FACULDADE DE TECNOLOGIA SENAI MATO GROSSO

JEAN CARLOS DA CRUZ

ANÁLISE DE MODELOS DE APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DE FALHA MECÂNICA DE EQUIPAMENTOS FLORESTAIS

CUIABÁ – MT

2022

JEAN CARLOS DA CRUZ

ANÁLISE DE MODELOS DE APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DE FALHA MECÂNICA DE EQUIPAMENTOS FLORESTAIS

Trabalho de Conclusão de Curso apresentado à Banca Examinadora do Curso de Pós Graduação MBA em Data Science da Faculdade de Tecnologia SENAI MT, como requisito parcial para obtenção do grau de Especialista na área de Gestão e Administração.

Aluno: Jean Carlos da Cruz

Orientador: Denise Prado Kronbauer

CUIABÁ – MT

2022

TRABALHO DE CONCLUSÃO DE CURSO

ESTUDANTE:_____

APROVADO EM : _____/_____/_____

BANCA EXAMINADORA

_____

Assinatura do(a) Docente Orientador(a)

_____

Assinatura do(a) Docente Co-orientador(a)

_____

Avaliador 1

_____

Avaliador 2

Nota: _____

AGRADECIMENTOS

A Deus pela vida, saúde e oportunidades em meu caminho.

À minha família, por todo o amor, incentivo e dedicação para a minha formação pessoal e profissional, por celebrarem minhas conquistas e me abraçarem nas adversidades.

À empresa Teak Resources Company, pela gentileza na concessão dos dados utilizados neste trabalho.

À professora Denise Prado Kronbauer, pela orientação e revisão deste trabalho.

RESUMO

Com a intensificação de uma economia globalizada, houve um crescimento na demanda por equipamentos e sistemas com melhor desempenho aliado ao baixo custo.Também surgiu a necessidade de redução na probabilidade de ocorrência de falhas nos produtos, considerando que essas falhas podem levar a um aumento dos custos dos produtos e até a acidentes. Sendo assim, a confiabilidade pode ser definida como a probabilidade de um sistema operar de maneira satisfatória (sem falhas) em um período de tempo conhecido e em condições definidas no projeto. Em algumas situações, a confiabilidade deve ter um nível extremamente alto, devido às consequências que a falha pode ocasionar. Assim, a falha pode ser definida como uma inoperância de um produto, que não executa a função para a qual foi projetado. Uma falha pode gerar uma situação indesejada como uma simples parada de máquina, prejuízos financeiros, e até algo pior, como o risco para as vidas humanas. Logo, não devem ser poupados esforços para minimizar e evitar os riscos de uma falha. Desse modo, visando aumentar a precisão das manutenções preventivas, utiliza-se da aplicação de técnicas de aprendizado de máquina, as quais podem ser descritas como técnicas que utilizam algoritmos capazes de aprender de acordo com as respostas esperadas, por meio de associações de diferentes dados, os quais podem ser números, imagens e tudo que possa ser identificado por essa tecnologia. Ao invés de se executarem as rotinas de software manualmente, é aplicado um set específico de instruções para completar uma tarefa em particular; a máquina é “treinada” usando uma quantidade grande de dados e algoritmos que lhe dão a habilidade de aprender como executar a tarefa. Objetivou-se a análise de modelos de aprendizado de máquina com a finalidade de predizer o tempo médio de falha, utilizando um dataset com 534 dados de manutenções realizadas com reparo de terminais e mangueiras hidráulicas, de janeiro de 2020 a janeiro de 2021 em uma empresa com 42 equipamentos, localizada no estado de Mato Grosso. Foram analisados os modelos de: regressão linear múltipla, árvore de decisão, floresta aleatória, gradient boosting e regressão K vizinhos, , a partir do que se concluiu que tais modelos não apresentam boa eficácia

para predição do tempo médio de falha, em função da alta volubilidade das variáveis independentes e de problemas estruturais nos dados obtidos.

Palavras-chave: aprendizado de máquina; regressão; árvore aleatória; manutenção de equipamentos florestais.

LISTA DE FIGURAS

SUMÁRIO

1 INTRODUÇÃO

A função Manutenção representa uma importante atividade em qualquer operação mecanizada, independentemente da posição hierárquica que ocupa na corporação. As máquinas não são eternas e, durante a sua vida útil, precisam de um acompanhamento minucioso para que possam apresentar bom desempenho, com disponibilidade e confiabilidade favoráveis. Uma máquina, quando está em operação, custa caro. Parada, então, custa mais caro ainda, além de nada produzir (PACCOLA, 2017).

Nesse contexto, sem estratégias visando a redução das manutenções corretivas, ou seja, aquelas que acontecem quando o equipamento efetivamente falha, não é possível cumprir os cronogramas de produção, ter um portfólio de produtos de qualidade e controlar (e tentar minimizar) os custos de produção. Esses fatores interferem diretamente na competitividade da indústria no mercado e na fidelidade dos clientes, que buscam produtos de qualidade e custo justo.

Portanto, não é uma surpresa que a manutenção preventiva tenha emergido rapidamente como um dos principais casos de uso da Indústria 4.0. Sua implementação possibilita monitorar a integridade dos ativos, otimizar os cronogramas de manutenção e obter alertas em tempo real dos riscos operacionais, benefícios estes que vão de encontro com as necessidades atuais das organizações (EZRA, 2018).

Dentre as técnicas utilizadas visando o aumento da manutenção preventiva, podemos citar as técnicas de *Machine Learning ou ML* (em português Aprendizagem de Máquina), que é um método de auto aprendizado computadorizado que reside no centro da maioria dos aplicativos de Inteligência Artificial. Os modelos de ML combinam habilidades avançadas de aprendizado de padrões, com a capacidade de se adaptar à medida que as mudanças ocorrem nos dados de entrada (EZRA, 2018). Ou seja, seus algoritmos aprendem a partir dos dados a eles submetidos, e, assim, as máquinas são treinadas para aprender a executar diferentes tarefas de forma autônoma. Logo, ao serem expostas a novos dados, elas se adaptam, a partir dos cálculos anteriores, e os padrões se moldam para oferecer respostas confiáveis (RICHTER, 2019).

Sendo assim, este projeto se propõe a analisar modelos de aprendizado de máquina aplicados à base de dados histórica de manutenção de empresas florestais, com o objetivo de predizer potencial momento de falha dos equipamentos, aumentando potencialmente a precisão das manutenções preventivas realizadas nos equipamentos.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 FALHA EM EQUIPAMENTOS MÓVEIS

Diante da intensificação de uma economia globalizada, houve um crescimento na demanda por equipamentos e sistemas com melhor desempenho aliado ao baixo custo. Também surgiu a necessidade de redução na probabilidade de ocorrência de falhas nos produtos, considerando que essas falhas podem levar a um aumento dos custos dos produtos e até a acidentes. Com a análise e minimização dos diversos tipos de falhas, busca-se o aumento da confiabilidade dos equipamentos (FOGLIATTO; RIBEIRO, 2009).

De acordo com Anzanello *et al.* (2003), a confiabilidade pode ser definida como a probabilidade de um sistema operar de maneira satisfatória (sem falhas) em um período de tempo conhecido e em condições definidas no projeto. Em algumas situações, a confiabilidade deve ter um nível extremamente alto, devido às consequências que a falha pode ocasionar. As ocorrências de falhas são toleradas em algumas situações, mas podem gerar danos à imagem de uma empresa.

Assim, a falha pode ser definida como uma inoperância de um produto, que não executa a função para a qual foi projetado (WUTTKE; SELLITO, 2008). Uma falha pode gerar uma situação indesejada como uma simples parada de máquina, prejuízos financeiros, e até algo pior, como o risco para as vidas humanas. Logo, não devem ser poupados esforços para minimizar e evitar os riscos de uma falha (LAFRAIA, 2001).

Nesse contexto, novas técnicas foram desenvolvidas no setor de manutenção, dentre elas a manutenção preditiva, que tem como objetivo minimizar ou evitar a queda no desempenho, seguindo um plano previamente elaborado, baseado nos intervalos definidos de tempos em tempos, sempre visando prolongar a vida útil das máquinas e equipamentos, garantindo assim o aumento da eficiência e da produtividade (KARDEC; NASCIF, 2009).

2.2 APRENDIZADO DE MÁQUINA

Aprendizado de Máquina (do Inglês *Machine Learning*) é o nome que se dá em Ciência da Computação à técnica baseada nos princípios do aprendizado indutivo, no qual algoritmos processam um conjunto de dados e extraem um modelo

capaz de representar os intervalos de dados. Tais modelos podem ser usados também para representar um dado não amostrado (PERES et al., 2012).

As principais formas de aprendizado de máquinas são o aprendizado supervisionado e o não supervisionado. No aprendizado supervisionado, o algoritmo aprende a extrair informações de dados previamente conhecidos e classificados, sendo que, após a execução do modelo, testa-se a eficácia do aprendizado em dados desconhecidos. Já o aprendizado não supervisionado, consiste na obtenção de informações diretamente a partir de dados desconhecidos, por meio da inferência de características e padrões de similaridade com os dados que são conhecidos (LUCIAN, 2020). Este trabalho teve como propósito o estudo de algoritmos de aprendizado supervisionado, sendo eles: regressão linear múltipla, árvore de decisão, floresta aleatória, gradient boosting e regressão K vizinhos.

2.2.1 REGRESSÃO LINEAR MÚLTIPLA

Análise de regressão linear estuda a relação entre uma variável dependente (ou seja, aquela que se deseja conhecer ou estimar) e uma ou várias variáveis independentes (também chamadas de regressoras). Essa relação se dá por meio de um modelo matemático que associa a variável dependente com as independentes. Em regressão linear múltipla, são consideradas duas ou mais variáveis regressoras (independentes), ou seja, assumimos que existe uma relação linear entre uma variável dependente e *n* variáveis independentes (RODRIGUES, 2012). A regressão linear múltipla é definida pela seguinte fórmula:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \ i = 1, \ldots, n \ ,$$

Na qual,

yi representa o valor da variável dependente;

xi1, xi2…xin,i=1, n são os valores da i-ésima observação das *p* variáveis explicativas;

$\beta 0, \beta 1, \ \beta 2, \ldots, \ \beta n$ são os parâmetros ou coeficientes de regressão;

$\varepsilon i, \ i = 1, \ldots$, correspondem aos erros aleatórios.

