

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

CARLOS EDUARDO BARBOSA MIRANDA

**APLICAÇÃO DA REGRESSÃO LOGÍSTICA BINÁRIA PARA MANUTENÇÃO
PREDITIVA EM MÁQUINAS DE RESSONÂNCIA MAGNÉTICA**

TOLEDO

2023

CARLOS EDUARDO BARBOSA MIRANDA

**APLICAÇÃO DA REGRESSÃO LOGÍSTICA BINÁRIA PARA MANUTENÇÃO
PREDITIVA EM MÁQUINAS DE RESSONÂNCIA MAGNÉTICA**

**Application of Binary Logistic Regression for Predictive Maintenance in
Magnetic Resonance Machines**

Trabalho de Conclusão de Curso de Graduação apresentado como requisito para obtenção do título de Bacharel em Engenharia Eletrônica do Curso de Bacharelado em Engenharia Eletrônica da Universidade Tecnológica Federal do Paraná.

Orientador: Prof. Dr. Marcos Roberto Bombacini

TOLEDO

2023



[4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Esta licença permite remixe, adaptação e criação a partir do trabalho, para fins não comerciais, desde que sejam atribuídos créditos ao(s) autor(es) e que licenciem as novas criações sob termos idênticos. Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

CARLOS EDUARDO BARBOSA MIRANDA

**APLICAÇÃO DA REGRESSÃO LOGÍSTICA BINÁRIA PARA MANUTENÇÃO
PREDITIVA EM MÁQUINAS DE RESSONÂNCIA MAGNÉTICA**

Trabalho de Conclusão de Curso de Graduação apresentado como requisito para obtenção do título de Bacharel em Engenharia Eletrônica do Curso de Bacharelado em Engenharia Eletrônica da Universidade Tecnológica Federal do Paraná.

Data de aprovação: 08/novembro/2023

Marcos Roberto Bombacini
Doutorado
Universidade Tecnológica Federal do Paraná

Evandro Andre Konopatzki
Doutorado
Universidade Tecnológica Federal do Paraná

Mauri Wagner de Sousa
Bacharelado
Universidade Estadual de Maringá

**TOLEDO
2023**

Dedico a Deus, pois acredito que ele tenha uma participação importante. À minha família, pois contribuíram para a realização deste trabalho. Aos meus amigos, poucos em números, mas incomensuráveis na qualidade. Ao orientador, pela sabedoria na orientação e por sua amizade dispensada.

AGRADECIMENTOS

Gostaria de agradecer a todas as pessoas que me ajudaram a concluir este trabalho de conclusão de curso. Primeiramente, gostaria de gratificar meu orientador Marcos Roberto Bombacini e os outros professores que fizeram parte de toda a graduação, que me forneceram orientações e ensinamentos valiosos além de críticas construtivas para aprimorar meu trabalho e me tornar um profissional melhor.

Também quero expressar minha gratidão aos meus pais e amigos, que me apoiaram ao longo deste processo desafiador, e me deram força e motivação para continuar até a conclusão do curso.

Finalmente, agradeço a todas as pessoas que participaram deste estudo e colaboraram para que eu pudesse coletar os dados necessários para a realização deste trabalho. Seu tempo e esforço são muito apreciados e foram fundamentais para a conclusão deste trabalho. Muito obrigado a todos que contribuíram para o sucesso deste projeto.

RESUMO

Este estudo foca na aplicação da regressão logística como uma técnica de *machine learning* para aprimorar a manutenção de máquinas de Ressonância Magnética General Electric (GE) 1.5 T. Dado o contexto de confiabilidade da máquina e continuidade associados as falhas dos equipamentos, a necessidade de otimizar a manutenção e minimizar interrupções inesperadas é crucial para garantir a eficiência operacional e a satisfação do cliente. Os objetivos do trabalho foram identificar as ferramentas e referências teóricas relacionadas à regressão logística binária e seus métodos. Realizar a coleta dos dados de sensores de temperatura, pressão e fluxo por meio de um *datalogger* instalado na máquina. Validar a capacidade do modelo de regressão logística no banco de dados utilizado. Para atingir esses objetivos, além da coleta de dados realizada, utilizou-se o Microsoft Excel para a construção do banco de dados, o *software* Orange para a aplicação do modelo de regressão. Para validação do modelo, a análise de métricas como a curva ROC e a matriz de confusão, comparando também com outros métodos de predição. Os resultados demonstraram a eficácia da regressão logística na previsão de possíveis falhas, evidenciando seu potencial como ferramenta de suporte à manutenção preditiva. A análise bem-sucedida dos dados de teste confirmou a capacidade do modelo de identificar o *status* operacional da máquina de Ressonância Magnética, contribuindo assim para a otimização da gestão de manutenção. A capacidade preditiva dessa abordagem fornece suporte para a tomada de decisões baseadas em dados e contribuem para a eficiência e confiabilidade dos serviços de Ressonância Magnética.

Palavras-chave: regressão logística binária; manutenção preditiva; custos de manutenção; ressonância magnética; cálculos probabilísticos.

ABSTRACT

This study focuses on the application of logistic regression as a machine learning technique to improve the maintenance of General Electric GE 1.5 T Magnetic Resonance Imaging machines. Given the context of machine reliability and continuity associated with equipment failures, the need to optimize maintenance and minimize unexpected interruptions is crucial to ensure operational efficiency and customer satisfaction. The objectives of the study were to identify tools and theoretical references related to binary logistic regression and its methods, to collect data from temperature, pressure, and flow sensors through a datalogger installed on the machine, and to validate the capacity of the logistic regression model in the utilized database. To achieve these objectives, in addition to the data collection, Microsoft Excel was used for constructing the database, and the software Orange was utilized for the application of the regression model. For model validation, metric analysis such as the ROC curve and the confusion matrix were employed, also comparing with other prediction methods. The results demonstrated the efficacy of logistic regression in predicting potential failures, highlighting its potential as a tool for supporting predictive maintenance. The successful analysis of the test data confirmed the model's ability to identify the operational status of the Magnetic Resonance Imaging machine, thus contributing to the optimization of maintenance management. The predictive capacity of this approach provides support for data-driven decision-making and contributes to the efficiency and reliability of Magnetic Resonance Imaging services.

Keywords: binary logistic regression; preventive maintenance; cost reduction ; magnetic resonance imaging; probabilistic calculations.

LISTA DE FIGURAS

Figura 1 – Sistema de Ressonância Magnética e componentes	13
Figura 2 – Exemplo de bobina Rádio frequência (RF) de joelho	14
Figura 3 – Sistema gradiente da Magnetic Resonance (RM) GE LX 1.5 Tesla (T)	14
Figura 4 – criogenia RM	15
Figura 5 – Compressor de hélio e <i>chiller</i> da RM GE LX 1.5 T	15
Figura 6 – Sistema de Ressonância Magnética General Eletric LX 1.5 T	16
Figura 7 – Exemplo de curva de <i>logit</i> x probabilidade	22
Figura 8 – Gráfico custo em relação ao nível de manutenção.	25
Figura 9 – Ponto de ótimo disponibilidade.	25
Figura 10 – Fluxograma dos passos realizados	29
Figura 11 – <i>Cold Head</i>	30
Figura 12 – Sistema <i>datalogger Magnet Monitor</i> da RM GE LX 1.5 T	31
Figura 13 – Sistema completo no <i>software</i> Orange	33
Figura 14 – Configuração Random Forest	35
Figura 15 – Configuração <i>Neural Networks</i>	36
Figura 16 – Exemplo de dados coletados com aspecto redundante da RM GE	37
Figura 17 – Definição das categorias das variáveis	38
Figura 18 – Estatísticas relacionadas as variáveis	39
Figura 19 – Correlação de Pearson	40
Figura 20 – Correlação de Spearman	41
Figura 21 – Parâmetros regressão logística	42
Figura 22 – <i>Widget</i> de Predição	44
Figura 23 – Matriz de confusão do modelo de regressão logística	49
Figura 24 – Curva Receiver Operator Characteristic Curve (ROC) gerada pelo modelo de regressão logística.	50
Figura 25 – <i>logs</i> de quebra do sistema de criogenia	51
Figura 26 – <i>Software</i> acertando as falhas	51
Figura 27 – Aplicação dos modelos nos dados.	52
Figura 28 – Curva ROC dos 3 modelos	53
Figura 29 – Planilha teste.	55

Figura 30 – Resultado da predição.	56
---	-----------

LISTA DE TABELAS

Tabela 1 – Dados coletados do <i>datalogger</i> e sensores externos do sistema de ressonância magnética GE LX 1.5 T	31
Tabela 2 – <i>Test and Score</i> regressão logística	48
Tabela 3 – <i>Test and Score</i> regressão logística	48
Tabela 4 – <i>Test and Score</i> dos métodos de <i>machine learning</i>	52
Tabela 5 – Tempo(segundos) do treinamento e teste dos modelos de <i>machine learning</i>	53
Tabela 6 – Coeficientes z regressão logística	54

LISTA DE ABREVIATURAS E SIGLAS

Siglas

AUC	Área sob a Curva ROC
CSV	<i>Comma-separated values</i>
GE	General Eletric
K	Kelvin
Lasso	Least Absolute Shrinkage and Selection Operator
NBR	Normas Brasileiras
RAM	Random Access Memory
RF	Rádio frequência
Ridge	Ridge Regression
RM	Magnetic Resonance
RMI	Magnetic Resonance Imaging
ROC	Receiver Operator Characteristic Curve
SGD	<i>Stochastic Gradient Descent</i>
T	Tesla
°C	Grau Celsius

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Contextualização	16
1.1.1	Definição do tema	16
1.1.2	Apresentação do problema	17
1.2	Objetivos	17
1.2.1	Objetivo geral	17
1.2.2	Objetivos específicos	17
1.3	Justificativa	18
1.4	Estrutura do trabalho	19
2	REFERENCIAL TEÓRICO	20
2.1	Ferramentas utilizadas	20
2.2	Regressão Logística	20
2.2.1	Função Logit	21
2.2.2	Regressão Logística Binária	23
2.2.3	Método de máxima verossimilhança	23
2.3	Tipos de Manutenção	24
2.3.1	Manutenção Corretiva	26
2.3.2	Manutenção Preventiva	26
2.3.3	Manutenção Preditiva	27
3	MATERIAIS E MÉTODOS	29
3.1	Materiais	30
3.2	Métodos	32
3.2.1	Implantação do sistema	33
3.2.2	Modelagem e implementação da regressão logística	36
4	RESULTADOS	45
4.1	Escopo do sistema	45
4.1.1	Funcionalidades	45
4.1.2	Limitações do sistema	47
4.2	Parametrização e Resultados	48
4.3	Resultados obtidos e comparação com outros modelos	52

5	CONCLUSÃO	57
	REFERÊNCIAS	58

1 INTRODUÇÃO

A Ressonância Magnética (RM) é uma técnica avançada de diagnóstico por imagem 2D-3D, diferenciando-se dos métodos convencionais, como raio-x e tomografia, que envolvem a exposição à radiação ionizante ou o uso de contraste iodado. Este exame não invasivo oferece uma notável capacidade de distinguir tecidos, com alta sensibilidade e especificidade na resolução de imagens. Sua aplicabilidade é ampla, abrangendo órgãos, artérias, nervos e tumores, possibilitando uma análise abrangente de diversas regiões do corpo humano.(SANTIAGO; SILVA, 2007; MAZZOLA, 2009; ARCURI; MCGUIRE, maio. 2001).

O contexto histórico dos primeiros estudos sobre ressonância magnética remonta a uma série de descobertas científicas e avanços tecnológicos ao longo de várias décadas. A base para a tecnologia desse utensílio começou com o princípio da ressonância magnética nuclear, descoberto por Felix Bloch e Edward Purcell em 1946, pelo qual foram posteriormente premiados com o Prêmio Nobel de Física em 1952. No entanto, a aplicação médica da RM só começou a se materializar nas décadas seguintes (RIBEIRO, 2018; ARMSTRONG; KEEVIL, jul. 1991; DAWSON, 2013).

A revolução da imagem por ressonância magnética iniciou-se nos anos 1970, com os trabalhos independentes de Paul Lauterbur e Sir Peter Mansfield, que foram fundamentais para o desenvolvimento das técnicas de imagem de ressonância magnética modernas. Em 1971, Lauterbur propôs o conceito de usar gradientes magnéticos para criar imagens, enquanto Mansfield, em 1973, introduziu a técnica de eco de spin, que permitiu a formação de imagens de ressonância magnética mais rápidas e precisas. Suas contribuições foram essenciais para a criação de imagens em duas e três dimensões, abrindo caminho para o uso clínico e diagnóstico da ressonância magnética (RIBEIRO, 2018).

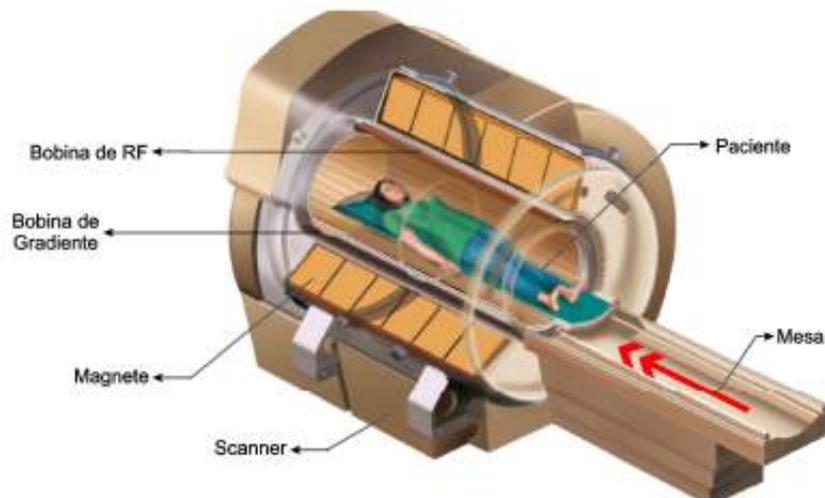
Desde então, a tecnologia de ressonância magnética continua a se aprimorar, impulsionada por avanços nas áreas de engenharia, física e ciência dos materiais. A pioneira Fonar produziu a primeira ressonância magnética humana e realizou a primeira utilização clínica comercial da ressonância magnética em 1980. Foi construída empregando um ímã permanente. Esses ímãs permanentes e os ímãs resistivos foram as tecnologias precursoras dos primeiros dispositivos comerciais de imagem por RM. A evolução dos gradientes e dos ímãs mais potentes, tem impulsionado significativamente a qualidade das imagens, incluindo os ímãs supercondutores, resultando em uma relação sinal-ruído superior e níveis de campo mais elevados. Com a evolução dos equipamentos é possível uma visualização detalhada de estruturas internas do corpo humano. Com o tempo, a ressonância magnética tornou-se uma ferramenta essencial no diagnóstico de uma ampla variedade de condições médicas, fornecendo aos médicos e pesquisadores informações vitais para o tratamento e a compreensão de doenças complexas (NATUKUNDA *et al.*, 2021; POSSES, 2023; SCHETTINI, 2016; BATHISTA, jul. 2005).

