acima. Gera a legenda das figuras
para identificar os pontos

plt.ylabel('Corrective Maintenance Cost')
    # Nome do eixo y
plt.xlabel('Year')
    # Nome do eixo x

plt.savefig("Custo de Manutenção Corretiva\corretivaXanoen.png")
    # Salva o gráfico com nome 'lucroXano' no
formato .png na pasta Custo de Manutenção Corretiva (precisa existir)

#### Corrective Maintenance Cost per Total Load Carried ####

plt.figure(figsize=(18, 9))

```

```

sns.scatterplot(x=X_traincmc[:,2], y=y_traincmc,s=100,color='black')
sns.scatterplot(x=X_testcmc[:,2], y=y_testcmc,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,2],y=gbrcmc.predict(X),s=100,color='blue')
sns.scatterplot(x=X_2022.to_numpy()[:,2],y=ycmc_2022,s=100,color='red')

plt.legend(legenda, ['Fit', 'Test', 'Predictions'], loc='upper left')

plt.ylabel('Corrective Maintenance Cost')
plt.xlabel('Total Load Carried')

plt.savefig("Custo de Manutenção
Corretiva\corretivaXtotaldecargaen.png")

### Corrective Maintenance Cost per Total Tons per Useful Kilometer
####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_traincmc[:,3], y=y_traincmc,s=100,color='black')
sns.scatterplot(x=X_testcmc[:,3], y=y_testcmc,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,3],y=gbrcmc.predict(X),s=100,color='blue')
sns.scatterplot(x=X_2022.to_numpy()[:,3],y=ycmc_2022,s=100,color='red')

plt.legend(legenda, ['Fit', 'Test', 'Predictions'], loc='upper left')

plt.ylabel('Corrective Maintenance Cost')
plt.xlabel('Total Tons per Useful Kilometer')

plt.savefig("Custo de Manutenção
Corretiva\corretivaXtoneladaquilometroutilen.png")

### Corrective Maintenance Cost per Number of Maintenance Stops ####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_traincmc[:,4], y=y_traincmc,s=100,color='black')
sns.scatterplot(x=X_testcmc[:,4], y=y_testcmc,s=100,color='orange')
sns.scatterplot(x=X.to_numpy()[:,4],y=gbrcmc.predict(X),s=100,color='blue')
sns.scatterplot(x=X_2022.to_numpy()[:,4],y=ycmc_2022,s=100,color='red')

plt.legend(legenda, ['Fit', 'Test', 'Predictions'], loc='upper left')

plt.ylabel('Corrective Maintenance Cost')
plt.xlabel('Number of Maintenance Stops')

plt.savefig("Custo de Manutenção
Corretiva\corretivaXnumerodeparadasen.png")

### Corrective Maintenance Cost per Service Personnel Cost ####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_traincmc[:,1], y=y_traincmc,s=100,color='black')
sns.scatterplot(x=X_testcmc[:,1], y=y_testcmc,s=100,color='orange')

```

```

sns.scatterplot(x=X.to_numpy()[ :,1],y=gbrcmc.predict(X),s=100,color='blue')
sns.scatterplot(x=X_2022.to_numpy()[ :,1],y=ycmc_2022,s=100,color='red')

plt.legend(legenda, ['Fit', 'Test', 'Predictions'], loc='upper left')

plt.ylabel('Corrective Maintenance Cost')
plt.xlabel('Service Personnel Cost')

plt.savefig("Custo de Manutenção
Corretiva\corretivaXcustopessoalservern.png")

#### Corrective Maintenance Cost per 2022 Data Groups ####

plt.figure(figsize=(18, 9))

sns.scatterplot(x=X_2022.to_numpy()[ :,5],y=ycmc_2022,s=100,color='red')

plt.ylabel('Corrective Maintenance Cost')
plt.xlabel('Data Groups')

plt.savefig("Custo de Manutenção Corretiva\grupo2022.png")