2.2.2 ÁRVORE DE DECISÃO

Algoritmos que induzem árvores de decisão pertencem à família de algoritmos *Top Down Induction of Decision Tree,* em que uma árvore de decisão constitui uma estrutura de dados que pode ser definida recursivamente, como um *nó folha* que corresponde a uma classe, ou um *nó de decisão* que contém um teste sobre algum atributo de interesse. Em cada resultado do teste existe uma possível aresta para um subárvore. Cada subárvore produzida tem a mesma estrutura que a árvore principal (MONARD; BARANAUSKAS, 2003).

As árvores de decisão são descritas como modelos supervisionados não paramétricos, utilizados tanto para classificação quanto para regressão. O objetivo é criar um modelo que possa predizer a variável dependente por meio do aprendizado de simples decisões baseadas em regras inferidas pelas variáveis independentes nos dados (PEDREGOSA *et al*., 2011).

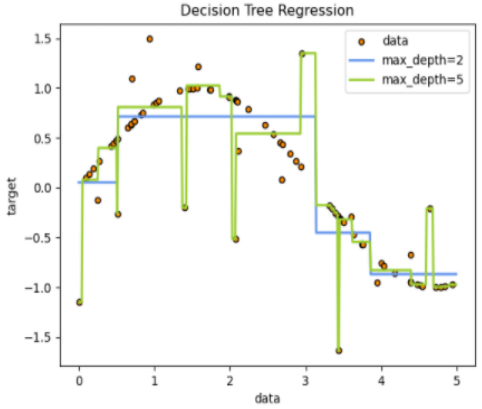

Figura 1 – Exemplo de Árvore de decisão aplicado a regressão (Fonte: Scikit-learn)

2.2.3 FLORESTA ALEATÓRIA

Florestas aleatórias são conjuntos de árvores de decisão criadas a partir de uma base de dados. Em florestas aleatórias, cada árvore de decisão é treinada com um subconjunto de instâncias da base de dados disponível, amostrado de forma aleatória repetidamente. Um algoritmo de floresta aleatória possui dois parâmetros principais: *ntree*, que corresponde ao número total de árvores que a floresta possui, e *mtry*, que corresponde à quantidade de variáveis amostradas em cada nó interno das árvores (RIQUETI; RIBEIRO; ZÁRATE, 2018).

As florestas aleatórias podem atingir boa generalização sem a necessidade de métodos de poda em suas árvores (reduções nas árvores de decisão realizadas para criar um poder de generalização artificial), o que torna o algoritmo mais eficiente. A classificação de novas instâncias se dá por um sistema de votação envolvendo cada árvore que compõe a floresta, de modo que a classe escolhida por mais árvores é assinalada na instância. Sendo assim, em linhas gerais, as florestas aleatórias atingem boa acurácia preditiva quando comparadas a outros métodos de aprendizado de máquina supervisionados. (FERNÁNDEZ-DELGADO *et al*., 2014).

2.2.4 GRADIENT BOOSTING

Em algoritmos de radient boosting, o processo de aprendizado ajusta novos modelos consecutivamente, para criar estimativas mais apuradas da variável dependente. A principal ideia é construir uma nova base de aprendizado para ser correlacionada ao máximo com o gradiente negativo da função de perda de todo o conjunto. A função de perda pode ser aplicada arbitrariamente, mas, de forma geral, a escolha da função de perda fica a cargo do analista ou da biblioteca que roda o modelo que assimila um valor padrão, sendo possível, também, a implementação de uma função de perda específica ajustada à tarefa em questão. Dessa forma, o gradient boosting proporciona muita liberdade durante a modelagem, tornando a escolha da função de perda mais apropriada uma questão de tentativa e erro (NATEKIN; KNOLL, 2013).

2.2.5 REGRESSÃO K VIZINHOS MAIS PRÓXIMOS

Esse método é do tipo não paramétrico, uma vez que não há um modelo a ser ajustado. O princípio que molda a regressão K vizinhos é a procura por um número predefinido de amostras de treinamento, que são as mais próximas em distância de um novo ponto, para dali predizer e catalogar. O número de amostras pode ser definido pelo usuário ou por bibliotecas de forma padrão, ou variar considerando a densidade local de pontos. Essa distância, em geral, pode ser qualquer métrica de medida, sendo a distância euclidiana a mais comum entre elas. Apesar de ser simples, *k vizinhos* obtêm bons resultados em problemas de classificação e regressão (PEDREGOSA *et al*., 2011). Em regressão, o conjunto formado pelos *k* padrões de treinamento mais próximos ao dado de entrada X.

Sendo assim, a saída para um novo dado de entrada X, pode ser escrita de uma forma geral como:

$$\hat{y}(\mathbf{x}') = \frac{\sum_{j=1}^{k} w_j y_j (\mathbf{x} \in \mathcal{N}_k(\mathbf{x}'))}{\sum w_j},$$

Em que,
wj,j=1,...,k representa o peso associado ao j-ésimo vizinho de X.

2.3 PYTHON

A linguagem orientada ao objeto Python tem em sua filosofia de design a legibilidade de código. Python busca auxiliar programadores por meio de escrita limpa, lógica e pequena, podendo ser utilizado em pequenos ou grandes projetos. (KUHLMAN, 2013). Criado por Guido van Rossum, no final dos anos 80, como sucessor da linguagem ABC, o Python teve sua primeira publicação em 1991 na versão 0.9.0 (ROSSUM, 2009). Python tem por definição a fácil legibilidade de código, sua formatação é organizada e comumente usa palavras em inglês, enquanto outras línguas utilizam pontuações. Python, em comparação a outras linguagens, não utiliza colchetes para delimitar blocos, apresenta também menos exceções sintáticas e casos especiais que linguagens como C ou Pascal costumam apresentar (PYTHON SOFTWARE FOUNDATION, 2012).

Atualmente, a grande quantidade de pacotes em bibliotecas disponíveis é considerada uma das maiores forças da linguagem Python (PIOTROWSKI, 2006). Em setembro de 2021, o repositório oficial de pacotes de terceiros continha mais de 300.000 pacotes com os mais diversos objetivos, tais quais: análise de dados, aprendizado de máquina, desenvolvimento de aplicativos mobile, processamento de texto, processamento de imagens, etc.

2.4 PYCARET

A biblioteca de código aberto em Python PyCaret, é uma ferramenta de baixo código que automatiza fluxos de trabalho de aprendizado de máquina. A biblioteca possui uma solução de ponta a ponta de gestão de modelos que acelera de forma

exponencial o ciclo de experimentação e aumenta a produtividade dos projetos que a utilizam (ALI, 2020).

Quando comparada a outras bibliotecas de aprendizado de máquina de código aberto, a PyCaret pode ser utilizada para substituir centenas de linhas de código com algumas poucas linhas. A biblioteca encapsula diversas outras bibliotecas de aprendizado de máquina, tais quais: scikit-learn, XGBoost, LightGBM, CatBoost, spaCy, Optuna, Hyperopt, Ray, etc. (ALI, 2020).

Seu design e simplicidade têm como fonte de inspiração a sua aplicação no contexto de ciência de dados, no qual tarefas analíticas de simples e moderada sofisticação podem ser performadas de maneira mais simples e rápida do que em outros modelos e até mesmo exigir menos expertise técnica (ALI, 2020).

3 MATERIAIS E MÉTODOS

3.1 LOCAL E PROCESSOS

O presente estudo foi realizado a partir de dados de manutenção mecânica realizados em equipamentos que atuam na etapa de colheita florestal de *Tectona grandis* L. F. (Teca) em três regiões do estado de Mato Grosso (Cáceres, Rosário Oeste e Tangará da Serra). A empresa dispõe de 42 equipamentos que atuam em três etapas no processo de colheita, sendo eles: corte, arraste, baldeio e carga. A etapa de corte consiste na efetiva derrubada de árvores com equipamentos florestais, como Harvesters ou Feller Bunchers. A etapa de arraste consiste na locomoção das árvores do local de queda até as linhas de operação em campo, sendo realizada por garras florestais acopladas em tratores agrícolas de 100 a 150 CV. O baldeio é realizado após o processamento das árvores em toras nas linhas de operação, atividade que locomove e empilha as toras nos carreadores das fazendas, processo realizado por carretas agrícolas autocarregáveis acopladas a tratores agrícolas de 100 a 150 CV. A carga, que é o processo final de colheita, no qual as toras são carregadas nos caminhões de transporte, é realizada por tratores agrícolas com mais de 150 CV acoplados com garras florestais.

A empresa possui departamento de manutenção próprio, composto por um time de mais de 10 pessoas (entre líderes, mecânicos e auxiliares). Essa equipe de manutenção se divide e fica alocada junto às frentes de colheita em regime de plantão enquanto a operação acontece. A equipe realiza as manutenções de duas formas, sendo elas: preventivas ou corretivas. As preventivas são realizadas em função do programa proposto pelo fabricante de cada equipamento e/ou alguma manutenção programada; as manutenções corretivas são aquelas que acontecem sempre que os equipamentos apresentam falhas mecânicas em operação.

O time de manutenção da empresa dispõe de software de mercado que auxilia na coleta de dados de manutenção em campo, por meio de um aplicativo android customizado que funciona com um formulário digital, no qual o auxiliar faz o apontamento dos dados referentes a cada manutenção. Uma vez inseridos, os dados são sincronizados com o banco de dados Microsoft SQL, server da companhia no qual as informações são armazenadas.