Devido à sua capacidade de produzir imagens de alta definição, as máquinas de ressonância magnética desempenham um papel indispensável no diagnóstico preciso de uma va-

riedade de doenças e lesões. Essa tecnologia não apenas aprimora a eficácia dos tratamentos médicos, mas também reduz significativamente erros diagnósticos, elevando a precisão em comparação com outros métodos de diagnóstico por imagem. (MAULAZ *et al.*, 2018; MAZZOLA, 2009; MINGUETTI, 2001).

O funcionamento das máquinas de Imagens Ressonância Magnética (Magnetic Resonance Imaging (RMI)) para gerar as imagens consiste em: trabalhar com as moléculas de água do corpo. A molécula de água é composta por um átomo de oxigênio e dois átomos de hidrogênio. O hidrogênio consiste em um próton e um elétron, esse próton é o núcleo da molécula e é ele que fornece o sinal de ressonância magnética, a partir de um campo magnético externo gerado pelo magneto conforme a Figura 1. Além disso, os spins da molécula se alinham com o campo e os prótons podem se alinhar paralelamente ou anti-paralelo conforme o estado do spin da molécula, para um maior alinhamento da molécula basta aumentar o campo magnético, no caso da máquina em estudo 1.5 Tesla (T). (RIBEIRO, 2018; HAGE; IWASAKI, 2009; MAZZOLA, 2009).

Figura 1 – Sistema de Ressonância Magnética e componentes



Fonte: (PINTO, 2017).

Após a excitação da molécula acontece o processo de relaxamento, onde a energia introduzida pelo campo magnético externo é emitida pelos prótons. Nisso, são gerados dois tipos relaxamentos: o longitudinal e o transversal. Esses relaxamentos geram o contraste da imagem, pois variam conforme o tecido e a intensidade do campo. (RIBEIRO, 2018; HAGE; IWASAKI, 2009; MAZZOLA, 2009).

Com intuito de mapear as imagens, a bobina de gradiente pode variar a intensidade, a direção e a linearidade do campo magnético. Desse modo, cada posição ao longo do eixo z, y ou x terá um valor diferente de campo, acionando um pulso de rádio frequência com valor igual à frequência de precessão dos prótons de hidrogênio existentes na região de mapeamento. Ademais, todos os prótons de hidrogênio da região receberão energia do pulso de Rádio Frequência (RF) e isso fará com que retorne sinal para a bobina de RF conforme da Figura 2.

Figura 2 – Exemplo de bobina RF de Joelho



Fonte: Autor.

A figura 3 mostra o sistema de gradientes da máquina RM GE LX 1.5 T, observa-se a fonte das bobinas e os 3 eixos (X, Y e Z), que variam a intensidade do campo magnético da máquina de ressonância.

Figura 3 – Sistema gradiente da RM GE LX 1.5 T

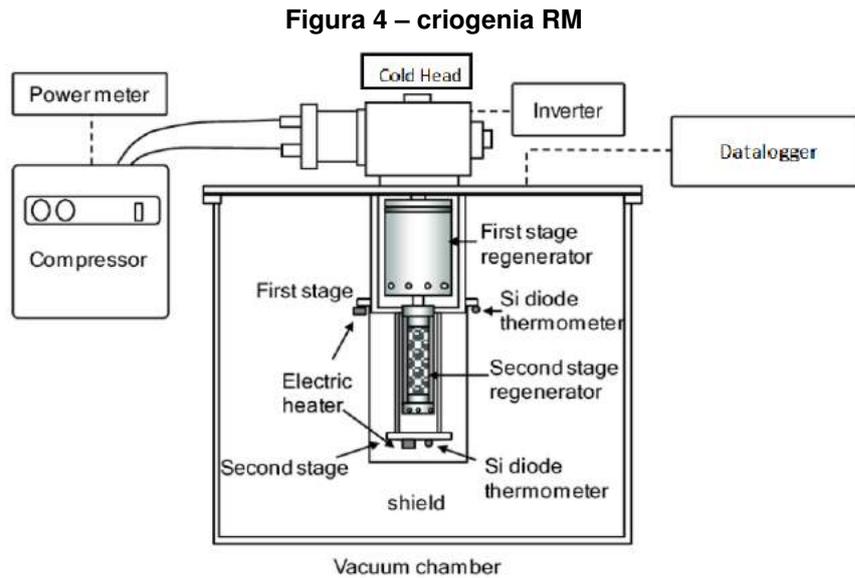


Fonte: Autor.

Para o funcionamento das etapas citadas, a criogenia da máquina em geral entra em ação, com o intuito de assegurar a estabilidade do núcleo do magneto. Por conta das suas

características supercondutoras, é necessário manter a temperatura em cerca de $-269,0^{\circ}\text{C}$. Isso permite que correntes elevadas fluam pelas bobinas, criando o campo magnético estático e intenso utilizado para gerar as imagens.

Os processos da criogenia e alguns dos sensores usados no estudo estão descritos na Figura 4, já a disposição do *Cold Head* está na Figura 6, localizada no canto superior esquerdo da máquina, na posição vertical como disposto na imagem 4.



Fonte: (MASUYAMA; NUMAZAWA, 2017), com alterações realizadas pelo autor.

A primeira etapa do processo de criogenia da RM é resfriar a água para fazer troca de calor no compressor de hélio através do *chiller*, enquanto o compressor comprime o hélio que envia para o *Cold Head* 11 em estado de hélio líquido, demonstrado pela Figura 4.

Figura 5 – Compressor de hélio e *chiller* da RM GE LX 1.5 T



Fonte: Autor.

O modelo da máquina em que foi realizada a análise de dados é uma Ressonância Magnética General Eletric LX 1.5 T, conforme a Figura 6. Existem 3 máquinas desse modelo,

as quais estão situadas na região de Toledo-PR e Cascavel-Pr, sendo acessadas com acompanhamento do engenheiro ou técnico responsável.

Figura 6 – Sistema de Ressonância Magnética General Eletric LX 1.5 T



Fonte: Autor.

1.1 Contextualização

Neste capítulo, serão abordados a problematização e a justificativa para a realização deste trabalho. Será discutido o problema a ser investigado, bem como as razões e motivações que levaram à escolha deste tema.

1.1.1 Definição do tema

A problemática desta pesquisa é obter um modelo preditivo a partir da análise de um banco de dados de sensores da máquina - dados esses recolhidos dos *datalogger* - utilizando modelo logístico binário através do *software* Orange Canvas, para prever a probabilidade de quebra e assim fornecer um modelo preditivo para auxiliar no cronograma de manutenções de uma máquina de ressonância magnética General Eletric GE LX 1.5 T em estudo.

O tema definido para o presente trabalho é: Análise de dados coletados através do *datalogger* da máquina de Ressonância Magnética RM e utilização do modelo logístico binário para definir a probabilidade de quebra do equipamento.

1.1.2 Apresentação do problema

Conforme (ALMEIDA, 2018), o uso das manutenções corretiva ou preventiva são dois extremos da área da manutenção e não definem com eficiência um modelo matemático para prever a quebra de uma máquina, como o uso da manutenção preditiva. O problema levantado é: como definir a probabilidade de falhas de uma RM utilizando o histórico de quebras anteriores e sensores por meio da regressão logística binária?.

1.2 Objetivos

Com base nos parágrafos precedentes, evidenciando a relevância do tema e os problemas levantados, o estudo teve como objetivo uma aplicação que possa contribuir com os métodos para a manutenção preditiva das máquinas de ressonância magnética do modelo GE LX 1.5 T, visando fornecer contribuições para o aprimoramento desses métodos.

1.2.1 Objetivo geral

Por meio de um banco de dados do histórico de quebras e sensores implantados nas máquinas, gerar um modelo preditivo capaz de prever a probabilidade de quebra das RM, com foco na criogenia, auxiliar no cronograma manutenção preditiva para as máquinas de RM, com finalidade de reduzir paradas indesejadas.

1.2.2 Objetivos específicos

Os objetivos específicos têm o intuito de descrever metas particulares alcançadas durante o estudo. Os objetivos descritos são formulados de forma precisa e coerente com a área de estudo, abrangendo ações ou resultados específicos obtidos.

- a) Coletar os dados de sensores de temperatura, pressão e fluxo através do *datalogger*, juntamente com histórico de quebras da RM GE LX 1.5 T.
- b) Tratar os erros, formação de banco de dados e amostra.
- c) Escolher um modelo matemático usando amostra.
- d) Realizar a validação do modelo com banco de dados completo, para prever a probabilidade de quebra.

1.3 Justificativa

Para maximizar o retorno econômico das empresas, é importante reduzir custos e aumentar o retorno de capital. Uma área com grande impacto para o custo das organizações é a manutenção, que dependendo do setor, pode ser responsável por 15% a 40% do custo do produto (MOBLEY, 2002).

Em um cenário competitivo, pequenos ganhos e perdas podem fazer toda a diferença, pois cada vantagem conquistada ou prejuízo sofrido pode afetar diretamente a posição de uma empresa em relação aos seus concorrentes. Nesse contexto, a manutenção adequada e eficiente desempenha um papel crucial para que os equipamentos, máquinas e processos estejam funcionando em seu melhor desempenho. Assim, a empresa pode evitar falhas, interrupções inesperadas e atrasos na produção. Isso significa que ela estará em condições de oferecer produtos ou serviços de qualidade superior, cumprir prazos com maior confiabilidade e manter a satisfação dos clientes em alta (NEPOMUCENO, 2014; ALMEIDA; ROCCA, 2015; SILVA *et al.*, 2020).

Conforme o autor (BORLIDO, 2017), a indústria 4.0 tem a habilidade de registrar e processar informações em tempo real, essa vantagem é resultado da fusão de tecnologias, da interação entre sistemas *cyber-físicos*, a Internet das Coisas e o Big Data, os benefícios desse avanço na indústria são favoráveis, como diminuição de despesas, elevação de proteção, diminuição de erros e a redução de desperdício. A transição para a Indústria 4.0 resulta em um crescimento de competição significativo entre as empresas, aumento de eficiência, aumento de receita e otimização dos processos produtivos, criação de tecnologias exponenciais, assim como uma melhor prestação de serviço ao cliente.

Neste contexto, destaca-se a manutenção preditiva, um tipo de manutenção cuja finalidade é antecipar a necessidade de serviços de manutenção em peças específicas, possibilitando seu máximo aproveitamento, aumentando o tempo de disponibilidade do equipamento, evitando emergências causadas por paradas imprevistas. A manutenção preditiva consiste em três etapas principais: aquisição de dados, que se baseia no monitoramento contínuo de uma máquina com sensores interconectados, o processamento dos dados e a terceira etapa consiste na tomada de decisão. (NEPOMUCENO, 2014; ALMEIDA, 2018; ALMEIDA; ROCCA, 2015; JARDINE; LIN; BANJEVIC, 2006).

Utilizando os conceitos da indústria 4.0, a aplicação da regressão logística como método de aprendizado de máquina nos dados coletados das máquinas de RM, é justificada por sua capacidade de proporcionar análises preditivas. Ao adotar esse modelo, é possível obter *feedback* valiosos sobre o desempenho das máquinas, identificar possíveis falhas e realizar manutenções preventivas. Além disso, a regressão logística, devido à sua adaptabilidade e eficácia na classificação de dados, pode ser instrumental na previsão de possíveis problemas ou anomalias no funcionamento das máquinas de ressonância magnética, contribuindo assim para a otimização da eficiência operacional e a redução de custos associados a falhas inesperadas.

Com a implementação da regressão logística, espera-se melhorar a tomada de decisões baseada em dados, promover a eficiência do sistema de manutenção e, por fim, aprimorar a qualidade e confiabilidade dos serviços de ressonância magnética na Indústria 4.0, resultando em benefícios tangíveis tanto para a saúde do equipamento (HIVAREKAR *et al.*, jul. 2020).

1.4 Estrutura do trabalho

A busca pela verdade, com base em critérios científicos, exige um esforço constante e dedicado, envolvendo práticas e conhecimentos relacionados aos fatos investigados. O processo de investigação, visando comprovar a veracidade, envolve a criação e a aceitação de símbolos, fórmulas e outros recursos utilizados para expressar os fenômenos, métodos e processos envolvidos.

No Capítulo 1, são apresentados os problemas discutidos, os objetivos do trabalho e a justificativa para a solução do problema proposto. O Capítulo 2, apresenta os conceitos teóricos que fundamentam os métodos utilizados. No Capítulo 3, são apresentados todos os materiais e métodos utilizados na pesquisa, explicando a escolha dos materiais e como eles contribuíram para a solução do tema do trabalho. O Capítulo 4 descreve os resultados após a coleta de dados e a implementação do modelo, fornecendo uma visão de como o modelo preditivo comportou-se em relação ao banco de dados.

2 REFERENCIAL TEÓRICO

Este capítulo aborda os assuntos que nortearam o desenvolvimento do trabalho. Inicialmente, é apresentado o uso do modelo de regressão logística binária e como deve ser implementado corretamente no contexto abordado no trabalho. Em seguida, é apresentada uma sustentação teórica sobre os tipos de manutenção e por que a manutenção preditiva foi escolhida para aplicação. Nos próximos capítulos, será introduzida a ferramenta computacional utilizada para realizar os cálculos e implementação do modelo de regressão logística binária em manutenções preditivas das máquinas de ressonância magnética nuclear RM.

2.1 Ferramentas utilizadas

A abordagem de aprendizado de máquinas para manutenção preditiva na indústria 4.0 é relativamente nova e se efetivou como uma importante ferramenta de produtividade e redução de custos a partir dos anos 70 (PIRES; OKADA, 2020). A regressão logística binária, juntamente com o aprendizado de máquina na manutenção preditiva, é usado de modo a poder estimar a probabilidade de que o equipamento irá falhar em um período futuro, utilizando seu histórico. No contexto da regressão, o objetivo é encontrar um modelo que calcule a vida útil restante de cada componente (PAOLANTI *et al.*, 2018).

2.2 Regressão Logística

Entre os modelos de regressão, existem duas categorias significativas. A primeira delas é composta pelos modelos de *logit*, nos quais a variável de resposta dependente Y é binária e pode ser relacionada a uma variável independente aleatória X . Nesse caso, segue a distribuição de Bernoulli, incluindo a regressão logística. A segunda categoria é composta pelos modelos *loglineares*, nos quais a variável de resposta é associada a uma variável aleatória Poisson (FIGUEIRA, 2006; FAVERO; FÁVERO, 2015).