#####Train and Test Groups r2 #####

train2_score = np.zeros((paramscmc["n_estimators"],), dtype=np.float64)
for i, y_traincmc in enumerate(gbrcmc.staged_predict(X_traincmc)):
    train2_score[i] = r2_score(y_traincmc, gbrcmc.predict(X_traincmc))

test_score = np.zeros((paramscmc["n_estimators"],), dtype=np.float64)
for i, y_traincmc in enumerate(gbrcmc.staged_predict(X_traincmc)):
    test_score[i] = r2_score(y_testcmc, gbrcmc.predict(X_testcmc))

fig = plt.figure(figsize=(9, 9))
plt.subplot(1, 1, 1)
plt.plot(
    np.arange(paramscmc["n_estimators"]) + 1, train2_score, "b-",
    label="Train Group  $r^2$ ",
)
plt.plot(
    np.arange(paramscmc["n_estimators"]) + 1, test_score, "r-",
    label="Test Group  $r^2$ "
)
plt.legend(loc="lower right")
plt.xlabel("Number of Iterations")
plt.ylabel(" $r^2$ ")
plt.xlim(0, 200)
plt.ylim(0.9, 1.01)
fig.tight_layout()

plt.savefig("Custo de Manutenção Corretiva\error2en.png")

#####parameter Relevance #####

feature_importance = gbrcmc.feature_importances_
# Comando para a relevância das variáveis, da própria
biblioteca do GBR
sorted_idx = np.argsort(feature_importance)

```

```

pos = np.arange(sorted_idx.shape[0]) + 0.5

plt.figure(figsize=(9, 9))

plt.barh(pos, feature_importance[sorted_idx], align="center")
plt.yticks(pos, np.array(['Ano', 'PS', 'TCT', 'TKU', 'NPM'])[sorted_idx])

plt.xlabel('Parameter Relevance')

plt.savefig("Custo de Manutenção Corretiva\importanciaen.png")

##### Variáveis Dependentes por Ano #####
fig, ax = plt.subplots(figsize=(18, 9))
fig.subplots_adjust(right=0.75)

twin1 = ax.twinx()
twin2 = ax.twinx()

twin2.spines.right.set_position(("axes", 1.2))

p1, = ax.plot(X.to_numpy()[:,0], gbrcmp.predict(X), "b-", marker="o",
label="Custo de Manutenção Preventiva")
p2, = twin1.plot(X.to_numpy()[:,0], gbrcmc.predict(X), "r-",
marker="o", label="Custo de Manutenção Corretiva")
p3, = twin2.plot(X.to_numpy()[:,0], X.to_numpy()[:,4], "g-",
marker="o", label='Número de Paradas para Manutenção')

ax.set_xlabel("Ano")
ax.set_ylabel("Custo de Manutenção Preventiva")
twin1.set_ylabel("Custo de Manutenção Corretiva")
twin2.set_ylabel("Número de Paradas para Manutenção")

ax.legend(handles=[p1, p2, p3])

plt.savefig("cmxnpmxanobr.png")

##### Dependent Variables per Year #####

fig, ax = plt.subplots(figsize=(18, 9))
fig.subplots_adjust(right=0.75)

twin1 = ax.twinx()
twin2 = ax.twinx()

twin2.spines.right.set_position(("axes", 1.2))

p1, = ax.plot(X.to_numpy()[:,0], gbrcmp.predict(X), "b-", marker="o",
label="Preventive Maintenance Cost")
p2, = twin1.plot(X.to_numpy()[:,0], gbrcmc.predict(X), "r-",
marker="o", label="Corrective Maintenance Cost")
p3, = twin2.plot(X.to_numpy()[:,0], X.to_numpy()[:,4], "g-",
marker="o", label='Number of Maintenance Stops')

ax.set_xlabel("Year")
ax.set_ylabel("Preventive Maintenance Cost")
twin1.set_ylabel("Corrective Maintenance Cost")

```

```

twin2.set_ylabel("Number of Maintenance Stops")

ax.legend(handles=[p1, p2, p3])

plt.savefig("cmxnpmxanoen.png")

##### Previsões para 2022 do Custo de Manutenção Preventiva e Corretiva
#####
plt.figure(figsize=(18, 9))

plt.ylabel('Custo de Manutenção Preventiva')
plt.xlabel('Grupos de Dados')
sns.lineplot(x=X_2022.to_numpy()[:,5],
y=ycmp_2022,color='red',marker="o")
#sns.lineplot(x=X.to_numpy()[:,0],
y=gbrcmc.predict(X),color='red',marker="o")

ax2 = plt.twinx()

sns.lineplot(x=X_2022.to_numpy()[:,5],
y=ycmc_2022,color='blue',ax=ax2,marker="o")

plt.legend(legenda2, ['Custo de Manutenção Preventiva', 'Custo de
Manutenção Corretiva'], loc='upper left')

plt.ylabel('Custo de Manutenção Corretiva')

plt.savefig("cmcp2022.png")