3.2 BASE DE DADOS

Ao total foram fornecidos 2.852 registros, coletados entre os meses de janeiro de 2020 a janeiro de 2022. Os dados foram fornecidos em formato .xlsx (Microsoft Excel). Os campos contidos no banco de dados são:

Campo	Tipo	Descrição
MAINTENANCE_ID	int	Código único utilizado para identificar cada manutenção
START_TIME	datetime	Início da Manutenção Realizada
END_TIME	datetime	Término da Manutenção Realizada
PART_TYPE	object	Tipo de peça utilizada na manutenção
SUB_SYSTEM	object	Subgrupo de sistema (Nomenclatura Interna)
MAINTENANCE_LOCAL	object	Local onde a Manutenção foi realizada
GROUP_NAME	object	Nome do grupo de manutenção
MAINTENANCE_TYPE	object	Tipo de manutenção realizada (Corretiva ou Preventiva)
SERVICE_NAME	object	O nome do serviço que foi realizado na manutenção
OBSERVATION	object	Observações que os mecânicos podem fazer em cada manutenção
PART_NUMBER	float	Número total de peças utilizado na manutenção
MACHINERY_ID	object	Código único utilizado para identificar cada equipamento
MACHINERY_HOURMETER	int	Horímetro do equipamento no momento do início da manutenção
MECHANIC	object	Nome do mecânico responsável pela manutenção
PURCHASE_DAT	datetim	Data em que o equipamento foi adquirido

E	e	
BUILT_YEAR	float	Ano que o equipamento foi montado
AGE	object	Idade dos equipamentos em meses quando a manutenção aconteceu
OPERATION	object	Operação Realizada pelo equipamento (Corte, Arraste, Baldeio ou Carga)

Quadro 1 – Tabela descritiva das variáveis disponíveis (Fonte: Teak Resources Company)

3.3 PROCESSAMENTO E MODELAGEM DOS DADOS

O processamento dos dados foi realizado utilizando a linguagem python e as seguintes bibliotecas open source: Numpy, Pandas, csv, datetime, seaborn e PyCaret. O código utilizado foi construído por meio do framework de notebook, utilizando a ferramenta Jupyter em nuvem, disponível no serviço grátis Google Colaboratory. Para ser realizado, o trabalho foi dividido em três etapas;a primeira delas sendo a etapa de limpeza e tratamento dos dados recebidos, em que dados nulos e duplicados foram tratados. Foram removidas as colunas *GROUP_NAME, MECHANIC, PURCHASE_DATE* e *BUILT_YEAR,* uma vez que essas não têm utilização prática para o objetivo deste trabalho. Na base de dados havia informação de manutenção de equipamentos que não estão relacionados com a colheita florestal, esses registros foram removidos do conjunto.

Após a limpeza e tratamento dos dados, realizou-se uma análise exploratória destes, análise que seguiu guiada por oito perguntas relevantes para o projeto, sendo elas:

1- Qual equipamento e operação apresentou o maior número de manutenções?;
2- Qual é o tempo médio para manutenção em horas por mês?;
3- Qual é o tempo médio entre manutenções por máquina e grupo de operação?;
4- Das manutenções realizadas, quantas foram corretivas e quantas preventivas?;
5- Quais são as principais peças usadas nas manutenções?;
6- Qual a estação operacional com mais manutenções registradas?;
7- Qual é o tempo médio para falha por máquina e grupo de operação?;
8- Qual é a idade média dos equipamentos por grupo de operação?

Com base na construção de respostas às perguntas criadas na etapa de análise exploratória, seguiu-se para a etapa de modelagem dos dados, utilizando a biblioteca PyCaret, por esta apresentar uma solução simples e de baixo código para teste e experimentação de modelos. Os modelos selecionados foram: regressão linear múltipla, árvore de decisão, floresta aleatória, gradient boosting e regressão K vizinhos, por se tratar de modelos de ampla utilização em nível acadêmico e privado. Definiu-se o erro absoluto médio e o erro quadrático médio como as métricas de sucesso dos modelos a serem analisados. A primeira métrica é utilizada comumente na avaliação de modelos, mede o erro entre pares de observações e é de fácil explicação, uma vez que é construída na mesma unidade dos dados. O erro quadrático também mede a magnitude do erro, porém, ao obter o quadrado antes da média, leva a uma maior ponderação de altos erros. Antes da efetiva aplicação dos modelos de regressão, testou-se a distribuição das variáveis independentes que foram utilizadas, bem como a correlação entre elas: o método de correlação utilizado foi o de Pearson.

Após o primeiro teste dos modelos aplicados aos dados, obteve-se os resultados das métricas de sucesso; com base nos resultados, selecionou-se o melhor modelo entre os analisados, e com base nessa seleção a ferramenta de “Tunning” do PyCaret foi utilizada com o objetivo de melhorar os resultados obtidos através do aumento do número de interações do modelo com os dados e alterando a métrica otimizadora para o erro médio absoluto. Após a etapa de tunning, realizou-se a análise de importância de cada variável independente utilizada no melhor modelo, bem como a análise de distribuição dos erros.

Por fim, com base em todos os resultados obtidos, tanto na etapa de exploração quanto na de modelagem, foi possível gerar as conclusões e recomendações do projeto.

4 RESULTADOS E DISCUSSÃO

4.1 RESULTADOS BASEADOS NA ANÁLISE EXPLORATÓRIA DOS DADOS

Por meio das perguntas formuladas na etapa de análise exploratória dos dados, observou-se padrões e tendências nos dados que serviram de embasamento para a seleção de variáveis independentes visando a predição do momento de falha.

4.1.1 QUAL EQUIPAMENTO E OPERAÇÃO APRESENTOU O MAIOR NÚMERO DE MANUTENÇÕES?

Para identificar a escala de grandeza de manutenções por cada equipamento contido nos dados, foi criado um gráfico de barras ordenando os equipamentos por quantidade de manutenções individuais (Figura 2).

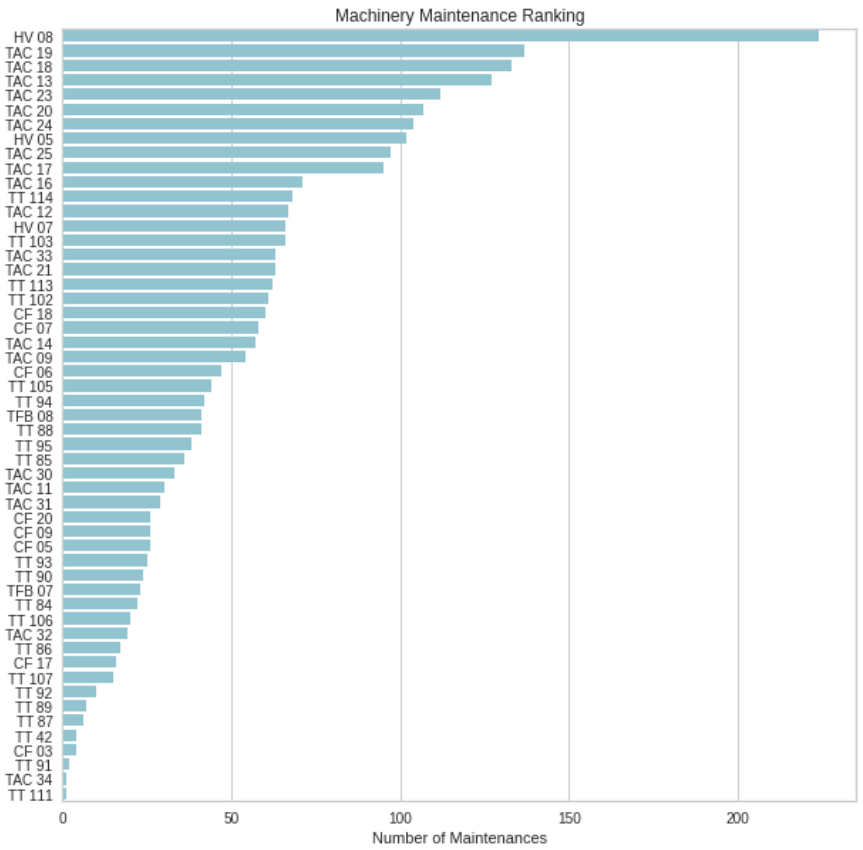

Figura 2 – Gráfico de barra ordenado para o número de manutenções por equipamento

Foi possível evidenciar que o equipamento HV08 é o equipamento com o maior número de manutenções registradas. Em linhas gerais, é possível evidenciar também que equipamentos com o código TAC, que são usados na atividade de

baldeio, mostram-se fortemente presentes entre os dez primeiros registros visualizados, indicando que essa grupo de atividade tem a maior frequência de manutenções entre todas as analisadas, representando mais de 50% do total de manutenções realizadas (Figura 3).

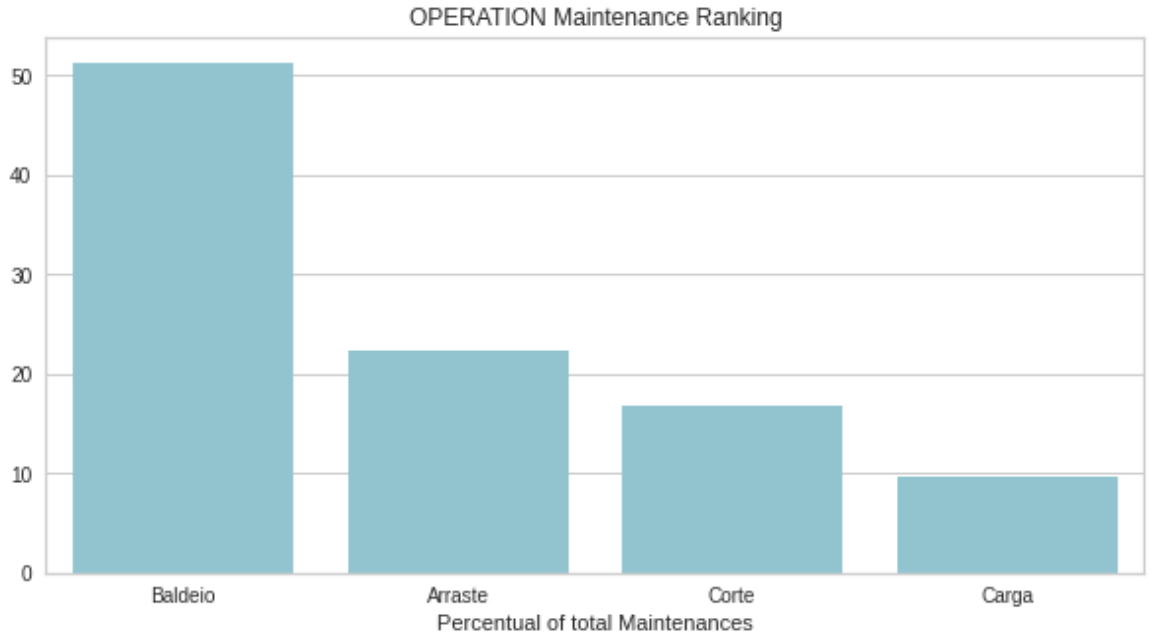

Figura 3 – Gráfico de barra ordenado para o número de manutenções por grupo de operação

4.1.2 QUAL É O TEMPO MÉDIO PARA MANUTENÇÃO EM HORAS POR MÊS?

O tempo médio para reparo nas manutenções dentro do conjunto de dados é de 43.96 horas. Ao analisar o resultado consolidado mês a mês, temos picos de aumento do tempo médio em relação à média; como exemplo, o mês de agosto de 2020, no qual o tempo médio para manutenção foi superior a 300 horas (Figura 4), indicando que algumas manutenções nesse período foram demasiado longas em relação ao habitual. O fenômeno provavelmente se explica por uma falha mais complexa que demandou maiores cuidados, ou aquisição de peça de reposição que não estava disponível em estoque no momento que a falha ocorreu.