Autores como (HARRELL *et al.*, 2001; GONZALEZ, 2018; FIGUEIRA, 2006) consideram a regressão logística uma técnica estatística amplamente empregada para realizar a classificação de padrões úteis via um banco de dados. Esses padrões poderiam ser ignorados por outros modelos. Sua finalidade é prever a probabilidade de um determinado evento ser “sim” ou “não”, “quebrado” ou “não quebrado”, de natureza binária (dicotômica), ou seja, a resposta possui apenas duas categorias e apenas uma variável independente envolvida. O modelo estatístico geralmente preferido para a análise de respostas binárias consegue obter a probabilidade de ocorrência de um determinado evento através da variável qualitativa e a influência independente de cada variável sobre o evento de “quebra” ou “não quebra”.

O modelo de regressão logística permite que sejam realizadas as seguintes ações:

- a) Modelar a probabilidade de um evento ocorrer dependendo dos valores das variáveis independentes, sejam elas categóricas ou contínuas.
- b) Estimar a probabilidade de observação estar em uma categoria determinada
- c) Determinar se um conjunto de variáveis é estatisticamente significativo para o resultado ou se não apresenta contribuição significativa para o resultado.
- d) Verificar a probabilidade de ocorrência de um evento em comparação com a probabilidade de não ocorrência do evento ao selecionar uma observação aleatória.

O método trabalho com a probabilidade do evento, de maneira descrita nas equações (1) e (2):

$$\text{probabilidade de evento}(x) = p \quad (1)$$

$$\text{probabilidade de não evento}(x) = 1 - p \quad (2)$$

2.2.1 Função Logit

O modelo *logit* é uma extensão do modelo de regressão linear, onde a variável de resposta é transformada através da função *logit*. Por conta do modelo logístico binário seguir a distribuição de Bernoulli, logístico: Método da máxima verossimilhança, é preciso realizar um elo entre as variáveis independentes X e as, dependente Y , estimando um p , de forma geral do modelo *logit* é dada pela fórmula (3):

$$\text{logit}(p) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (3)$$

Onde o $\text{logit}(p)$ é o logaritmo da razão entre a probabilidade do evento ocorrer p e a probabilidade do evento não ocorrer, $(1 - p)$, $\beta_0, \beta_1, \beta_2, \dots, \beta_K$ são os coeficientes estimados e X_1, X_2, \dots, X_k são as variáveis explicativas.

Os coeficientes $\beta_0, \beta_1, \beta_2, \dots, \beta_K$ são estimados usando métodos de maximização da verossimilhança ou técnicas de mínimos quadrados. Uma vez estimados, esses coeficientes podem ser usados para fazer previsões sobre a probabilidade de ocorrência do evento binário com base nos valores das variáveis explicativas.

A função *logit* é definida como o logaritmo da razão entre a probabilidade do evento ocorrer e a probabilidade do evento não ocorrer como descrito, onde a razão da probabilidade

pode ser chamado de chance, como mostrado na fórmula (4). Essa transformação permite que a variável de resposta seja modelada em termos de uma combinação linear de variáveis explicativas.

$$\ln(\text{chance}) \rightarrow \ln\left(\frac{p}{1-p}\right) \quad (4)$$

Através da fórmula (4) de *logit*, percebe-se que após a substituição de valores na fórmula como $p=1$, $p=0,5$ e $p=0$, resultará em uma curva característica. Essa curva fica nos intervalos entre 0 e 1. Dessa forma, a probabilidade está nos 0% e 100%. Quando $p=0,5$, a equação (4) retornará um valor de $\ln(1)=0$. Isso demonstra que a probabilidade de quebra e não quebra da máquina são 50%, ambas iguais. A equação que demonstra o *logit* é (5):

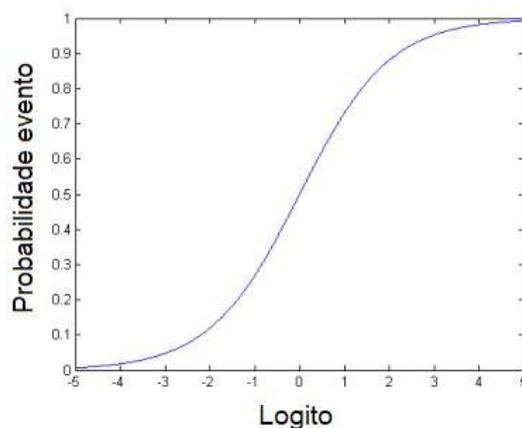
$$\text{logit}(z) = \frac{e^z}{1 - e^z} \quad (5)$$

Para a plotagem do gráfico é necessário usar a inversa da equação (5) usando a regra de Bayes', que pode ser definida apenas por inverter a função do *logit* em 90°, gerando a equação (6):

$$\text{logit}^{-1}(z) = \frac{e^z}{1 + e^z} \quad (6)$$

Pois desse modo a inversa da função *logit* retornará a probabilidade da variável dependente ser 1 ou 0. Observando a Figura 7, obtemos as conclusões descritas.

Figura 7 – Exemplo de curva de *logit* x probabilidade



Fonte: Autor.

A curva da Figura 7 pode ser chamada de curva de sigmoide, e em resumo é apenas a inversão dos eixos x e y.

2.2.2 Regressão Logística Binária

A regressão logística binária representa o caso em que a variável dependente Y é dicotômica ou binária, aceita apenas duas categorias 1 ou 0. e partindo da função do *logit* definimos que a probabilidade do evento é conforme a equação (7):

$$P = \frac{e^z}{1 + e^z} \quad (7)$$

reescrevendo a equação (7) e rearranjando seus coeficientes, chegamos a equação (8) :

$$P = \frac{1}{1 + e^{-z}} \quad (8)$$

colocando a equação de z (9) em função da equação (8) conforme a função descrita na (8), é gerado a equação de probabilidade do evento acontecer conforme (10).

$$z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (9)$$

$$P = \frac{1}{1 + e^{-\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}} \quad (10)$$

O modelo logístico binário a partir das equações tem a probabilidade do evento acontecer é conforme a equação (10), as variáveis independentes são X_1, X_2, \dots, X_k e os parâmetros a serem definidos são α, β_1, β_2 até β_k através do método de máxima verossimilhança.

2.2.3 Método de máxima verossimilhança

O método de máxima verossimilhança é amplamente utilizado para ajustar o modelo de regressão logística binária, visando estimar os coeficientes $\alpha + \beta_1 + \beta_2 + \dots + \beta_k$ com base nos dados coletados. Essa abordagem estatística visa encontrar os valores dos parâmetros que maximizem a probabilidade de replicar o padrão de observação dos dados amostrados, utilizando a função de verossimilhança. Essas estimativas dos parâmetros são consideradas as melhores estimativas no contexto do modelo, ao serem aquelas que maximizam a probabilidade de replicar o padrão observado nos dados amostrados.

Conforme o autor (FIGUEIRA, 2006), um caso específico da regressão logística binária, é a equação de verossimilhança (11):

$$L(\beta|x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n) = \prod_{i=1}^n [\pi(x_i)]^{y_i} \cdot [1 - \pi(x_i)]^{1-y_i} \quad (11)$$

e a realização da transformação da função de verossimilhança que facilita o cálculo e a maximização numérica. Aplicando o logarítmico natural em ambos os lados da equação de verossimilhança (11), a log-verossimilhança é obtida tomando o logaritmo natural da função de verossimilhança conforme (12):

$$l(\beta) = \ln(L(\beta)) = \sum_{i=1}^n [y_i \cdot \ln(\pi(x_i))] + (1 - y_i) \cdot \ln(1 - \pi(x_i)) \quad (12)$$

O valor de β que maximiza o logaritmo natural da função de verossimilhança, $\ln[L(\beta)]$, é obtido através da derivação (12) de $l(\beta)$ em relação aos parâmetros (β_0, β_1) , tal que para (β_0)

$$\frac{\partial \ln(L(\beta))}{\beta_0} = \sum_{i=1}^n [y_i - \pi(x_i)] \quad (13)$$

e para (β_1) (14)

$$\frac{\partial \ln(L(\beta))}{\beta_1} = \sum_{i=1}^n [y_i - \pi(x_i)] \quad (14)$$

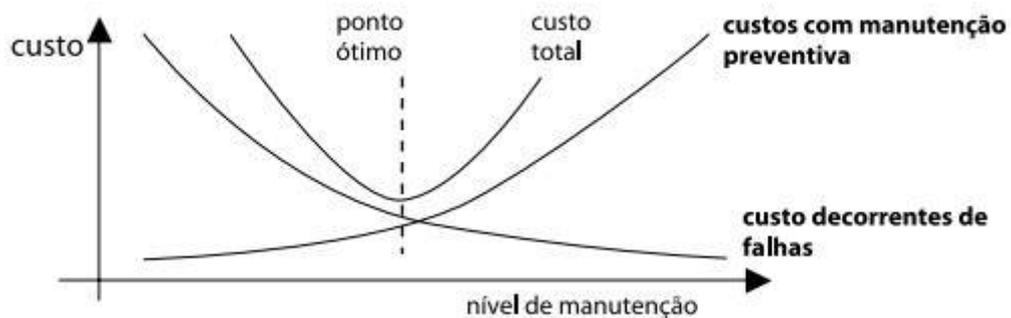
O estimador β_1 mede a taxa de variação do *logit* em outras palavras, o logaritmo da razão de chances conforme (4), para uma unidade de variação na variável independente x_i . Isso significa que β_1 indica o quanto o *logit* muda quando a variável independente aumenta em uma unidade, mantendo todas as outras variáveis constantes. O estimador β_0 , por sua vez, representa o valor do *logit* quando a variável independente é igual a zero. Ele indica o valor inicial do *logit* quando todas as variáveis independentes são nulas, portanto, os estimadores β_0, β_1 fornecem informações importantes sobre a relação entre as variáveis dependentes e independentes e podem ser interpretados como a inclinação da linha de regressão nesse contexto.

2.3 Tipos de Manutenção

Conforme o autor (NEPOMUCENO, 2014) a manutenção é um “mal necessário”, e sua necessidade é indiscutivelmente importante para qualquer tipo de maquinário ou processo de fabricação. (MARCORIN; LIMA, 2003). A falta de uma política de manutenção pode acarretar custos sobre quebra do equipamento, contudo existem os custos invisíveis, são eles: perda de qualidade do produto, penalidades comerciais, custo de maquinário indisponível que remete ao não produção. Em um mercado competitivo, custos excedentes trazem prejuízo e perda

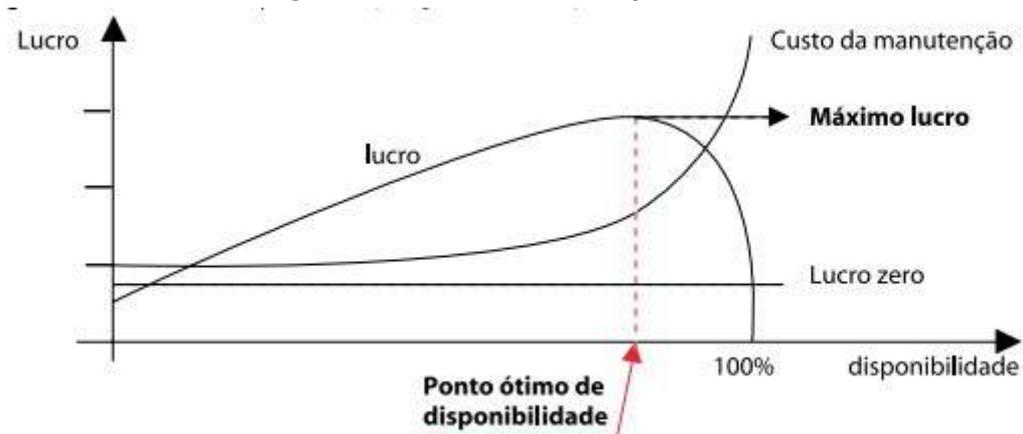
de poder de concorrência, a finalidade dos planos de manutenção é a diminuição de despesas, uma relação mais vantajosa entre custo e benefício ao abordar a conservação preventivamente, em vez de lidar com circunstâncias descontroladas no processo de produção devido à ausência de conservação (CHIU; HUANG, feb. 1996). Conforme o gráfico da Figura 8 o aumento de planejamento de manutenção acarretam redução de custos decorrentes de falhas, conforme (MIRSHAWKA; OLMEDO, 1993), o aumento de planejamento manutenção traz um aumento de custo, e conforme o (MURTY; NAIKAN, 1995) existe um ponto de equilíbrio entre investimento de manutenção, lucro gerado pela máquina e sua disponibilidade, essa a afirmação fica melhor exemplificada na Figura 9.

Figura 8 – Gráfico custo em relação ao nível de manutenção.



Fonte: (MIRSHAWKA; OLMEDO, 1993).

Figura 9 – Ponto de ótimo disponibilidade.



Fonte: (MURTY; NAIKAN, 1995).

Para definir o ponto de equilíbrio é necessário saber qual modelo de manutenção vai se adequar melhor ao maquinário, o que acarretam as falhas, custos da manutenção, esses parâmetros se alteram dependendo do maquinário. Os modelos mais conhecidos são os de manutenção corretiva, preventiva e preditiva, que serão abordados nas próximas subsucessões.

2.3.1 Manutenção Corretiva

Conforme (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1994), página 7, 2.8.8, define Manutenção Corretiva como:

- a) Manutenção efetuada após a ocorrência de uma pane destinada a recolocar um item em condições de executar uma função requerida.

Além da definição da Normas Brasileiras (NBR) sobre manutenção corretiva, outros autores como (MARCORIN; LIMA, 2003), (PIRES; OKADA, 2020) e (SOUZA, 2006) descrevem o que é a manutenção corretiva. A manutenção corretiva é uma abordagem de manutenção que visa restaurar um equipamento ou sistema ao seu estado operacional normal após a ocorrência de uma falha, ou quebra, sendo realizada de forma reativa e imediata para minimizar o tempo de paralisação e os impactos operacionais. É realizada reativamente, em resposta direta ao mau funcionamento ou paralisação do equipamento. Conforme o gráfico 8, essa manutenção possui custos de planejamento mínimos, estando localizada à esquerda do gráfico 8.

Esse tipo de manutenção é realizada quando não há outras opções imediatas disponíveis ou quando os custos associados à paralisação do equipamento são menores do que os custos de prevenção ou de outros tipos de manutenção.

A manutenção corretiva pode ser classificada em dois tipos principais:

- a) Manutenção corretiva não planejada: nesse caso, a intervenção ocorre de maneira imprevista e emergencial, sem um planejamento prévio. A ação é tomada assim que a falha é detectada, visando restaurar o funcionamento normal do equipamento o mais rápido possível.
- b) Manutenção corretiva planejada: Nesse tipo de manutenção corretiva, embora a falha seja identificada, a intervenção é agendada e planejada para ocorrer em um momento mais conveniente, minimizando o impacto na operação e produtividade. Isso permite que os recursos e materiais necessários sejam organizados antecipadamente, visando uma resolução eficiente e eficaz do problema.