##### Preventive and Corrective Maintenance Cost Predictions for 2022
#####

plt.figure(figsize=(18, 9))

plt.ylabel('Preventive Maintenance Cost')
plt.xlabel('Data Groups')
sns.lineplot(x=X_2022.to_numpy()[:,5],
y=ycmp_2022,color='red',marker="o")
#sns.lineplot(x=X.to_numpy()[:,0],
y=gbrcmc.predict(X),color='red',marker="o")

ax2 = plt.twinx()

sns.lineplot(x=X_2022.to_numpy()[:,5],
y=ycmc_2022,color='blue',ax=ax2,marker="o")

plt.legend(legenda2, ['Preventive Maintenance Cost', 'Corrective
Maintenance Cost'], loc='upper left')

plt.ylabel('Corrective Maintenance Cost')

plt.savefig("cmcp2022en.png")

##### Previsões para 2022 o Lucro Cessante e os Custos de Manutenção
Preventiva e Corretiva #####
plt.figure(figsize=(18, 9))

```



```

plt.ylabel('Custo de Manutenção Preventiva')
plt.xlabel('Grupos de Dados')
sns.lineplot(x=X_2022.to_numpy()[:,5],
y=ycmp_2022,color='orange',marker="o")
#sns.lineplot(x=X.to_numpy()[:,0],
y=gbrcmc.predict(X),color='red',marker="o")

ax2 = plt.twinx()

sns.lineplot(x=X_2022.to_numpy()[:,5],
y=ycmc_2022,color='red',ax=ax2,marker="o")
sns.lineplot(x=X_2022.to_numpy()[:,5],
y=ylic_2022,color='blue',ax=ax2,marker="o")

plt.legend(legenda, ['Lucro Cessante','Custo de Manutenção Preventiva',
'Custo de Manutenção Corretiva'], loc='upper left')

plt.ylabel('Custo de Manutenção Corretiva/Lucro Cessante')

plt.savefig("cmpclc2022.png")

##### Operating Profit/Loss and Preventive and Corrective Maintenance
Cost Predictions for 2022 #####

plt.figure(figsize=(18, 9))

plt.ylabel('Preventive Maintenance Cost')
plt.xlabel('Data Groups')
sns.lineplot(x=X_2022.to_numpy()[:,5],
y=ycmp_2022,color='orange',marker="o")
#sns.lineplot(x=X.to_numpy()[:,0],
y=gbrcmc.predict(X),color='red',marker="o")

ax2 = plt.twinx()

sns.lineplot(x=X_2022.to_numpy()[:,5],
y=ycmc_2022,color='red',ax=ax2,marker="o")
sns.lineplot(x=X_2022.to_numpy()[:,5],
y=ylic_2022,color='blue',ax=ax2,marker="o")

plt.legend(legenda, ['Operational Profit/Loss','Preventive Maintenance
Cost', 'Corrective Maintenance Cost'], loc='upper left')

plt.ylabel('Corrective Maintenance Cost/Operational Profit/Loss')

plt.savefig("cmpclc2022en.png")

##### Matriz de Correlação #####
dcorr = do[['Ano','TU','TKU','NPM','PS','LC','CMP','CMC']]
# Define as variáveis que serão utilizadas

plt.figure(figsize=(18, 15))

corrMatrix = dcorr.corr()
sns.heatmap(corrMatrix, annot=True)

```

```
plt.savefig("corrmatrix.png")

##### Correlation Matrix #####
den = pd.read_csv('dadosen.txt') #
Importa os dados do arquivo dadosen.txt, com os nomes das variáveis em
inglês

dcorren = den[['Year', 'TU', 'TKU', 'NMS', 'PS', 'OP/L', 'PMC', 'CMC']]
# Define as variáveis que serão utilizadas

plt.figure(figsize=(18, 15))

corrMatrix = dcorren.corr() #
Gera a matriz de correlação de todas as variáveis utilizadas
sns.heatmap(corrMatrix, annot=True)
# Define o padrão de cores da matriz de correlação

plt.savefig("corrmatrixen.png")
```