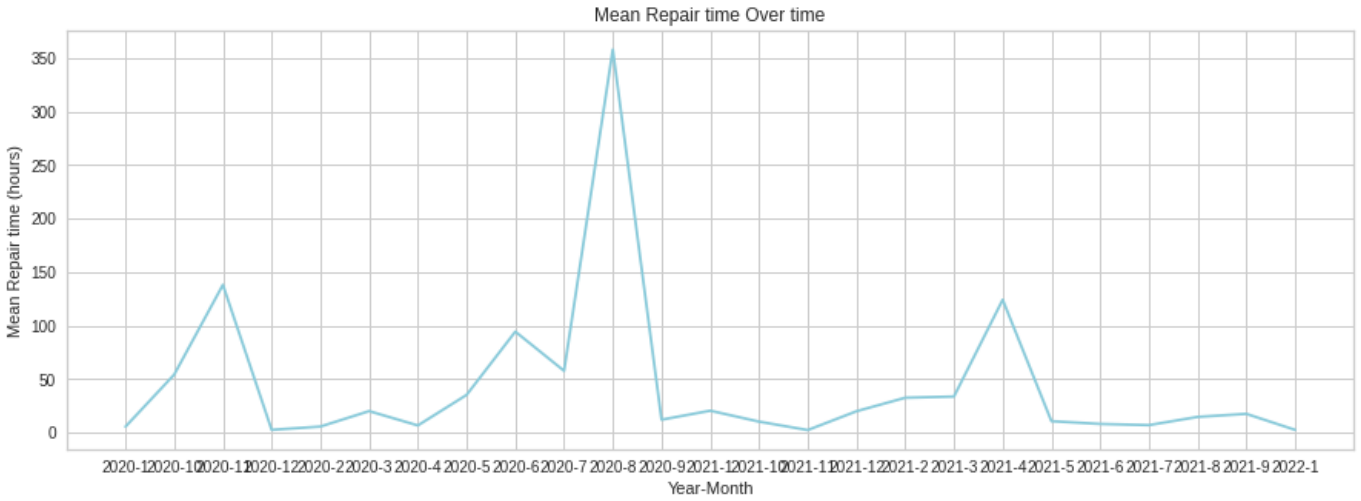

Figura 4 – Gráfico de linha ordenado por mês para o tempo médio para manutenção

4.1.3 QUAL É O TEMPO MÉDIO ENTRE MANUTENÇÕES POR MÁQUINA E GRUPO DE OPERAÇÃO?

Para o cálculo do tempo médio entre manutenções, um novo campo foi calculado, uma vez que essa informação não estava disponível de forma explícita no conjunto de dados. Essa variável foi obtida por meio da diferença entre os horímetros registrados da manutenção analisada e da anterior, gerando tal diferença em horas. O cálculo dessa variável gerou os primeiros indicadores, tornando possível evidenciar resultados incoerentes; como exemplo, tempos médios negativos (Figura 5): esse fato indicou que existia algum problema nos dados de horímetro apontado em cada manutenção.

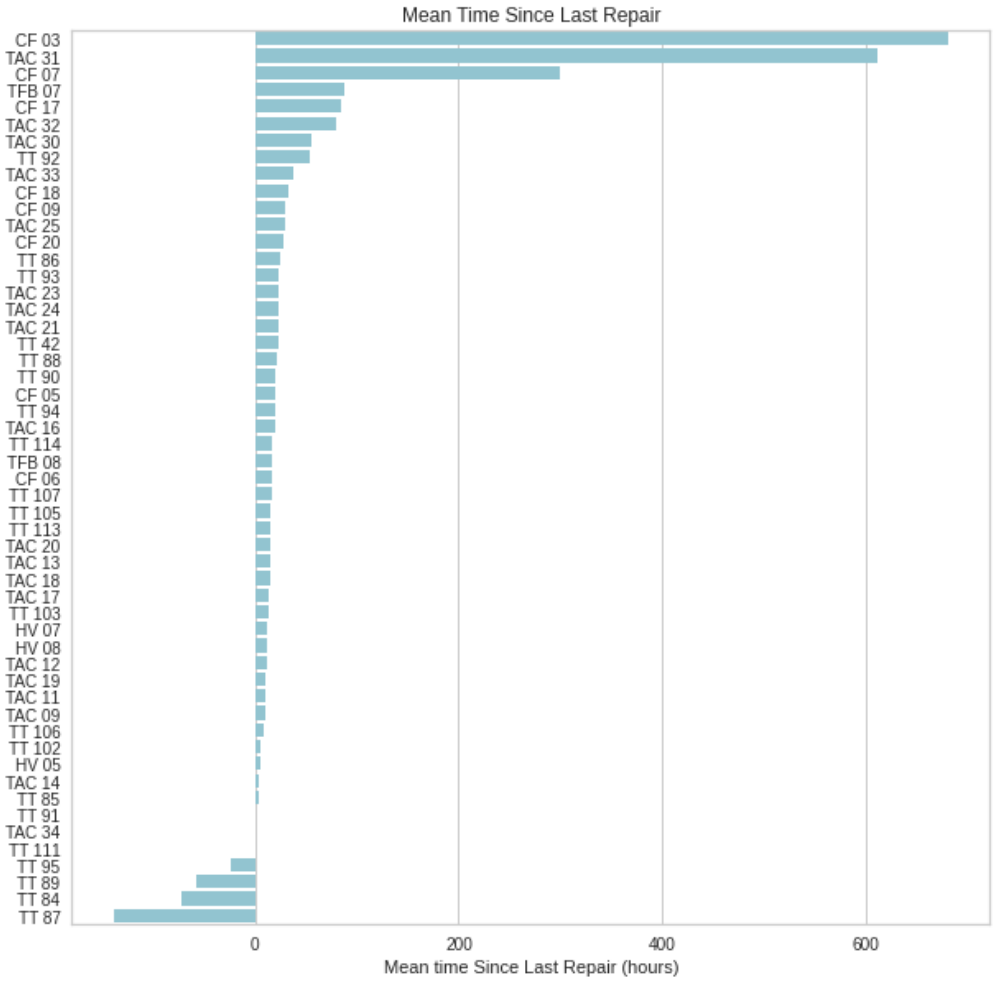

Figura 5 – Gráfico de linha ordenado por máquina para o tempo médio entre manutenções

Sendo assim, não poderíamos utilizar tais números da maneira que foram disponibilizados na modelagem futura, sendo requerida uma análise mais a fundo e possíveis correções. Para tal, plotamos a dispersão do tempo médio entre manutenções por máquina (Figura 6).

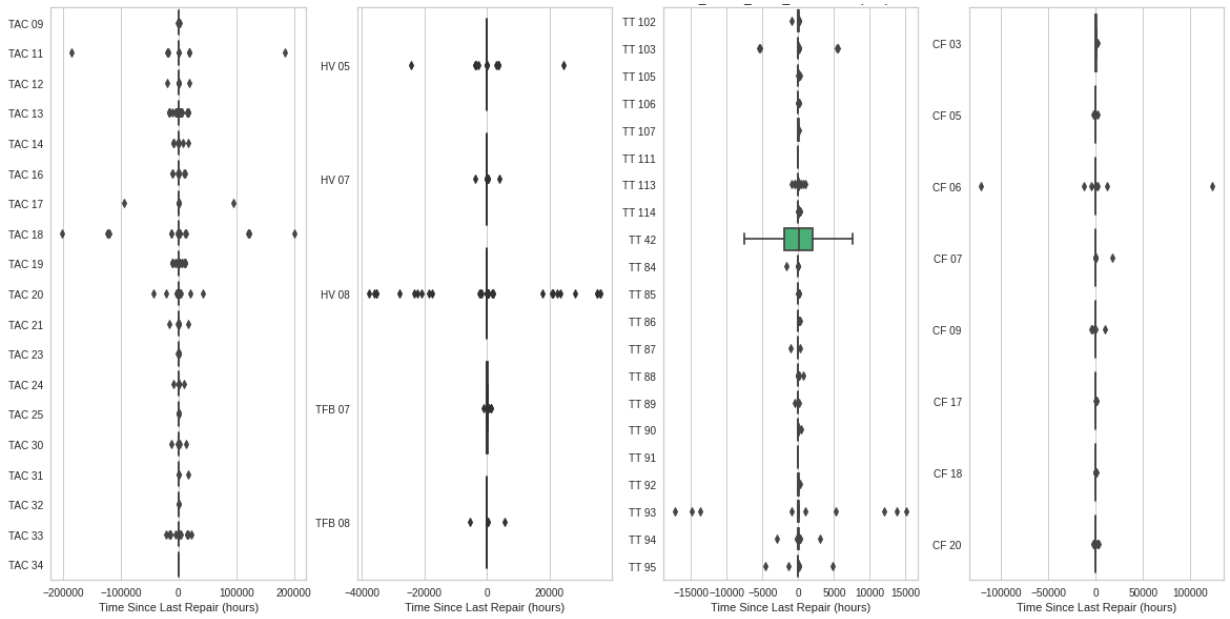

Figura 6 – Gráfico de dispersão por máquina para o tempo médio entre manutenções

Pela Figura 6 fica evidente que existem muitos outliers nos dados. Provavelmente esses problemas se devem a falhas na digitação, erros de sincronização, erro entre o apontamento em papel e o aplicativo, falha da marcação do horímetro na máquina, etc. De qualquer maneira, para que essa variável pudesse ser utilizada posteriormente, foi necessário fazer a correção desses valores no conjunto de dados. Essa correção foi realizada em associação com os responsáveis pela manutenção da empresa, sendo que foram realizados 262 ajustes nas informações de horímetro, e um novo gráfico de dispersão foi criado (Figura 7). Com as correções, os resultados agora refletem a realidade enfrentada em campo.