2.3.2 Manutenção Preventiva

Conforme a (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1994), página 7, 2.8.8, define Manutenção Preventiva como:

- Manutenção 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.

Autores (PIRES; OKADA, 2020; SOUZA, 2006; NEPOMUCENO, 2014) expõem sua explicação sobre manutenção preventiva. A manutenção preventiva é uma abordagem sistemática de manutenção que visa evitar falhas e minimizar a degradação do desempenho dos equipamentos. Essa estratégia envolve a realização de ações planejadas e periódicas em intervalos pré-determinados, considerando, fatores como recomendações do fabricante, requisitos regulatórios e condições operacionais, visando inspecionar, ajustar, lubrificar, substituir peças desgastadas e realizar outras atividades preventivas para manter os equipamentos em condições adequadas de funcionamento. A manutenção preventiva é uma opção mais vantajosa em comparação à manutenção corretiva em máquinas ligadas diretamente ao processo, por reduzir o risco de paralisações não programadas devido a falhas no equipamento, pode ser classificada por:

- a) Manutenção preventiva baseada no tempo: nesse caso, as atividades de manutenção são realizadas com base em intervalos fixos de tempo. Por exemplo, determina-se que determinada inspeção ou substituição de peças seja feita a cada três meses, independentemente do estado real do equipamento.

Ao adotar a manutenção preventiva, as organizações podem reduzir o risco de falhas inesperadas, evitar paralisações não programadas, prolongar a vida útil dos equipamentos e aumentar a confiabilidade das operações. Porém, conforme o gráfico 8, essa manutenção possui custos de planejamento altos, e se não for bem planejada, pode se desviar do "Ponto ótimo" conforme (MIRSHAWKA; OLMEDO, 1993). Outro empecilho desse tipo de manutenção é realizar inspeções e substituições periódicas de componentes, pois não consideram a condição real do equipamento, o que pode resultar em custos desnecessários.

2.3.3 Manutenção Preditiva

Conforme os autores (PIRES; OKADA, 2020; MARCORIN; LIMA, 2003; NEPOMUCENO, 2014), a manutenção preditiva é uma abordagem sistemática e baseada em informações para a manutenção de equipamentos e sistemas. Essas informações podem incluir sinais como temperatura, vibração, pressão, tempo de uso, fluxo e outros parâmetros relevantes. Essa abordagem utiliza técnicas de monitoramento, coleta de dados e análise para determinar o estado atual de funcionamento e antecipar possíveis falhas ou problemas futuros. Com isso, a equipe responsável pela manutenção pode se programar para a intervenção e aquisição de peças, resultando em redução de despesas com estoque e evitando paradas desnecessárias na linha de produção, levando a uma redução dos custos.

Com base nas análises, é possível tomar decisões informadas sobre a manutenção a ser realizada. As intervenções são planejadas, para evitar falhas iminentes e realizar reparos ou substituições de componentes antes que ocorram danos significativos, ou paralisações não programadas.

Ao adotar a manutenção preditiva, as empresas podem alcançar uma série de benefícios, como redução de custos operacionais, otimização do uso de recursos, aumento da disponibilidade e confiabilidade dos equipamentos, além de maximização da vida útil dos componentes, porém esse modelo de manutenção necessita de uma mão-de-obra mais qualificada para o trabalho e alguns aparelhos ou instrumentos de medição, mesmo com o aumento de gastos com mão de obra e equipamento, temos uma diminuição de gastos de troca desnecessário de peças com vida útil e manutenção corretiva não planejadas, fazendo com que esse modelo de manutenção situando-se mais próximo do “ponto ótimo” conforme a Figura 8 da relação custo-benefício em equipamentos cuja parada traz grandes prejuízos ao processo, como as RM.

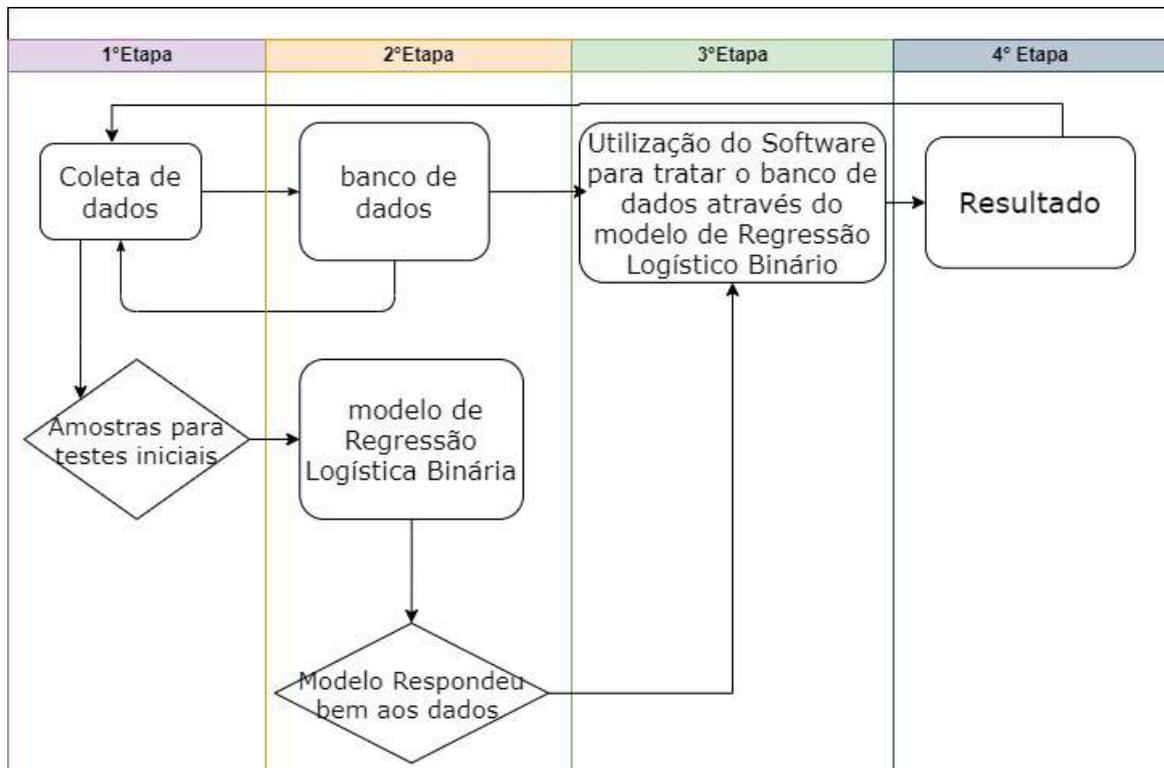
Em resumo, a manutenção preditiva é uma estratégia proativa baseada em dados e análises que permite a detecção precoce de falhas e a implementação de ações corretivas antes que ocorram paralisações ou danos significativos aos equipamentos.

3 MATERIAIS E MÉTODOS

A manutenção preditiva desempenha um papel crucial na eficiência e no funcionamento adequado das máquinas (MARCORIN; LIMA, 2003). Com o avanço da tecnologia, é viável empregar métodos estatísticos para antever falhas e estabelecer a probabilidade de quebra de uma máquina ou peças. Nesse contexto, o trabalho desenvolveu um modelo para prever falhas de uma RM por meio da utilização do método estatístico de regressão logística binária, mediante a coleta de dados de falhas anteriores, sensores de temperatura, sensores de pressão, entre outros dispositivos. Por meio da aplicação do método de regressão logística binária em um *software* estatístico (GONZALEZ, 2018), analisando os dados coletados e a probabilidade de falha dos equipamentos. Essa análise dos resultados possibilitou a definição de uma possível falha e conseqüentemente uma intervenção de manutenção preditiva mais eficiente e a redução de custos com manutenções preventivas e corretivas (MARCORIN; LIMA, 2003; NEPOMUCENO, 2014).

Adicionalmente, a busca por dados complementares de outras máquinas do mesmo modelo foi conduzida para ampliar o banco de dados e aprimorar a precisão dos cálculos (PAOLANTI *et al.*, 2018). Além da descrição, utilizou o cronograma conforme a Figura 10 para auxiliar na execução dos métodos.

Figura 10 – Fluxograma dos passos realizados



Fonte: Autor.

3.1 Materiais

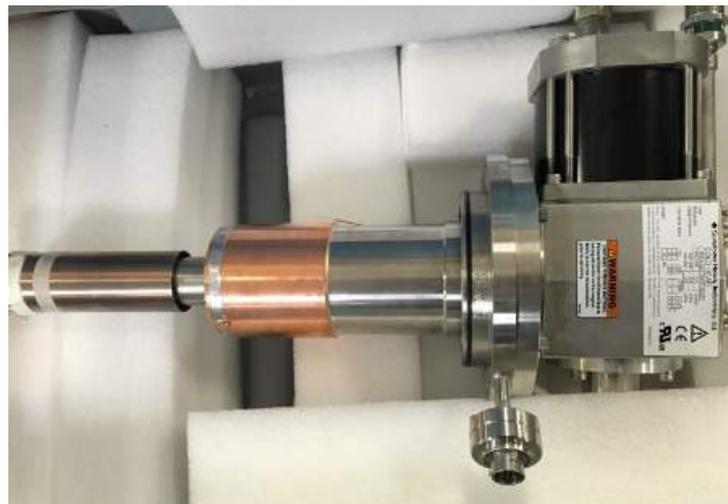
A conexão entre parte teórica (2) e sua aplicação prática é concretizada pelo uso das ferramentas adequadas, tais materiais e ferramentas desempenham um papel crucial nos passos representados no diagrama mostrado na Figura 10.

O emprego de materiais apropriados evidencia o compromisso acadêmico e a excelência profissional do estudo, ressaltando a vital contribuição desses elementos na concretização da pesquisa. Esses materiais oferecem as ferramentas e tecnologias essenciais para a coleta de dados, análise de informações, construção de modelos e execução de experimentos. A referência aos materiais possibilita que leitores e pesquisadores possam reproduzir e validar os resultados alcançados, bem como de avaliar a consistência dos mesmos.

Os dados coletados da RM são referentes ao *datalogger* das máquinas e sensores externos conforme a Tabela 1, o *datalogger* conforme a Figura 12 armazena todas as medidas dos sensores da parte de criogenia, os sensores são: Nível de Hélio, pressão de hélio, temperatura do *Cold Head* em Kelvin (K), temperatura do *Shield* em K, temperatura água do *Chiller* em Grau Celsius (°C), Recon_ru e Recon_si também medidos em K.

O *Cold Head* 11 tem a função de assegurar a estabilidade do núcleo do magneto mantendo-o na temperatura de funcionamento de -269,0 °C, minimizando o consumo de hélio líquido, toda a criogenia opera de forma contínua, sem interrupções, durante todos os dias do ano. Por isso, é comum que a peça se desgaste ao longo do tempo, reduzindo sua vida útil do *Cold Head* para aproximadamente 2 anos, tornando necessária a sua substituição, a ser realizada quando necessário.

Figura 11 – Cold Head



Fonte: Autor.

Figura 12 – Sistema *datalogger* Magnet Monitor da RM GE LX 1.5 T



Fonte: Autor.

Os sensores externos são: umidade da sala de máquinas, temperatura de componentes internos, temperatura da sala de máquinas em °C e temperatura sala operador em °C . Os dados históricos são referentes ao tempo de uso da máquina, quantos exames realizou por mês, além dos dados já existentes como o histórico de quebras e trocas de componentes devido ao desgaste, conforme demonstrado na Tabela 1.

Tabela 1 – Dados coletados do *datalogger* e sensores externos do sistema de ressonância magnética GE LX 1.5 T

Máquina	M1	M2	M3
Nível de Hélio	82.46%	60.66%	78,29%
Pressão de Hélio	2,66	4,35	1,44
Temperatura K <i>Cold Head</i>	4,27	4,35	4,28
Temperatura K de <i>Shield</i>	41,34	43,69	40,095
Temperatura °C Água <i>Chiller</i>	10,79	17,73	17,76
Temperatura K Recon_ru	4,344	4,129	4,446
Temperatura K Recon_si	3,56	3,455	4,22
Tempo de uso (em meses)	95
Quantidade de exames	10120	9920	...
Temperatura °C sala de máquinas	21	24	24
Umidade sala de máquinas	46,5	47,3	50,7

Fonte: Autor.

As ferramentas e materiais empregados para o desenvolvimento serão listados a seguir:

- Computador com memória disponível para a aquisição e armazenamento de dados.
- Máquinas General Electric LX 1.5 T (três unidades) para coleta de dados.
- Planilha eletrônica (ou *software* similar) para criação do banco de dados.

- d) *Software Orange Canvas*, para a aplicação do método de regressão logística binária.
- e) Dados de quebras anteriores de sensores das máquinas.

O fato de ser necessário um computador com memória disponível, é para evitar possíveis falhas de processamento do banco de dados, ou afetar as simulações é essencial memória Random Access Memory (RAM) compatível com o tamanho do banco de dados a ser processado ou realizado as simulações no *software Orange*. O computador desempenhou um papel indispensável nesta pesquisa, pois por meio dele que todos os dados foram coletados e armazenados. Além disso, o computador abriga os *softwares* necessários para a implementação da regressão logística binária, além da planilha eletrônica, de extrema importância para a organização dos dados e a implementação do modelo.

Nesse sentido, foi crucial realizar visitas às clínicas que possuem as máquinas do modelo GE LX 1.5 T, buscando os dados dos sensores e das quebras anteriores. Esses dados foram adicionados ao banco de dados. É importante ressaltar que, na região onde a pesquisa foi conduzida, existem três máquinas do modelo GE LX 1.5 T, no entanto, para futuras pesquisas, é recomendado ampliar o banco de dados, coletando dados de outras máquinas do mesmo modelo. Isso contribuirá para aumentar a confiabilidade e o *feedback* do modelo de regressão logística.

Portanto, o *notebook*, os *softwares*, as visitas às clínicas para ampliação do banco de dados e a aplicação do modelo e comparação com resultados de quebras anteriores foram etapas essenciais para obter confiabilidade ao modelo de previsão, visando construir um modelo de manutenção preditiva eficiente utilizando o método da regressão logística binária.

3.2 Métodos

A seguir é definido o plano do trabalho, conforme já descrito no diagrama da Figura 10, estão descritos etapas realizadas e a ordem em que foram efetuadas, citadas a seguir e com uma explicação do que foi feito em cada uma.

a) Coleta de Dados:

- Aquisição dos dados de quebras anteriores a partir dos sensores das máquinas.
- Utilização de um computador para realizar a coleta.

b) Criação do Banco de Dados:

- Adição dos dados coletados em uma planilha eletrônica (Excel).
- Organização e armazenamento dos dados na planilha.

c) Aplicação do Método de Regressão Logística Binária:

- A implementação do método de regressão logística binária no *software* estatístico, Orange.

d) Análise dos Resultados:

- Avaliação dos resultados da previsão através da matriz de confusão e da curva ROC.

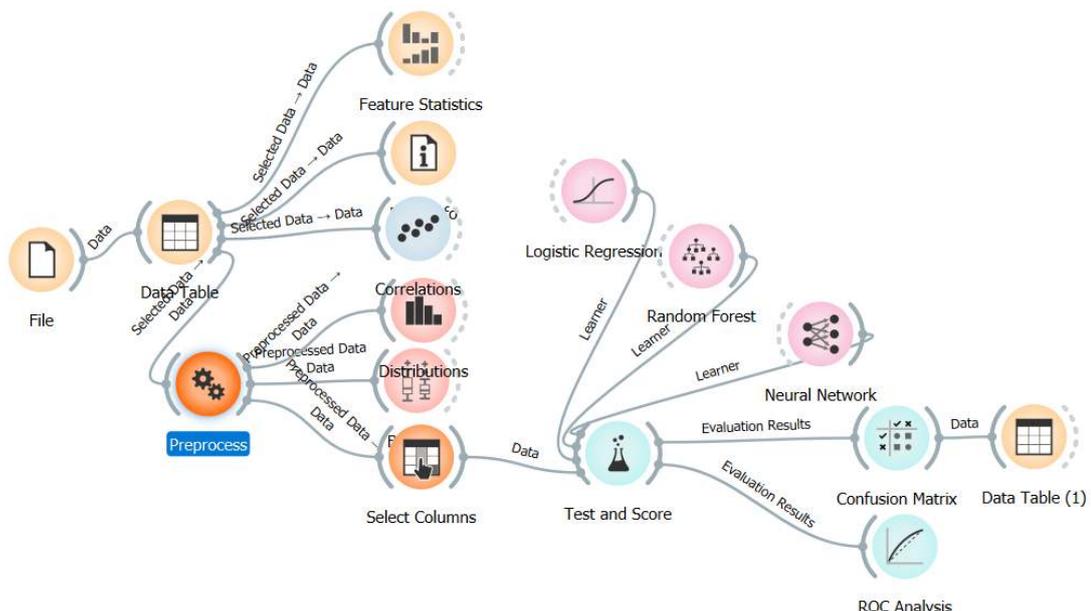
3.2.1 Implantação do sistema

O modelo completo e todos os *widget* utilizados estão na Figura 13. Os resultados do *machine learning* dos modelos são descritos no Capítulo 4, juntamente com todas as ressalvas relacionadas à aplicação do modelo e ao banco de dados, bem como a interpretação das métricas estatísticas.

Na aplicação do modelo inicialmente, é necessário converter o *log* retirado do *datalogger* da máquina de arquivo de texto para um formato *Comma-separated values* (CSV). No Excel foi aplicado um filtro de amostragem por motivos de processamento e redundância de dados. A amostragem escolhida para os dados foi de meia em meia hora, conforme demonstrado na Figura 25, em vez de minuto a minuto, como é quando o *log* é retirado da máquina, conforme ilustrado na Figura 16.

O próximo passo é realizar o *upload* do arquivo filtrado no *software* Orange. Dentro do Orange, deve configurar todos os *widget*, conforme apresentado na Figura 13, e realizar os ajustes mencionados para obter os melhores resultados para o banco de dados utilizado. Finalmente, procede-se à coleta dos resultados obtidos, e então o aumento do banco de dados.

Figura 13 – Sistema completo no *software* Orange



Fonte: Autor.

Observa-se que no modelo completo da Figura 13 foram implementados mais dois modelos, são eles *Neural Networks* e *Random Forest*, ambos serviram para comparação entre modelos de *machine learning*, e verificar se o modelo de regressão logística foi o melhor método utilizado para os *logs* obtidos, no Capítulo 4 os resultados foram avaliados e apresentados.

Random Forest é um modelo ensemble de aprendizado de máquina, construído a partir de árvores de decisão. Esse método pode ser aplicado tanto em tarefas de classificação quanto em tarefas de regressão, uma vez que combina as previsões de várias árvores de decisão em um único modelo.

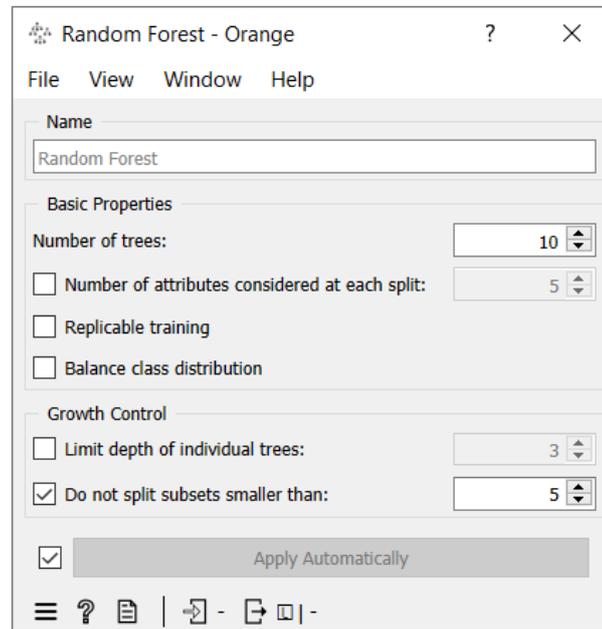
Quanto ao desempenho, a *Random Forest* é reconhecido por sua eficácia em diversos cenários, mesmo quando se trata de dados com alta dimensionalidade ou características não lineares. Suas aplicações são amplas e incluem classificação de imagens, diagnóstico médico, previsão de preços de ações e muitas outras.

A *Random Forest* é constituída por um conjunto de preditores de árvores, onde cada árvore depende dos valores de um vetor aleatório amostrado de maneira independente e com a mesma distribuição para todas as árvores na floresta. Conforme o número de árvores na floresta aumenta, o erro de generalização da floresta converge para um limite. Esse erro de generalização depende da qualidade das árvores individuais na floresta e da correlação entre elas (NASCIMENTO, 2022).

Um aspecto interessante da *Random Forest* é a seleção aleatória de recursos para dividir cada nó da árvore, resultando em taxas de erro comparáveis ao *Adaboost* outro algoritmo de *machine learning*, porém com maior robustez em relação ao ruído. Além disso, estimativas internas são usadas para monitorar o erro, a força e a correlação, permitindo avaliar o impacto do aumento do número de recursos na divisão. Essas estimativas internas também são úteis para medir a importância das variáveis, sendo aplicáveis tanto em problemas de classificação quanto em problemas de regressão (MARINS, 2016).

Através da Figura 14, observam-se as configurações utilizadas para a aplicação do modelo *Random Forest*. O número de árvores (*number of trees*) escolhido foi 10 árvores de decisão na floresta, e optou-se por não dividir subconjuntos menores que 5. Como o foco do trabalho não é a aplicação do *Random Forest* nem das *Neural Networks*, as configurações de pré-processamento e controle de crescimento foram mantidas nas configurações do *software* Orange.

Figura 14 – Configuração Random Forest

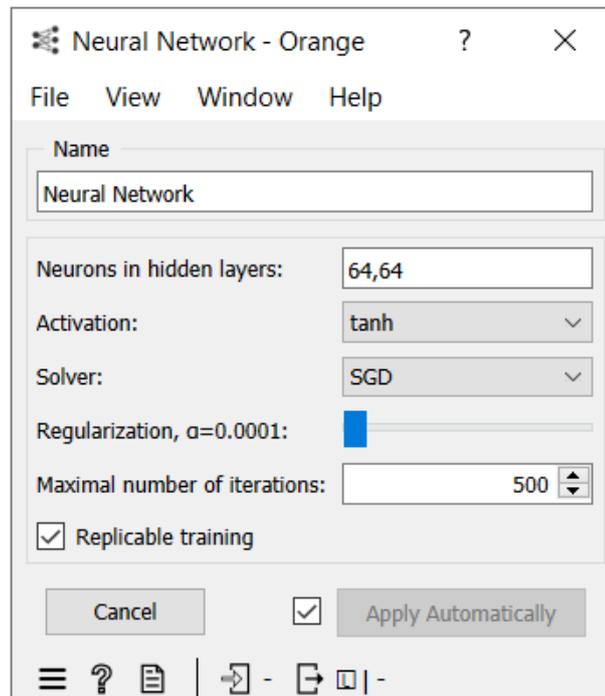


Fonte: Autor.

As redes neurais, ou *Neural Networks*, são modelos de aprendizado de máquina que consistem em múltiplos neurônios interconectados por pesos adaptáveis, organizados em uma estrutura paralela de neuro-computação. Esse arranjo permite o processamento de dados paralelamente, diferenciando-se do processamento sequencial adotado em abordagens tradicionais. Devido a essa característica, a falha de alguns neurônios não afeta de maneira significativa o desempenho do sistema, conferindo-lhe uma importante qualidade denominada tolerância a falhas (MENDONÇA; NÉIA; BRINHOLE, 2002).

As redes neurais podem se adaptar e aprender a partir do ambiente, essa habilidade possibilita que os modelos de redes neurais lidem com dados imprecisos e situações não completamente definidas, tornando-os robustos. Mesmo na presença de pequenas falhas, a rede neural consegue reconhecer padrões específicos, mesmo que haja distorções em suas entradas (CERQUEIRA *et al.*, mar. 2001).

As configurações escolhidas para *Neural Networks* estão exemplificadas na 15, *Neurons in hidden layers* Neurônios por camada oculta definido como o *i*-ésimo elemento representa o número de neurônios na *i*-ésima camada oculta, foi escolhido 64,64. A função de ativação "*Activation*" escolhida foi a "*tanh*" função tan hiperbólica. A função solucionador para otimização de peso "*solver*", foi utilizado o modo "*Stochastic Gradient Descent* (SGD)" descida gradiente estocástica. Para *Maximal number of iterations* número máximo de interação que o modelo pode fazer foi escolhido 500. Outros parâmetros como pré-processamento foi deixado como vem no modelo originalmente.

Figura 15 – Configuração Neural Networks

Fonte: Autor.

3.2.2 Modelagem e implementação da regressão logística

Como mencionado anteriormente, a modelagem do sistema foi realizada utilizando o *software* Orange. Nesta seção, descreve-se em detalhes como a modelagem do sistema foi conduzida, juntamente com a implementação do método de regressão logística, após a coleta dos dados e sua conversão para o formato CSV.

Com diversos testes e observações do sistema, foi identificado que a filtragem inicial da planilha no Excel seria benéfica para o processamento dos dados pelo *software* Orange. Portanto, optou-se por classificar todo o banco de dados, mantendo na planilha apenas as linhas com valores registrados a cada meia hora. Isso resultou em uma redução significativa no tamanho do arquivo, de 15.728.640 bytes para 393.216 bytes. Essa abordagem foi escolhida porque o *datalogger* registra os dados minuto a minuto, conforme mostrado na Figura 16, o que tornava alguns dos dados redundantes.

Figura 16 – Exemplo de dados coletados com aspecto redundante da RM GE

Date	Time	He_Level	Water_Temp	Shield_Si410	Recon_RuO	Recon_Si410	Coldhead_RuO	He_Pressure
11-Apr-23	08:29	82.34	11.570	39.970	4.374	3.563	4.268	3.075
11-Apr-23	08:30	82.34	11.843	39.970	4.374	3.563	4.268	3.075
11-Apr-23	08:31	82.34	11.911	39.970	4.374	3.563	4.268	3.075
11-Apr-23	08:32	82.34	11.870	39.970	4.374	3.563	4.268	3.075
11-Apr-23	08:33	82.34	11.761	39.970	4.374	3.563	4.268	3.075
11-Apr-23	08:34	82.34	11.529	39.970	4.374	3.563	4.268	3.075
11-Apr-23	08:35	82.34	11.202	39.970	4.374	3.563	4.268	3.075
11-Apr-23	08:36	82.34	10.916	39.970	4.374	3.563	4.268	3.075
11-Apr-23	08:37	82.34	10.643	39.970	4.374	3.563	4.268	3.075
11-Apr-23	08:38	82.34	10.480	39.970	4.374	3.563	4.268	3.075
11-Apr-23	08:39	82.34	10.275	39.970	4.371	3.563	4.268	3.075
11-Apr-23	08:40	82.34	10.152	39.970	4.371	3.563	4.268	3.075
11-Apr-23	08:41	82.34	10.084	39.970	4.371	3.563	4.268	3.079
11-Apr-23	08:42	82.34	10.057	39.970	4.371	3.563	4.268	3.079
11-Apr-23	08:43	82.34	9.989	39.970	4.371	3.563	4.268	3.079
11-Apr-23	08:44	82.34	9.962	39.970	4.371	3.563	4.353	3.079
11-Apr-23	08:45	82.34	9.962	39.970	4.371	3.563	4.353	3.079
11-Apr-23	08:46	82.34	10.125	41.586	4.382	3.563	4.232	3.079
11-Apr-23	08:47	82.34	10.316	39.970	4.374	3.563	4.299	3.079
11-Apr-23	08:48	82.34	10.643	39.970	4.374	3.563	4.314	3.079
11-Apr-23	08:49	82.34	10.929	42.578	4.260	3.563	4.337	3.075
11-Apr-23	08:50	82.34	11.379	39.970	4.377	3.563	4.273	3.079
11-Apr-23	08:51	82.34	11.856	39.970	4.379	3.563	4.232	3.075

Fonte: Autor.

Ao abrir o *software* Orange, é necessário importar a planilha para o *software* através do *widget File*. O Orange possui um menu lateral onde estão disponíveis todas as ferramentas necessárias. No conjunto de ferramentas chamado “*Data*”, utiliza-se o *Data Table* para exibir a tabela importada. Antes de prosseguir, é fundamental definir o tipo de variável para cada coluna, conforme a Figura 17.

No caso, o nome da máquina, a data e a hora não são relevantes para a regressão logística; portanto, define-se essas variáveis como texto e sua função como *meta*, o que significa que elas não têm valor significativo para o modelo. Por outro lado, as variáveis de nível de hélio *He Level*, temperatura da água que circula no trocador de calor *Water Temp*, temperatura da capa externa do magneto *Shield*, Recon superior e inferior, *Cold Head* e pressão de hélio são do tipo numérico, por representarem valores medidos pelos sensores de temperatura, pressão e nível.

Por fim, a variável de falha que nos interessa é categórica, mas sua especificação é o alvo (*target*) que planejamos alcançar: ‘sim’ para falha e ‘não’ para máquina em funcionamento normal.