De acordo com os dados, é possível verificar que o tempo médio entre manutenções gerais é de 19 horas. Quando analisamos o tempo médio entre manutenções por grupo de operação, podemos ver que o maior tempo foi evidenciado na operação de carga (26 horas), seguido por baldeio (21 horas), arraste (16 horas) e corte (13 horas), respectivamente.

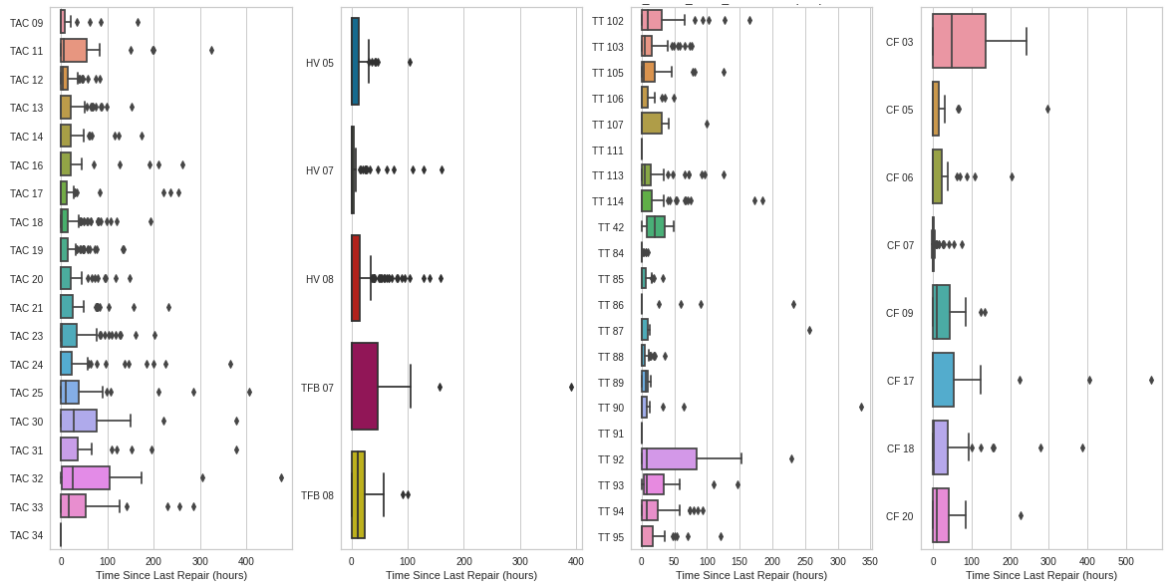

Figura 7 – Gráfico de dispersão ajustado por máquina para o tempo médio entre manutenções

4.1.4 DAS MANUTENÇÕES REALIZADAS, QUANTAS SÃO CORRETIVAS E QUANTAS PREVENTIVAS?

Do total de manutenções realizadas no conjunto de dados, 68,4% são manutenções corretivas, 29,6% são manutenções preventivas e 2%, manutenções periódicas. Quando visualizadas em uma escala mensal de tempo, podemos ver uma grande heterogeneidade no número de manutenções corretivas e preventivas, e estabilidade no número de periódicas (Figura 8).

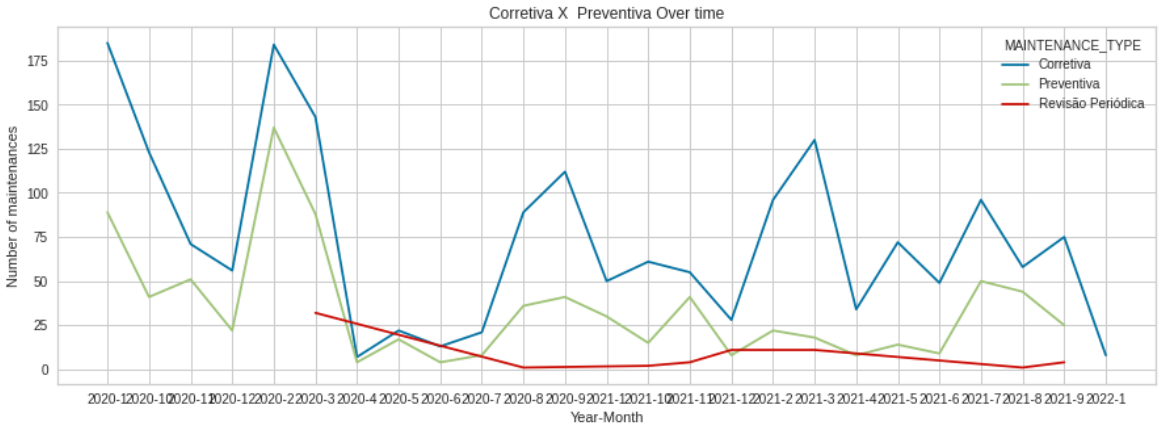

Figura 8 – Gráfico de linha ordenado por mês para o número de manutenções realizadas por tipo

Abril, maio e junho de 2020 tiveram o menor número em cada tipo de manutenção, provavelmente devido a pandemia de COVID, quando muitas

operações foram paralisadas. É possível evidenciar também que janeiro, fevereiro e março de 2020, foram os meses com o maior número de manutenções preventivas realizadas; após o período de fechamentos por conta da pandemia, tais números nunca retomaram ao mesmo patamar.

4.1.5 QUAIS AS PRINCIPAIS PEÇAS USADAS NAS MANUTENÇÕES?

Terminal Hidráulico e Mangueira Hidráulica são as peças com maior frequência de utilização no conjunto de dados: aproximadamente 40% do total das manutenções, quando combinadas (Figura 9). Como foi possível evidenciar que temos diversas outras peças utilizadas nas manutenções, decidiu-se que consideraríamos, para o futuro modelo de predição, apenas as manutenções desses dois tipos de peça, assim reduzimos a dispersão nos dados e foi possível focar no tipo de manutenção mais frequente, ou seja, o maior problema da companhia.

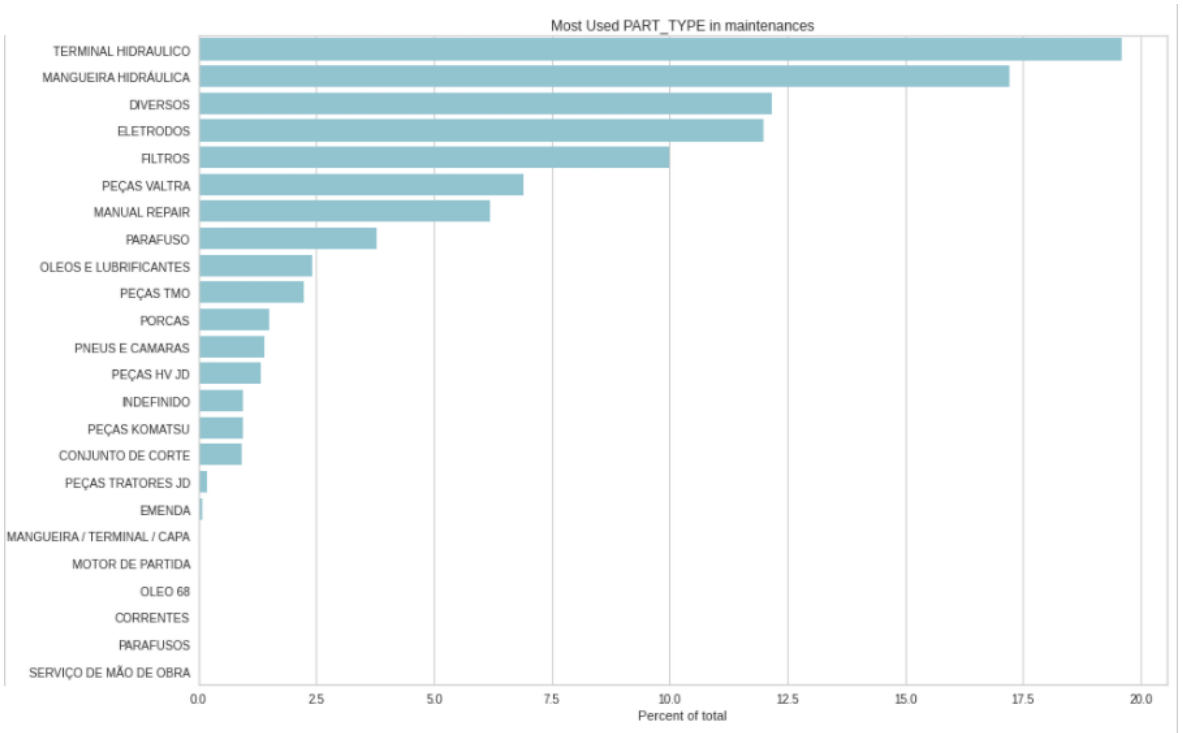

Figura 9 – Gráfico de barra ordenado por peça para o percentual do total de manutenções realizadas

4.1.6 QUAL A ESTAÇÃO OPERACIONAL COM MAIS MANUTENÇÕES REGISTRADAS?

A estação operacional é definida pela companhia como estação seca, que corresponde aos meses de abril, maio, junho, julho, agosto e setembro, período em

que temos os meses mais secos do ano no estado do Mato Grosso. A estação chuvosa é definida pelos meses de outubro, novembro, dezembro, janeiro, fevereiro e março. Foi possível evidenciar que 65% do total de manutenções ocorreram na estação chuvosa (Figura 10). Esse resultado faz sentido, uma vez que, com o período chuvoso, o terreno florestal e as próprias toras ficam mais pesadas em função da água acumulada, exigindo maior esforço dos equipamentos. Entretanto, esse mesmo resultado pode apresentar algum viés, uma vez que evidenciamos no item 4.1.4 que na estação seca de 2020 ocorreram os piores meses da pandemia de COVID, e algumas operações foram paralisadas.