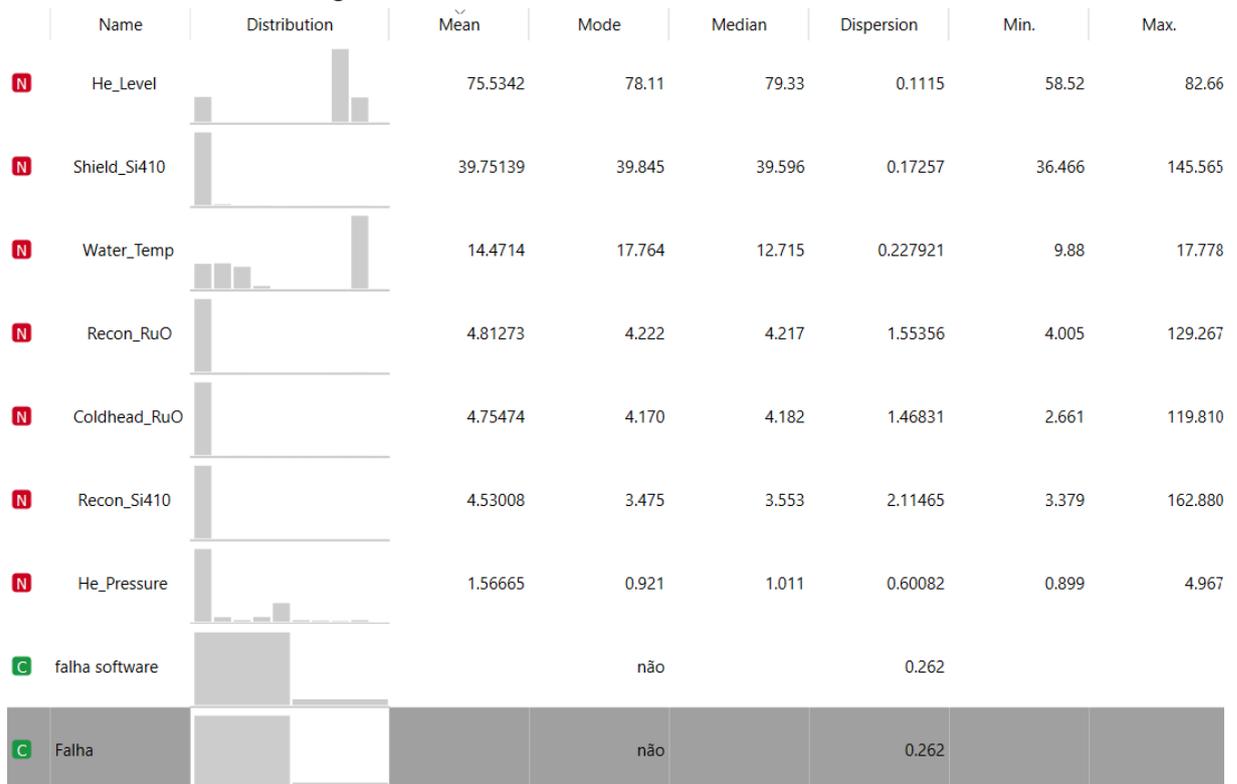
Figura 17 – Definição das categorias das variáveis

Info			
6735 instances			
12 features (no missing values)			
Data has no target variable.			
0 meta attributes			
Columns (Double click to edit)			
	Name	Type	Role
1	Maquina	S text	meta
2	Date	S text	meta
3	Time	S text	meta
4	He_Level	N numeric	feature
5	Water_Temp	N numeric	feature
6	Shield_Si410	N numeric	feature
7	Recon_RuO	N numeric	feature
8	Recon_Si410	N numeric	feature
9	Coldhead_RuO	N numeric	feature
10	He_Pressure	N numeric	feature
11	falha software	C categorical	feature
12	Falha	C categorical	target

Fonte: Autor.

No próprio conjunto *Data* existe uma ferramenta *Feature Statistics* onde é possível verificar as estatísticas de cada variável. Ao analisar os dados da tabela da Figura 18, é importante compreender algumas medidas estatísticas essenciais. A distribuição dos valores descreve como as observações estão espalhadas no conjunto de dados, podendo apresentar diferentes padrões, como uniformidade ou assimetria. A moda representa o valor mais frequente, enquanto a média é a total dos valores dividida pelo número de observações, indicando o valor médio. A mediana é o valor no meio dos dados quando organizados em ordem crescente e não é afetada por valores extremos. A dispersão revela a extensão da variação dos dados em relação à medida de tendência central, incluindo medidas como a amplitude, desvio padrão e variância. Por fim, o valor máximo é o maior valor observado na amostra, enquanto o valor mínimo é o menor valor. Essas medidas estatísticas são cruciais para compreender a natureza dos dados, identificar padrões e tendências, e interpretar os resultados obtidos por meio da regressão logística.

Figura 18 – Estatísticas relacionadas as variáveis



Fonte: Autor.

Na Figura 18 percebe-se uma grande variação dos valores máximos e mínimos das temperaturas do *Cold Head* 11, Recon e *Shield*, notando que já houve um *log* que apresentou uma falha/troca na parte de criogenia de uma das máquinas.

A falha na parte de criogenia da máquina de RM é, indiscutivelmente, temida e indesejada. Isso porque a criogenia desempenha um papel fundamental no perfeito funcionamento dos equipamentos de ressonância magnética de alto campo. O sistema é responsável por manter o campo magnético ativo por meio do resfriamento da bobina supercondutora. Um supercondutor é um material que, quando resfriado a temperaturas extremamente baixas, perde toda a sua resistência elétrica, permitindo a passagem de corrente elétrica sem perdas significativas de energia. Portanto, as bobinas supercondutoras são cruciais para criar campos magnéticos intensos e estáveis em máquinas de RM.

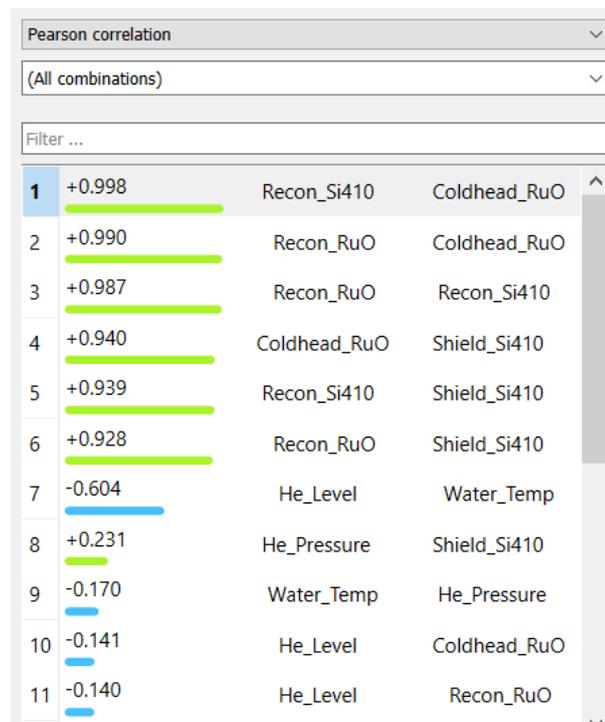
Qualquer falha nesse sistema pode resultar em interrupções significativas nos procedimentos de Ressonância Magnética, afetando a qualidade das imagens ou até mesmo o aborto do exame e a parada do equipamento, comprometendo potencialmente o diagnóstico médico ou a estrutura da máquina. Além disso, uma falha nesse componente pode ser extremamente custosa de ser reparada. Se não for corrigida rapidamente, pode levar a erros irreversíveis nos equipamentos, exigindo um alto investimento na manutenção do equipamento. Portanto, a manutenção adequada e a prevenção de falhas na criogenia é de extrema importância para garantir o desempenho confiável desses equipamentos de RM e minimizar os custos associados à sua operação.

Observando mais profundamente os dados da Figura 18 e da Tabela 16, nota-se que o aumento de algumas temperaturas não influenciam outras. Com uma análise mais detalhada usando o *software* Orange, é possível identificar a correlação entre essas medições conforme a Figura 19 e 20.

A correlação positiva é definida quando duas variáveis aumentam juntas, ou seja, uma variável influencia na outra, já a correlação negativa é definida quando uma variável aumenta enquanto a outra diminui.

A correlação de Pearson dada pela Figura 19 é uma medida de correlação usada para avaliar a relação linear entre duas variáveis contínuas. O coeficiente de correlação de Pearson varia de -1 a 1, onde -1 representa uma correlação negativa perfeita, 1 representa uma correlação positiva perfeita e 0 representa nenhuma correlação linear.

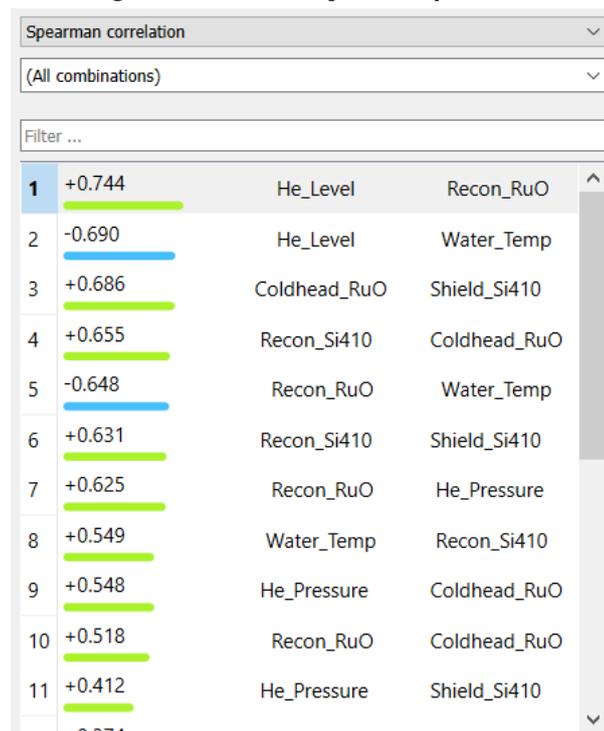
Figura 19 – Correlação de Pearson



Fonte: Autor.

A correlação de Spearman dada pela Figura 20 pode ser descrita como uma medida de correlação que avalia a relação entre duas variáveis, mas não assume uma relação linear. É útil quando as variáveis não são necessariamente distribuídas normalmente ou quando há presença de *outliers* como no caso quando ouve a falha na máquina.

Figura 20 – Correlação de Spearman



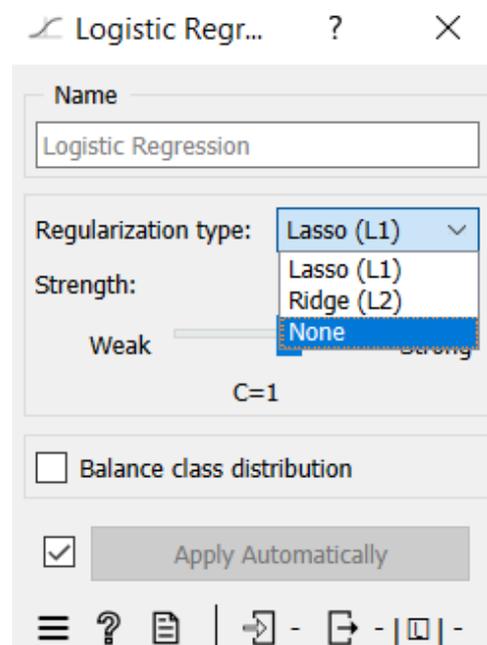
Fonte: Autor.

Observando as Figuras 19 e 20, nota-se que o *Cold Head* 11 e o *Shield* são duas das variáveis com uma forte correlação. De fato, o *software* acertou nisso, pois o *Shield* é uma proteção para o *Cold Head*, e à medida que a temperatura aumenta, consequentemente, a temperatura da sua proteção também sobe. Outra correlação, mas negativa, é entre o nível de hélio e a temperatura da água *Water Temp*.

O próximo passo na aplicação do modelo é o pré-processamento dos dados, uma etapa importante que visa colocar os valores em escalas semelhantes, de modo a trazer todos os dados para o mesmo intervalo e evitar distorções entre as diferentes variáveis. O pré-processamento desempenha um papel crucial na prevenção de futuros problemas e é uma prática essencial para garantir que o modelo de *machine learning* não seja tendencioso em relação às variáveis com maior ordem de grandeza. No trabalho, realizou-se o pré-processamento utilizando a normalização, o que resultou na padronização de todas as variáveis para a mesma ordem de grandeza, com média igual a 0 e desvio padrão igual a 1, isso ajuda a garantir que todas as variáveis tenham um impacto equilibrado no modelo e contribuam justamente para a previsão da quebra de equipamentos.

Adicionando o *widget* da regressão conforme a Figura 13, deve ser realizado ajustes para a aplicação do modelo no *software* e no banco de dados, uma vez que ele envolve a definição de parâmetros de regularização do modelo, como ilustrado na Figura 21.

Figura 21 – Parâmetros regressão logística



Fonte: Autor.

Os algoritmos de classificação de regressão logística com regularização Least Absolute Shrinkage and Selection Operator (Lasso) (L1) ou Ridge Regression (Ridge) (L2) são técnicas usadas em modelos de regressão para evitar o *overfitting* e melhorar o desempenho do modelo. Ambas são métodos de regularização que adicionam penalidades aos coeficientes do modelo com base nos valores desses coeficientes. A regularização Ridge é eficaz na redução do impacto de coeficientes altamente correlacionados, evitando que eles se tornem muito grandes e, assim, reduzindo a chance de *overfitting*. A regularização Ridge é útil quando há multicolinearidade (alta correlação entre variáveis de entrada) nos dados. Por outro lado, a regularização Lasso consegue forçar alguns dos coeficientes do modelo a serem exatamente iguais a zero, o que a torna útil para seleção de recursos, eliminando variáveis menos importantes.

Com todos os dados processados, na aba de *Evaluate* existe uma ferramenta chamada *Test and score*, essa ferramenta dirá as estatísticas do modelo de regressão logística como de outros métodos sendo possível comparar qual modelo é mais eficiente perante o banco de dados, uma das métricas é a acurácia, essa medida indica a proporção de previsões corretas feitas por um modelo em relação ao total de previsões feitas. Em outras palavras, a acurácia mede a precisão geral do modelo em classificar corretamente os exemplos em um conjunto de dados.

A acurácia muito é utilizada em problemas de classificação ou categorização, sua fórmula é dada pela equação (15), acaba sendo uma métrica útil para avaliar o desempenho geral de um modelo de classificação, mas não é a única métrica a ser considerada. Em alguns casos, a acurácia pode ser enganosa, especialmente quando as classes estão desequilibradas (ou seja, uma classe tem muito mais exemplos do que a outra). Nesses casos, outras métricas,

como precisão (17), *recall* (16) (ou sensibilidade) e *F1-score* (18) podem fornecer uma visão mais completa do desempenho do modelo.

$$CA = \frac{\text{Numerodeprevisoescorretas}}{\text{totaldeprevisoes}} \quad (15)$$

$$\text{Recall} = \frac{\text{VerdadeirosPositivos}}{\text{VerdadeirosPositivos} + \text{FalsosNegativos}} \quad (16)$$

$$P = \frac{\text{VerdadeirosPositivos}}{\text{VerdadeirosPositivos} + \text{FalsosNegativos}} \quad (17)$$

$$\text{F1-score} = \frac{2 * \text{precisao} * \text{recall}}{\text{Precisao} + \text{Recall}} \quad (18)$$

Outras métricas observadas no *Test and score* são o *recall* que mede a proporção de exemplos positivos reais que foram corretamente identificados pelo modelo, o *recall* dado pela equação (16) sua aplicação é útil quando é importante detectar a maioria dos exemplos positivos reais, mesmo que isso signifique ter alguns falsos positivos. A precisão dada pela equação (17) que mede a proporção de exemplos classificados como positivos pelo modelo que realmente são positivos, a precisão é útil quando deseja ter certeza de quando o modelo faz uma previsão positiva, ela é confiável. Já o *F1-score* dado pela equação (18), é uma métrica que combina precisão e *recall* em um único número que equilibra essas duas medidas, o *F1-score* é útil quando deseja-se encontrar um equilíbrio entre precisão e *recall*, sendo especialmente útil quando as classes estão desequilibradas, ao considerar tanto os falsos positivos quanto os falsos negativos. Os valores das métricas obtidas pelo *software* estão na Tabela 2.

O Área sob a Curva ROC (AUC), onde ROC *Receiver Operating Characteristic*, é uma métrica usada para avaliar o desempenho de modelos de classificação binária e validar o teste. Ele representa a área sob a curva ROC, sendo um gráfico que mostra a taxa de verdadeiros positivos *Recall* em relação à taxa de falsos positivos.

1. Resposta curva ROC

a) AUC igual a 0,5:

- Indica um modelo que está realizando previsões aleatórias, essencialmente não é melhor do que uma escolha aleatória.

b) AUC entre 0,5 e 1:

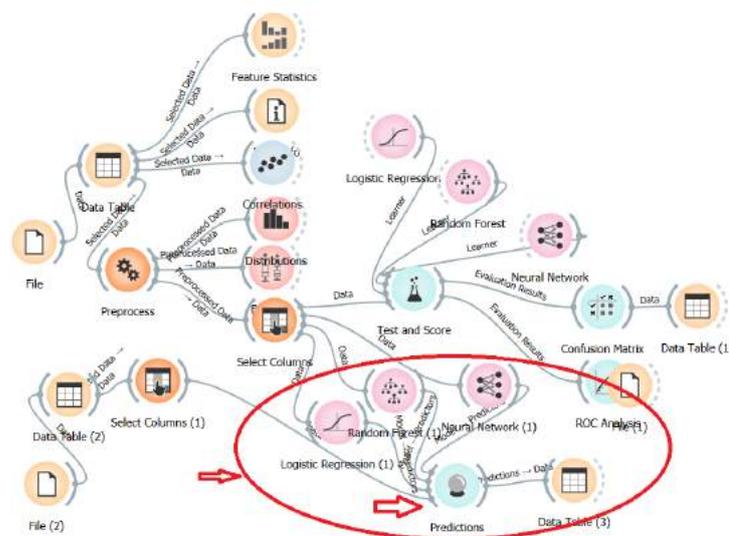
- Indica a capacidade discriminativa do modelo. Quanto mais próximo de 1, melhor o modelo em distinguir entre as classes positiva e negativa. Um valor de 1 indica um modelo perfeito.

c) AUC abaixo de 0,5 :

- Indica que o modelo está pior do que uma escolha aleatória. Isso geralmente ocorre quando o modelo está fazendo previsões invertidas, trocando as classes.

Para ser realizada a predição deve-se adicionar um *widget* chamado “*Predictions*” conforme apontado pela seta na Figura 22, e conectar tanto a planilha teste quanto os modelos de classificação, desse modo o *software* entende que deve tentar prever qual o *target* referente ao dados da linha, o modo de ligação está demonstrado na Figura 22.

Figura 22 – Widget de Predição



Fonte: Autor.

4 RESULTADOS

Através da sustentação teórica apresentada no Capítulo 2 e juntamente com os materiais e métodos do Capítulo 3, este trabalho apresenta os resultados obtidos após a aplicação do modelo de regressão logística binária nos dados coletados. O objetivo é prever a probabilidade de quebra de equipamentos com base na coleta de amostras dos dados de Ressonância Magnética (RM), conforme ilustrado no fluxograma da Figura 10. Os resultados esperados são fundamentados nas discussões realizadas e nas soluções propostas para as questões apresentadas na seção de considerações iniciais seção 1.1.

Para obter resultados mais precisos, os registros *logs* durante esse período, foram geradas 202 mil linhas de registros, detalhando minuto a minuto o funcionamento da criogenia de três máquinas do modelo RM General Electric GE LX 1.5 T.

Os registros das máquinas foram originalmente obtidos em formato de arquivo de texto, sendo uma estrutura de dados sequencial composta por linhas de texto. Posteriormente, esses dados foram convertidos para o formato CSV (Comma-Separated Values), também conhecido como “valores separados por vírgula”. O formato CSV é regulamentado pelo *Request for Comments* 4180 e utiliza vírgulas para separar os valores em uma tabela. A conversão foi realizada com o auxílio do *software* Excel, utilizado para ler e armazenar os dados obtidos.

4.1 Escopo do sistema

O estudo avalia a eficácia da regressão logística na detecção de falhas das RM, utilizando dados coletados por sensores e armazenados nos *dataloggers* das máquinas. O Comparações com outros métodos de aprendizado de máquina são realizados para validar o desempenho do modelo de treinamento. É importante notar que para aplicar este modelo em outros equipamentos, é necessário realizar ajustes tanto no modelo quanto na base de dados, uma vez que a regressão logística foi adaptada especificamente para os dados das ressonâncias GE 1.5 T. O estudo é direcionado a pesquisadores e profissionais interessados em aplicar a regressão logística em diferentes tipos de equipamentos, visando prever falhas com base em dados de sensores e histórico de quebra.

4.1.1 Funcionalidades

a) Coleta de Dados:

- coleta de dados A coleta de dados ocorreu da seguinte maneira; após a medição da variável pelo sensor, essas medições são armazenadas e monitoradas no *datalogger* da máquina por um período de 30 dias, após o qual são auto-

maticamente eliminadas. Para realizar a coleta desses dados, é necessária uma abordagem presencial.

O procedimento envolve a conexão do *notebook* ao *datalogger*, utilizando um cabo *crossover*, que deve ser conectado tanto no *notebook* quanto no *datalogger*. Após a realização das configurações e ajustes necessários e a conclusão da conexão, a coleta de dados segue um procedimento específico. A coleta é realizada incrementalmente devido às restrições do *software*.

Por exemplo, suponhamos que hoje seja 6 de outubro. Nesse caso, a coleta de dados se inicia com os registros mais recentes disponíveis, que remontam em até 168 horas antes da data atual. Portanto, no dia 6 de outubro, é possível coletar dados até o dia 29 de setembro.

Para coletar um período de 30 dias completo, o procedimento é repetido em quatro etapas. Isso significa que, após a primeira coleta, o *software* da GE é ajustado para permitir a coleta a partir do dia 29 de setembro, e assim por diante, até abranger os 30 dias completos.

b) Criação do Banco de Dados

- A criação do banco de dados tem como finalidade armazenar todos os *logs* coletados em um único local e desempenha um papel essencial em garantir que as informações sejam gerenciadas, protegidas, acessadas e utilizadas eficazmente, beneficiando uma ampla gama de operações organizacionais e estratégicas. O *software* utilizado para a criação do banco de dados foi o *software* Excel. Além de sua utilidade na criação do banco de dados, o *software* permitiu a realização das primeiras filtragens dos dados.

Visto que os dados são coletados minuto a minuto, isso acarreta dados repetidos e sem utilidade. Dessa forma, é possível, no *software*, filtrar os dados, deixando apenas os registros a cada meia hora. Isso reduz consideravelmente o tamanho da tabela e permite que o Orange processe os dados com mais rapidez e eficiência.

c) Aplicação do Método de Regressão Logística Binária:

- A escolha do método de regressão logística foi feita devido à sua eficácia como uma técnica de *machine learning* para análise de dados. Essa abordagem utiliza princípios matemáticos e probabilísticos para identificar relações entre diferentes fatores. Essas relações são então aplicadas para prever o valor de um desses fatores com base nas informações dos outros. Geralmente, essa previsão resulta em um número finito de resultados, como 'sim' ou 'não'.

No contexto do estudo, utilizou a regressão logística para determinar se uma máquina está ou não com falha, comparando seus dados com informações semelhantes ao momento da quebra real.

- A adaptação do método de regressão logística binária em um *software* estatístico foi realizada no Orange, utilizado para análise dos dados e cálculo da probabilidade de quebra. Isso permitiu a análise dos dados e a obtenção de estatísticas relacionadas ao *machine learning* do método, bem como a validação dos resultados.

4.1.2 Limitações do sistema

a) Dependência de Dados de Entrada:

- Dado que a coleta de dados é um processo manual, uma das limitações deste projeto é a necessidade de deslocamento até o ambiente de estudo para extrair as informações do *datalogger* da máquina. No entanto, essa etapa de coleta é crucial, uma vez que a aplicação bem-sucedida do modelo de regressão depende inteiramente da qualidade e disponibilidade dos dados da RM obtidos. Como mencionado anteriormente, para enriquecer o banco de dados, é necessário realizar coletas periódicas e contínuas.

b) Limitação ao Modelo de Máquina Específico:

- O sistema foi especificamente projetado para as máquinas de ressonância magnética do modelo GE LX 1.5 T, embora sua aplicação em outros modelos seja viável após ajustes tanto no *software* quanto na abordagem dos dados coletados. A escolha do modelo GE foi motivada pela presença de três exemplares dessa máquina na região oeste do Paraná, o que possibilitou a expansão do banco de dados. Além disso, a existência de um *datalogger*, conforme ilustrado na imagem 12, facilitou a coleta de dados. A combinação de três máquinas e a presença do *datalogger* simplifica o estudo e a extração de medições reais de funcionamento, tanto em condições regulares quanto irregulares da máquina, permitindo, assim, a aplicação bem-sucedida do modelo de regressão logística.

c) Resultados Indicativos:

- Os resultados gerados pelo sistema são indicativos e não devem ser utilizados como diagnóstico definitivo. A aplicação da técnica de regressão logística binária aos dados visa ao cálculo de coeficientes de regressão, distribuições e acurácia, entre outras métricas. Portanto, a geração de gráficos e tabelas

para visualizar os resultados da análise, assim como a interpretação dos resultados da regressão logística binária em termos de sua relevância para a previsão de quebra de equipamentos, constitui aspectos fundamentais deste estudo.

d) Requisitos de Software:

- O sistema exige o uso do *software* Excel para a conversão e armazenamento dos dados em formato CSV. Os dados coletados, originalmente em formato de arquivo de texto, devem ser convertidos para o formato CSV a fim de facilitar a análise e a aplicação no *software* Orange. Além disso, é necessário que o sistema tenha requisitos de memória suficientes para o armazenamento dos dados coletados e, caso seja realizada a criação de um banco de dados adicional, é necessário um processador e memória *RAM* superiores.

4.2 Parametrização e Resultados

Conforme as métricas obtidas no modelo descritas na Tabela 2, o AUC está em 0.833, um valor aceitável, demonstrando que o modelo consegue distinguir as classes de falha 'sim' e 'não'.

Tabela 2 – Test and Score regressão logística

Modelo	AUC	CA	F1	Prec	Recall
Logistic regression	0.833	0.933	0.905	0.934	0.933

Fonte: Autor.

Ao comparar ambas as regularizações no Orange conforme a Figura 21, para determinar qual se adapta melhor ao banco de dados e qual delas apresenta métricas mais favoráveis para o *machine learning*, observa-se na comparação dos resultados que a regularização Lasso obteve uma pequena vantagem sobre a Ridge, como mostrado na Tabela 3.

Tabela 3 – Test and Score regressão logística

Modelo	AUC	CA	F1	Prec	Recall
Logistic regression Lasso	0.833	0.933	0.905	0.934	0.933
Logistic regression Ridge	0.833	0.932	0.905	0.934	0.932

Fonte: Autor.

Para avaliar o desempenho do *machine learning* do método de regressão logística foi utilizado a curva ROC e matriz de confusão, a matriz de confusão é uma tabela, conforme a Figura 23 que permite avaliar o desempenho de um modelo de classificação comparando suas previsões com os resultados reais.

Figura 23 – Matriz de confusão do modelo de regressão logística

		Predicted		Σ
		não	sim	
Actual	não	93620	30	93650
	sim	6790	610	7400
Σ		100410	640	101050

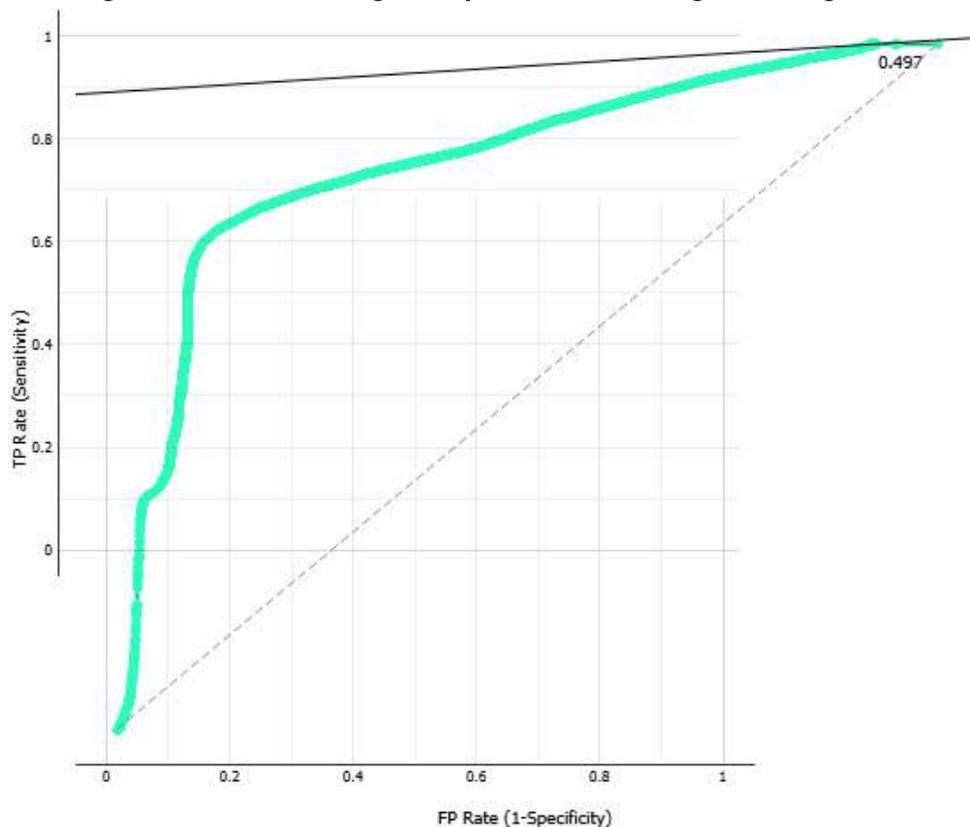
Fonte: Autor.