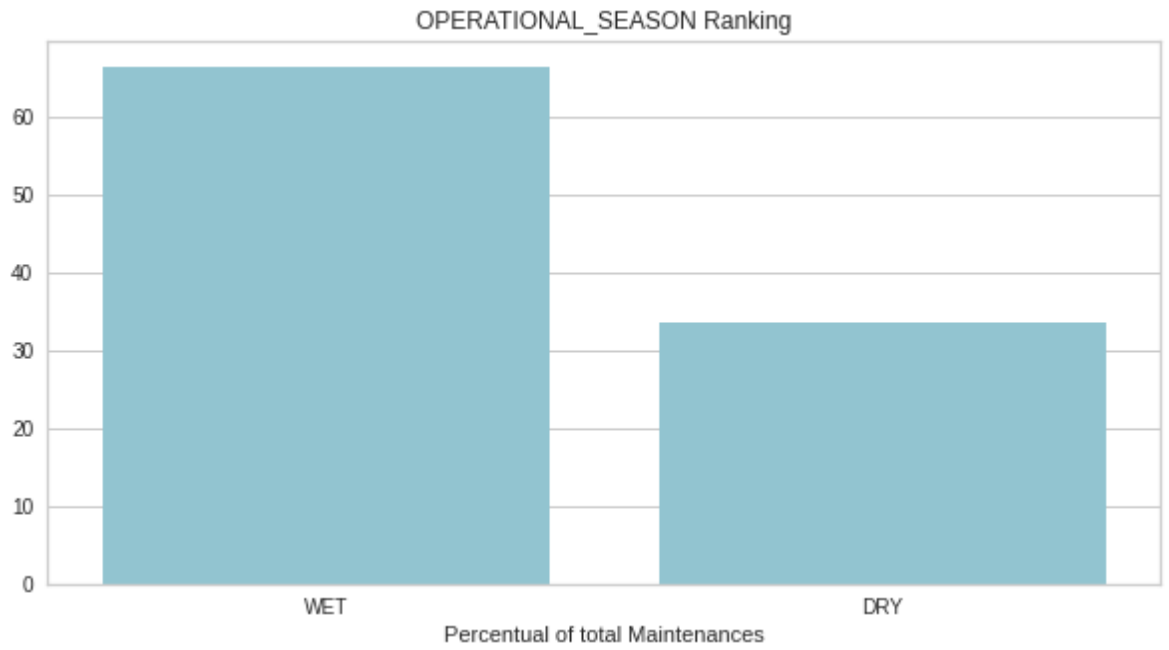

Figura 10 – Gráfico de barra ordenado por estação operacional

4.1.7 QUAL É O TEMPO MÉDIO PARA FALHA POR MÁQUINA E GRUPO DE OPERAÇÃO?

Para o cálculo do tempo médio para falha um novo campo foi calculado, de maneira muito similar ao calculado no item 4.1.3, no qual o tempo médio para falha foi dado pela diferença entre o horímetro registrado na manutenção seguinte e na manutenção atual, dado em horas. Uma vez que os resultados são similares ao item 4.1.3, com diferença da posição de linha no conjunto de dados, a variável de tempo médio para falha constitui-se em nossa variável dependente no futuro modelo preditivo.

4.1.8 QUAL É A IDADE MÉDIA DOS EQUIPAMENTOS POR GRUPO DE OPERAÇÃO?

A idade média dos equipamentos é de 65 meses, resultado que indica que a maior parte dos equipamentos estão próximos ou além do seu ciclo de vida estimado pelos fabricantes (em média 60 meses).

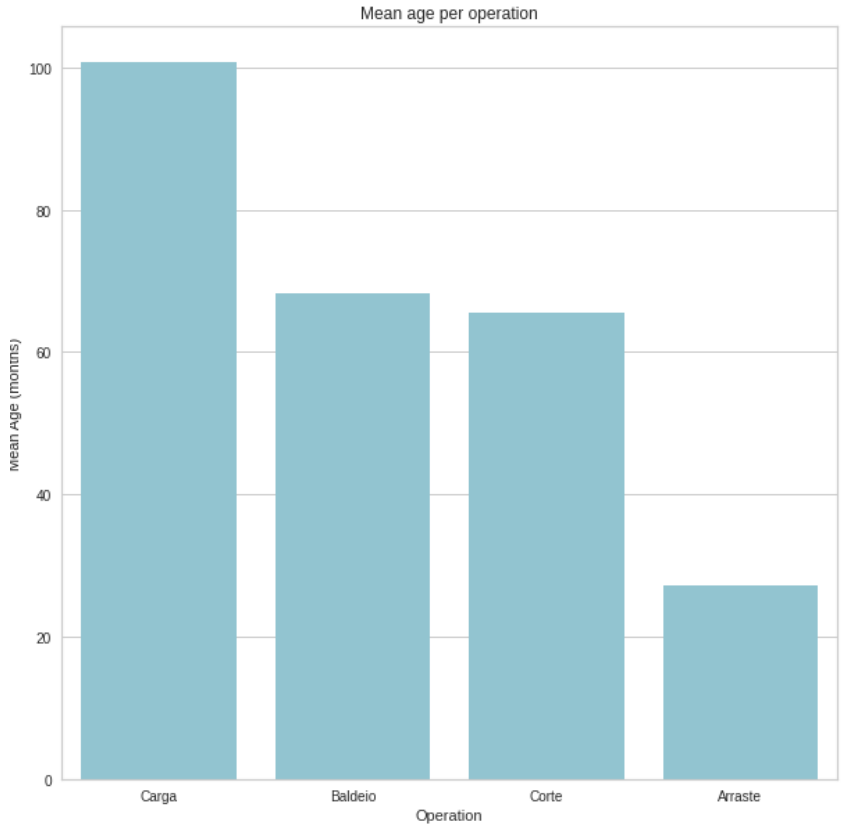

Figura 11 – Gráfico de barra ordenado por grupo de operação para a idade média em meses

Quando comparamos os resultados por grupo de operação (Figura 11), a operação de arraste tem os equipamentos mais novos (27 meses), enquanto os equipamentos de carga são os mais velhos (100 meses). Devido a essa grande dispersão dentro da idade média, utilizamos essa variável como independente na construção do modelo preditivo.

4.2 MODELAGEM

4.2.1 ESTATÍSTICA DESCRITIVA

Por meio da análise descritiva apresentada na Figura 12, observou-se que, após todos os ajustes e definições realizados até aqui, temos o total de 534

registros de manutenções disponíveis para a análise de modelos de preditivos. A média de horímetro é de 8473 horas, com desvio padrão de 9211 horas, ou seja, os dados de horímetro são bastante dispersos, fato esse que corrobora o disposto no item 4.1.8, no qual evidenciamos que a média de idade dos equipamentos é bastante dispersa.

	MAINTENANCE_ID	MACHINERY_HOURMETER	AGE	TIME_SINCE_LAST_REPAIR	TIME_TO_NEXT_REPAIR
count	534.000000	534.000000	534.000000	534.000000	534.000000
mean	81032.046816	8473.466292	63.172285	72.750936	72.750936
std	1376.985152	9211.408321	54.113704	128.150195	128.150195
min	78950.000000	19.000000	1.000000	0.000000	0.000000
25%	79567.250000	1427.000000	18.000000	0.000000	0.000000
50%	81205.500000	3077.000000	35.000000	26.000000	26.000000
75%	82409.750000	13206.750000	95.000000	83.500000	83.500000
max	83155.000000	31565.000000	181.000000	997.000000	997.000000

Figura 12 – Estatística descritiva das variáveis utilizadas na análise de modelos

4.2.2 DISPERSÃO E HISTOGRAMAS

Foram realizados gráficos de dispersão e histogramas, de modo que foi possível evidenciar que os dados não apresentam uma distribuição normal para o horímetro e idade. O tempo médio entre manutenções e o tempo médio para falha são enviesados à direita (Figura 13).

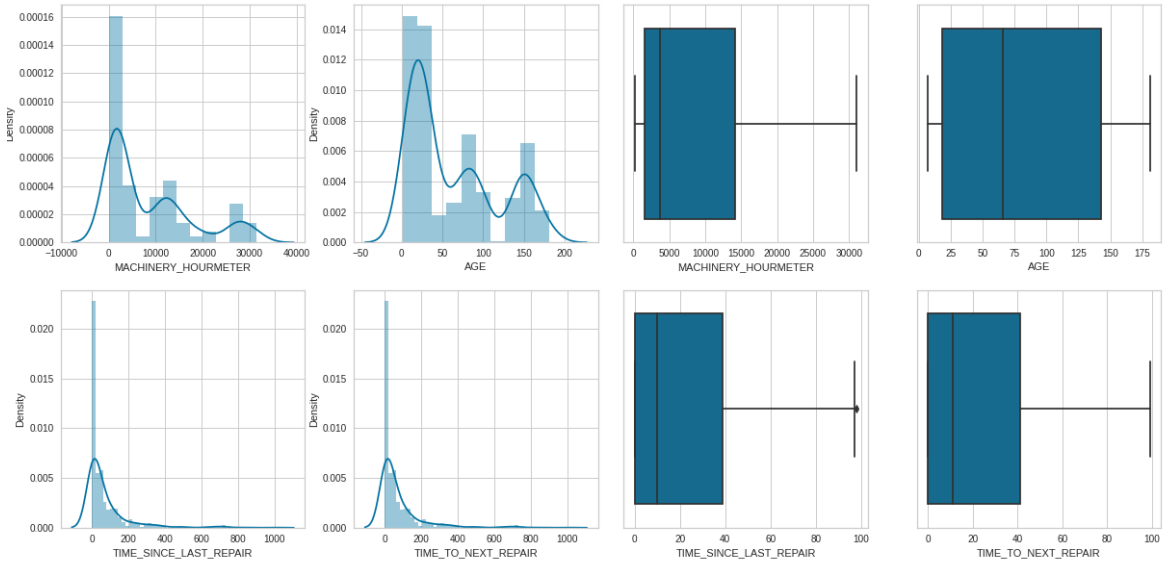

Figura 13 – Histograma e “Box-Plot” das variáveis utilizadas na análise do modelo.

Após a remoção de outliers, podemos ver através de gráficos de dispersão do tipo “Box-Plot” (Figura 13), como as variáveis estão distribuídas dentro dos quartis e suas medianas. É possível evidenciar que, para o horímetro, a mediana está abaixo das 5.000 horas, enquanto o valor máximo supera as 30.000 horas, corroborando os resultados apresentados no item 4.2.1. O mesmo efeito é visível para as variáveis de tempo médio entre manutenções e tempo médio para falha, em que a mediana encontra-se abaixo de 20 horas e os valores máximos se aproximam das 100 horas.

4.2.3 CORRELAÇÃO

Foi possível evidenciar uma forte correlação positiva entre as variáveis horímetro e idade dos equipamentos (Figura 14), resultado que se apresenta bastante coerente, uma vez que, quanto mais velho o equipamento, provavelmente mais horas o mesmo trabalhou. As demais variáveis apresentam correlação negativa, próximas a zero, resultado que pode indicar que a idade do equipamento, de forma isolada, não é um bom preditor para o tempo médio para a próxima falha, uma vez que, possivelmente, existem diversos outros fatores que não estão considerados no conjunto de dados disponível.

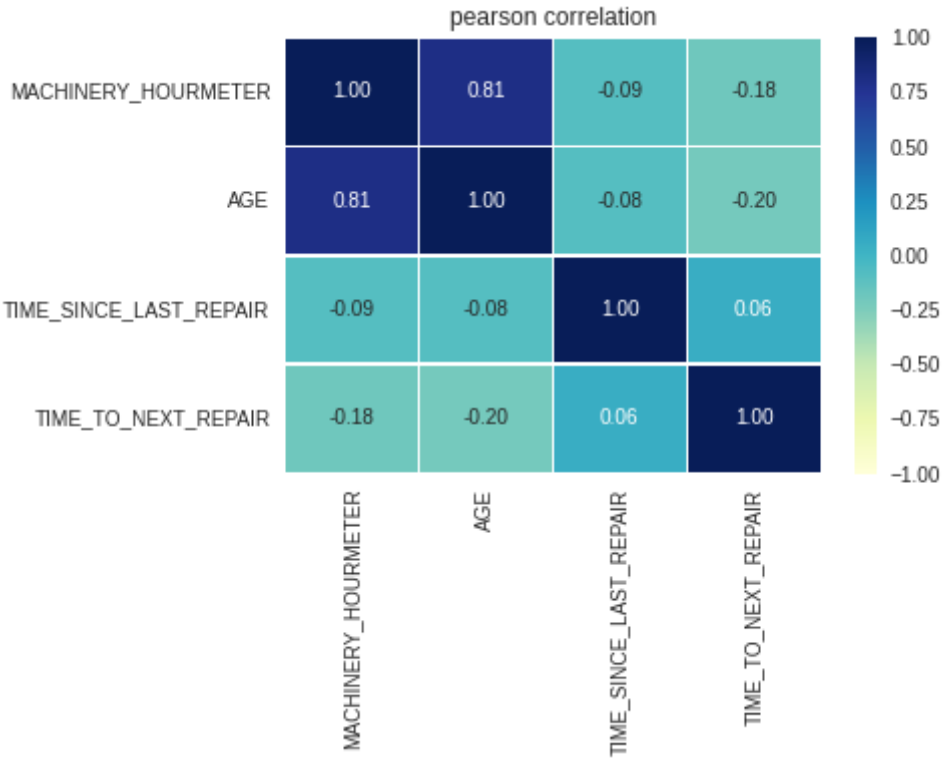

Figura 14 – Matriz de correlação entre as variáveis utilizadas na análise de modelos

4.2.4. CONFIGURAÇÕES DA ANÁLISE DE MODELOS

Utilizando a biblioteca Python PyCaret, definiu-se a variável tempo médio para falha como variável dependente, e as variáveis horímetro, idade e tempo médio entre manutenções como variáveis independentes; a biblioteca foi configurada para realizar a normalização dos dados. Na Figura 15, apresentam-se as configurações gerais da biblioteca, onde é possível evidenciar que não existem dados faltantes, três variáveis numéricas, três variáveis categóricas, foi definido a proporção de 70%/30% entre conjunto de dados entre treinamento e testes. O número de interações padrão é de 10 e a normalização dos dados se dá por Z-Score.

	Description	Value		Description	Value		Description	Value		Description	Value
0	session_id	7412	16	Use GPU	False	32	PCA Method	None	48	Polynomial Threshold	None
1	Target	TIME_TO_NEXT_REPAIR	17	Log Experiment	False	33	PCA Components	None	49	Group Features	False
2	Original Data	(185, 8)	18	Experiment Name	reg-default-name	34	Ignore Low Variance	False	50	Feature Selection	False
3	Missing Values	False	19	USI	cea5	35	Combine Rare Levels	False	51	Feature Selection Method	classic
4	Numeric Features	3	20	Imputation Type	simple	36	Rare Level Threshold	None	52	Features Selection Threshold	None
5	Categorical Features	3	21	Iterative Imputation Iteration	None	37	Numeric Binning	False	53	Feature Interaction	False
6	Ordinal Features	False	22	Numeric Imputer	mean	38	Remove Outliers	False	54	Feature Ratio	False
7	High Cardinality Features	False	23	Iterative Imputation Numeric Model	None	39	Outliers Threshold	None	55	Interaction Threshold	None
8	High Cardinality Method	None	24	Categorical Imputer	constant	40	Remove Multicollinearity	False	56	Transform Target	False
9	Transformed Train Set	(129, 9)	25	Iterative Imputation Categorical Model	None	41	Multicollinearity Threshold	None	57	Transform Target Method	box-cox
10	Transformed Test Set	(56, 9)	26	Unknown Categoricals Handling	least_frequent	42	Remove Perfect Collinearity	True			
11	Shuffle Train-Test	True	27	Normalize	True	43	Clustering	False			
12	Stratify Train-Test	False	28	Normalize Method	zscore	44	Clustering Iteration	None			
13	Fold Generator	KFold	29	Transformation	False	45	Polynomial Features	False			
14	Fold Number	10	30	Transformation Method	None	46	Polynomial Degree	None			
15	CPU Jobs	-1	31	PCA	False	47	Trignometry Features	False			

Figura 15 – Configuração da análise de modelos utilizando PyCaret

4.2.5 RESULTADO DOS MODELOS ANALISADOS

Entre os cinco modelos analisados, o que apresentou o melhor resultado, considerando as métricas de sucesso definidas anteriormente, foi o modelo de regressão de floresta aleatória (Figura 16).

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
knn	K Neighbors Regressor	101.1399	21846.2540	140.4576	-0.4541	1.5796	7.1654	0.080
lr	Linear Regression	98.9857	19155.2832	131.6107	-0.5063	1.6275	8.8708	0.025
rf	Random Forest Regressor	97.8355	22723.7639	144.5608	-0.8065	1.5811	9.8049	0.567
gbr	Gradient Boosting Regressor	101.1751	27753.8204	161.8613	-1.4531	1.5514	9.6592	0.068
dt	Decision Tree Regressor	128.0263	39296.8596	193.6178	-3.0360	1.9178	9.5279	0.040

Figura 16 – Resultado dos modelos analisados

O modelo de floresta aleatória obteve erro absoluto médio de 97.83 horas e o erro quadrático médio de 22723,76, mesmo sendo o melhor resultado entre os modelos analisados. Fica evidente que o resultado é fraco para o objetivo de predizer o tempo médio para falha, tanto no modelo de floresta aleatória, quanto em todos os outros aqui analisados, haja vista que um erro médio de 97 horas, entre estimado e real, não configura um erro aceitável para os padrões do negócio. O coeficiente de correlação, neste trabalho, não definido como métrica de sucesso, mostra resultados negativos, indicando que os modelos não conseguem explicar de maneira satisfatória a variável dependente de acordo com as variáveis independentes disponíveis.

4.2.6 OTIMIZAÇÃO DO MELHOR MODELO ANALISADO

Com o objetivo de melhorar os resultados obtidos no tópico 4.2.5, realizou-se a separação do melhor modelo obtido, floresta aleatória, sendo otimizados seus hyper parâmetros com o auxílio da biblioteca PyCaret, buscando o melhor erro absoluto médio possível; os resultados são apresentados na Figura 17:

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	48.2692	3718.3654	60.9784	0.3863	0.9348	1.4283
1	76.7308	16414.0577	128.1174	0.5091	0.8071	0.9589
2	75.0000	10684.8462	103.3675	-1.0979	1.9772	13.9900
3	93.3846	35327.0769	187.9550	-0.1042	1.5197	3.7373
4	102.1923	30258.2885	173.9491	-0.0146	1.2294	1.7512
5	104.3077	25895.0769	160.9195	-0.5104	1.5384	4.1485
6	42.9231	3260.3077	57.0991	0.5896	1.2868	3.5919
7	70.6154	11624.6538	107.8177	-3.9355	1.8414	27.3811
8	120.3846	25433.7692	159.4797	-0.8188	1.7739	6.7616
9	142.4583	53227.8542	230.7116	-0.4766	1.6231	1.1963
Mean	87.6266	21584.4296	137.0395	-0.5473	1.4532	6.4945
SD	29.5265	14827.3467	52.9585	1.2499	0.3650	7.8798

Figura 17 – Resultado do modelo de floresta aleatória otimizado

A otimização melhorou o erro absoluto médio, que agora é de 87.62 horas, uma diferença de aproximadamente 10 horas em comparação ao modelo não otimizado. Entretanto, esse erro é inaceitável do ponto de vista do negócio, em que

qualquer tomada de decisão baseada nesses resultados poderá ocasionar em erros operacionais.

4.2.7 NÍVEL DE IMPORTÂNCIA DAS VARIÁVEIS INDEPENDENTES

Com base nos dados disponíveis e nos modelos analisados, o horímetro da máquina é a principal variável de importância na predição do tempo médio para falha, seguido pela idade do equipamento em meses e do tempo desde o último reparo. As demais variáveis não apresentam importância estatística (Figura 18).

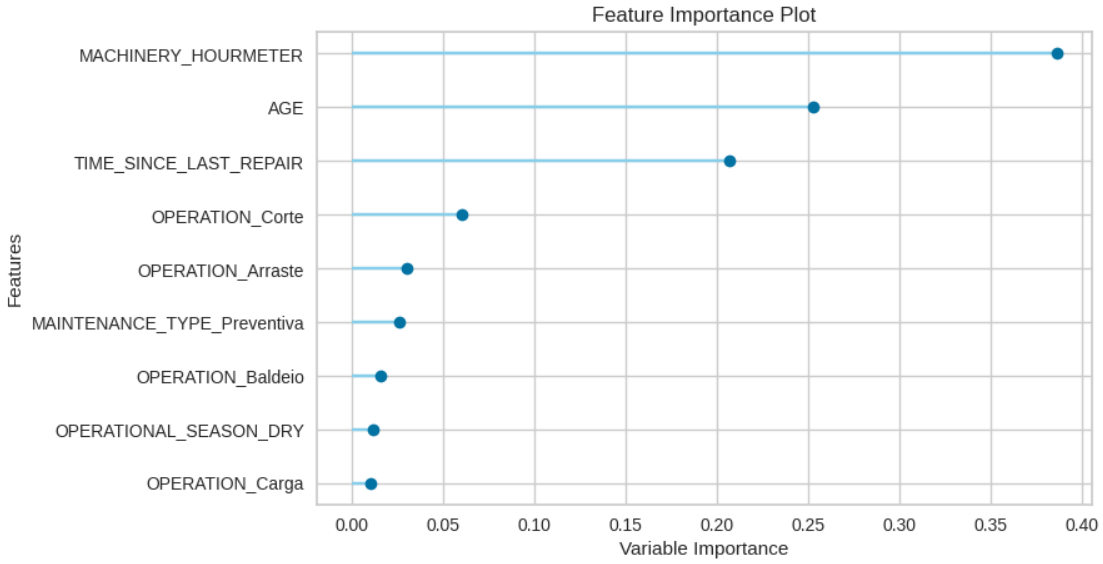

Figura 18 – Resultado de importância das variáveis utilizadas no modelo

Esse resultado corrobora com os resultados obtidos no item 4.2.5, que identificam que a utilização de modelos de aprendizado de máquina aqui avaliados não se mostrou satisfatória para predizer o tempo médio para falha. Uma vez que na etapa de análise exploratória dos dados, notou-se que existe diferença nos dados entre o tipo de operação e a estação operacional, entretanto, tal diferença não foi expressa no modelo de melhor resultado.

5 CONCLUSÕES

A análise realizada neste estudo foi bastante interessante de ser trabalhada. Dentre os resultados obtidos, constatou-se que a atividade de baldeio representa mais da metade do total de registros de manutenção, sendo maior que o dobro da segunda operação, que é a de arraste. O tempo médio para reparo é de 43,96 horas, sendo um número considerado alto, pois algumas manutenções no conjunto de dados foram mais longas que a média (em alguns casos superior a 300 horas), devido a falhas complexas ou dificuldades de aquisição de peças de reposição.

O conjunto dos dados tem sérios problemas em relação aos apontamentos de horímetro. Foram necessárias 262 interações de correção com auxílio do time de negócios da companhia para ajuste dos números. Isso, além de oneroso em tempo, pode ser considerado um sinal de desconfiança em relação a qualidade dos dados disponibilizados. De qualquer maneira, após as correções, foi possível evidenciar um tempo médio para reparos de aproximadamente 19 horas, dentre os quais a atividade de corte apresenta o menor resultado (13 horas), indicando que essa operação apresenta mais necessidade de manutenção que as demais. Isso corrobora o resultado obtido no item 4.1.1, no qual foi possível evidenciar que o equipamento HV 08 é aquele com maior número de manutenções registradas.

Pelos resultados obtidos, evidenciou-se que 68,4% do total de manutenções foram corretivas, e 29,6%, preventivas, indicando que, na maioria das vezes, as manutenções acontecem após a falha do equipamento, incorrendo em perdas de produção por conta de equipamento parado de forma não programada. As peças Terminal Hidráulico e Mangueira Hidráulica são as mais frequentes nas manutenções, somando 40% do total quando combinadas. Sendo assim, decidiu-se, neste trabalho, focar a modelagem nas manutenções que utilizaram esse tipo de peça, por constituírem o maior problema da empresa em relação a falhas e manutenção mecânica.

Mais de 65% das manutenções, no conjunto dos dados, aconteceram na temporada chuvosa. Entretanto, o resultado aqui obtido, pode conter algum viés em função da pandemia de COVID, quando algumas operações foram paralisadas temporariamente. A idade média dos equipamentos é de 65 meses, indicando que a maior parte dos equipamentos está no final do seu ciclo de vida (geralmente 60

meses); a atividade de arraste tem os equipamentos mais novos, e a de carga, os mais antigos.

As métricas utilizadas para avaliar o sucesso dos modelos analisados foram o erro médio absoluto e o erro quadrático, e o modelo com o melhor resultado foi o de floresta aleatória, com erro médio absoluto após otimização de 87,62 horas e erro quadrático de 21584,43. Tais resultados mostram que a aplicação de modelos de aprendizado de máquina para predição do tempo médio para falha no conjunto de dados analisados não tem resultados satisfatórios, não podendo ser utilizada em escala de produção para auxílio na tomada de decisão da companhia.

Como recomendações baseadas nos resultados obtidos, sugere-se que o negócio invista em aumento da governança de dados, objetivando maior qualidade nos dados obtidos, uma vez que esses são a chave para o sucesso de qualquer análise exploratória ou modelagem preditiva. Os resultados obtidos no presente trabalho deixaram claro que os dados atuais apresentam falhas de governança, que podem ter impactado o resultado final. Recomenda-se, também, a continuidade do trabalho, por meio do incremento da engenharia de variáveis, com o objetivo de entender melhor a variabilidade dos dados. É aconselhável investigar a estratificação dos modelos e gerar resultados baseados em grupos de operação ou idade dos equipamentos, o que pode apresentar erros menores que os aqui obtidos.

6 REFERÊNCIAS

ALI, M. PyCaret: An open source, low-code machine learning library in Python. **PyCartet version 1.0**, 2020. Disponível em: https://www.pycaret.org. Acesso em: 27 abr. 2022.

ANZANELLO, M. J.; SILVA, P. R. S. da; RIBEIRO, J. L. D.; FOGLIATTO, F. S. Proposição de modelo de degradação para capacitores submetidos a ensaios acelerados. *In*: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 23., 2003, Ouro Preto. **Anais** […]. Ouro Preto: ABEPRO, 2003.

EZRA, O. Achieving Manufacturing Excellence with Predictive Maintenance and Machine Learning. **Industry 4.0 Insights**, [*s. l.*], 2018.

FERNÁNDEZ-DELGADO, M. *et al*. Do we need hundreds of classifiers to solve real world classification problems. **Journal of Machine Learning Research**, [*s. l.*], v. 15, n. (1), p. 3133-3181, 2014.

FOGLIATTO, F. S.; RIBEIRO, J. L. D. **Confiabilidade e manutenção industrial**. Rio de Janeiro: Elsevier, 2009. 265 p.

KARDEC, A.; NASCIF, J. de A. **Manutenção**: Função estratégica. Rio de Janeiro: Qualitymark, 2009.

KUHLMAN, D. Introductions Etc. *In*: KUHLMAN, D. **A Python Book**: Beginning Python, Advanced Python, and Python Exercises. [*S. l.*: *s. n.*], 2013.

LAFRAIA, J. R. B. **Manual de confiabilidade, mantenabilidade e disponibilidade**. 2. ed. Rio de Janeiro: Qualitymark, 2001.

LUCIAN, L. M. **Fundamentos de Aprendizagem de Máquina**. Porto Alegre: SAGAH, 2020.

MITCHELL, T. M. **Machine Learning**. New York: McGraw-Hill, 1997.

MONARD, M. C.; BARANAUSKAS, J. A. **Conceitos Sobre Aprendizado de Máquina. Sistemas Inteligentes Fundamentos e Aplicações**. Barueri, SP: Manole Ltda, 2003. p. 89-114.

NATEKIN, A.; KNOLL, A. **Gradient boosting machines, a tutorial**. Munich, Alemanha: Department of Informatics, Technical University Munich, Garching, 2013.

PACCOLA, J. E. **Manutenção e Operação de Equipamentos Móveis**. São José dos Campos: JAC, 2017.

PEDREGOSA, F. *et al*. **Journal of Machine Learning Research**, v. 12, p. 2825-2830, 2011.

PERES, S. M. *et al*. Tutorial sobre fuzzy-c-means e fuzzy learning vector quantizations: Abordagens híbridas para tarefas de agrupamento e classificação. **Revista de Informática Teórica e Aplicada**, v. 19, n. 1, p. 120-163, 2012.

PIOTROWSKI, P. Build a Rapid Web Development Environment for Python Server Pages and Oracle. **Oracle Technology Network**, 2006. Retirado do original em 2 abr. 2019. Acessado novamente em 12 mar. 2012.

PYTHON SOFTWARE FOUNDATION. Is Python a good language for beginning programmers? **General Python FAQ**. Retirado do original em 24 out. 2012. Acessado novamente em jan. 2022.

RICHTER, I. Achieving Zero Unplanned Downtime with Predictive Maintenance Analytics. **Industry 4.0 Insights**, [*s. l.*], 2019.

RIQUETI, G. A.; RIBEIRO, C. E.; ZÁRATE, L. E. **Classificando perfis de longevidade de bases de dados longitudinais usando Floresta Aleatória**. [*S. l.*]: Symposium on Knowledge Discovery, Mining and Learning, KDMILE, 2018.

ROSSUM, G. van. **The History of Python**: A Brief Timeline of Python. [*S. l.*: *s. n.*], 2009. Retirado do original em 2022.

RODRIGUES, S. C. **Modelo de Regressão Linear e suas Aplicações**. Covilhã: [*s. n.*], 2012.

WUTTKE, R. A.; SELLITO, M. A. Cálculo da disponibilidade e da posição na curva da banheira de uma válvula de processo petroquímico. **Revista Produção Online**, v. 8, n. 4, p. 1-23, 2008.