Observando a matriz de confusão, pode-se notar que o método de regressão logística apresentou alguns erros. No entanto, ao interpretar os resultados e considerar o conhecimento sobre os dados, é possível deduzir que os erros de classificação do modelo foram causados pela inclusão de uma coluna na planilha que registra falhas de *software* ou *hardware* da máquina, não relacionadas à falha na criogenia. Isso não significa que a máquina não tenha parado; a questão de parada devido a problemas de *software* ou *hardware* envolve muito mais do que a parte relacionada à criogenia, como mostrado no Capítulo 2.

A parada pode estar associada a quedas de energia, falta de manutenção preventiva, perturbações elétricas na rede, entre outros fatores elétricos ou mecânicos. Além disso, como pode ser observado, o conjunto denominado RM 'Ressonância Magnética' é composto por diversas partes, como ilustrado nas Figuras 3 e 6, que podem interromper a máquina ou causar falhas. Foi aplicada a classificação de paradas por *software* e *hardware* aos dados em que houve falha nos dias em que a máquina parou, a fim de verificar se havia alguma correlação entre a parada e o aumento de temperatura na criogenia da máquina.

A curva ROC conforme a Figura 24, mostra a taxa de verdadeiros positivos *Recall* em relação à taxa de falsos positivos para diferentes pontos de corte em um modelo de classificação, ajudando a visualizar e medir a capacidade de um modelo de classificação de distinguir entre as classes positiva e negativa, como visto na Tabela 3. Quanto mais próxima à curva ROC estiver do canto superior esquerdo do gráfico, melhor será o desempenho do modelo, a curva ROC do modelo de regressão logística, variando os parâmetros do modelo de regressão como Lasso ou Ridge geraram a mesma curva, e como AUC gerado pelo modelo de regressão logística na Tabela 3 apresentam um valor entre 0.5 e 1 pode se dizer que o método tem capacidade discriminar e distinguir as classes de sim ou não para falha da máquina.

Figura 24 – Curva ROC gerada pelo modelo de regressão logística.



Fonte: Autor.

A matriz de confusão fornece uma visão detalhada das previsões do modelo, enquanto a curva ROC ajuda a avaliar o desempenho geral do modelo em diferentes limiares de decisão. Ambas as ferramentas são essenciais para entender como o modelo de regressão logística de classificação está se saindo em problemas de classificação binária. O aumento do banco de dados relacionado às quebras, principalmente na criogenia da máquina, pode ser usado para ajustar e melhorar o desempenho do modelo conforme necessário.

Como foram selecionadas três máquinas de RM do mesmo modelo General Electric GE LX 1.5 T, e apenas uma apresentou uma quebra na criogenia, como pode-se notar na Figura 25 que nos dias 23 a 26 de julho, as temperaturas do *shield* até o *Cold Head* estão elevadas 163,15K em torno de -110°C , enquanto o funcionamento normal, conforme mostrado na Figura 16, é 4.2 K em torno de -269°C . Fica evidente que, para melhorar o *machine learning* e suas métricas, é necessário expandir o banco de dados, tornando o método mais eficiente. No entanto, mesmo com poucos *logs* de quebra, conforme ilustrado na Figura 25, a regressão conseguiu obter uma alta taxa de acertos, como observado nas métricas de *test and score*.

Figura 25 – logs de quebra do sistema de criogenia

Maquina	Date	Time	He_Level	Water_Temp	Shield_Si	Recon_Ri	Recon_Si	Coldhead	He_Pr	falha softw	Falha
máquina 3	23-Jul-23	08:30	58.74	17,737	144.192	129.267	162.880	119.810	2.268	não	sim
máquina 3	26-Jul-23	04:00	59.96	17,737	145.565	129.267	161.571	118.976	4.760	não	sim
máquina 3	23-Jul-23	08:00	58.71	17,737	142.521	128.387	160.785	118.146	2.156	não	sim
máquina 3	23-Jul-23	07:30	58.71	17,737	140.547	126.638	158.684	116.501	2.052	não	sim
máquina 3	26-Jul-23	03:30	59.96	17,737	143.013	126.638	158.421	116.501	4.643	não	sim
máquina 3	25-Jul-23	03:00	59.25	17,751	139.756	124.906	156.575	115.685	3.575	não	sim
máquina 3	23-Jul-23	07:00	58.65	17,737	138.468	124.906	156.047	114.874	1.948	não	sim
máquina 3	26-Jul-23	03:00	59.93	17,737	140.350	123.191	154.989	114.069	4.521	não	sim
máquina 3	25-Jul-23	02:30	59.25	17,737	137.375	122.339	153.399	113.268	3.471	não	sim
máquina 3	23-Jul-23	06:30	58.65	17,737	136.181	122.339	153.133	112.473	1.849	não	sim
máquina 3	26-Jul-23	02:30	59.93	17,737	137.574	120.649	150.739	110.899	4.399	não	sim
máquina 3	23-Jul-23	06:00	58.65	17,737	133.483	118.976	149.405	110.121	1.746	não	sim
máquina 3	25-Jul-23	02:00	59.22	17,751	134.684	119.810	149.672	110.121	3.359	não	sim
máquina 3	26-Jul-23	02:00	59.87	17,737	134.184	115.685	145.383	107.071	4.282	não	sim
máquina 3	23-Jul-23	05:30	58.65	17,737	130.268	115.685	144.575	106.325	1.651	não	sim
máquina 3	25-Jul-23	01:30	59.22	17,751	131.376	115.685	144.575	106.325	3.255	não	sim
máquina 3	23-Jul-23	05:00	58.66	17,737	126.423	110.121	138.342	102.008	1.556	não	sim
máquina 3	25-Jul-23	03:30	59.45	17,723	131.275	110.899	139.159	102.008	3.678	não	sim
máquina 3	26-Jul-23	01:30	59.87	17,737	129.865	110.121	138.070	102.008	4.170	não	sim

Fonte: Autor.

O uso do *machine learning*, especificamente a regressão logística, tem benefícios significativos para a previsão de falhas na máquina em comparação com métodos tradicionais de manutenção preventiva. Enquanto os métodos tradicionais muitas vezes dependem de agendamentos fixos de manutenção ou de análises periódicas baseadas em intervalos de tempo predefinidos, o *machine learning* permite uma abordagem mais adaptativa e orientada por dados.

O modelo de regressão logística analisa os dados coletados, identificando padrões e correlações que podem ser sutis ou não evidentes para os métodos convencionais, análises de dados resultam em previsões mais precisas e oportunas, permitindo que a manutenção seja realizada com base na real necessidade, tornando o *machine learning* uma ferramenta valiosa para melhorar a confiabilidade e a disponibilidade de equipamentos críticos.

Figura 26 – Software acertando as falhas

Date	Time	a(Logistic Regress	p(não)	p(sim)	Selected
23-Jul-23	04:30:00	sim	0	1	No
24-Jul-23	23:30:00	sim	7.75358e-12	1	No
23-Jul-23	06:30:00	sim	0	1	No
24-Jul-23	23:00:00	sim	1.63602e-05	0.999984	No
26-Jul-23	05:00:00	sim	0.450425	0.549575	No
25-Jul-23	02:30:00	sim	0	1	No
25-Jul-23	03:00:00	sim	0	1	No
24-Jul-23	22:00:00	sim	0.0149996	0.985	No
26-Jul-23	03:30:00	sim	0	1	No
26-Jul-23	02:00:00	sim	0	1	No
23-Jul-23	04:00:00	sim	4.44089e-16	1	No
26-Jul-23	08:00:00	sim	0.0218257	0.978174	No
26-Jul-23	04:30:00	sim	2.22045e-16	1	No
25-Jul-23	00:00:00	sim	2.22045e-15	1	No
23-Jul-23	09:00:00	sim	0	1	No
23-Jul-23	07:00:00	sim	0	1	No
23-Jul-23	02:00:00	sim	0.000140481	0.99986	No

Fonte: Autor.

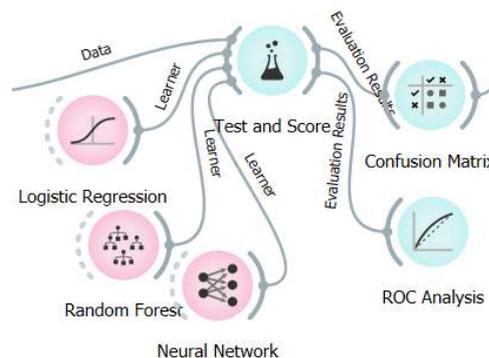
Como as máquinas de ressonância magnética, ao mesmo tempo, em que reduz o risco de paradas inesperadas. Na Figura 26 observa-se o modelo de regressão logística acertando corretamente à falha da máquina 3, por conta *machine learning*.

4.3 Resultados obtidos e comparação com outros modelos

No *software* Orange, além do método de *regressão logística*, existem outros métodos de *machine learning* que foram utilizados para comparação conforme a Figura 27 e as métricas obtidas estão na Tabela 4, a fim de verificar se outras abordagens apresentam métricas melhores do que o modelo de *regressão logística*. Os métodos aplicados foram o *Random Forest* e *Neural Networks*.

Esses modelos podem ser empregados tanto em tarefas de classificação quanto de regressão, sendo particularmente adequados para lidar com problemas complexos e não lineares. No entanto, as redes neurais tendem a ser menos interpretáveis em comparação com a Regressão Logística.

Figura 27 – Aplicação dos modelos nos dados.



Fonte: Autor.

Pode-se observar que as métricas do modelo Random Forest se destacou melhor perante aos outros métodos na Tabela 4.

Tabela 4 – Test and Score dos métodos de machine learning

Modelo	AUC	CA	F1	Prec	Recall
Logistic regression	0.833	0.933	0.905	0.934	0.933
Random Forest	0.951	0.939	0.935	0.931	0.939
Neural Network	0.904	0.933	0.907	0.931	0.933

Fonte: Autor.

Principalmente AUC, métrica usada para avaliar o desempenho dos modelos de classificação binária e validar o *machine learning*. A curva ROC da Figura 28 comparando os 3 modelos mostra que o modelo *Random Forest* erra menos nas predições. Durante o processamento dos dados, observou que o modelo de *neural network* levou um tempo considerável

a mais conforme a Tabela 5 que os outros dois modelos no processamento dos dados e não apresentou métricas superiores.

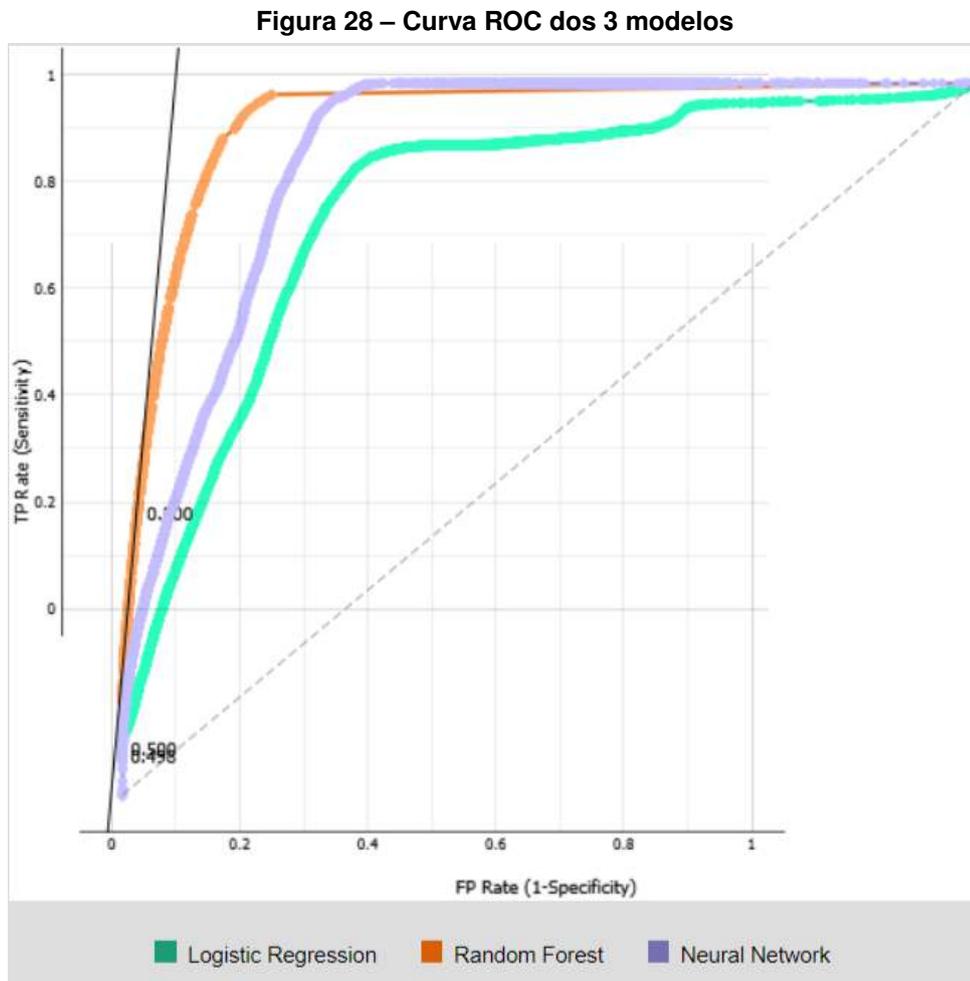


Tabela 5 – Tempo(segundos) do treinamento e teste dos modelos de *machine learning*

Modelo	Train	Test
Logistic regression	15,720	0.294
Random Forest	4.546	0.960
Neural Network	1256.86	1.379

Fonte: Autor.

Os tempos de treinamento de cada modelo variam conforme a utilização e variação dos parâmetros definidos no capítulo Capítulo 3, conforme a Figura 21 os parâmetros utilizados para regressão logística, na Figura 14 as utilizadas no modelo *Random Forest* e por último a Figura 15 que mostra as configurações utilizadas para a *Neural Networks*. Os tempos de treinamento podem variar conforme o banco de dados e a adaptação do modelo de *machine learning* sobre o banco de dados.

Conforme mencionado anteriormente no Capítulo 2, a regressão logística binária representa o caso em que a variável dependente Y é dicotômica ou binária, admitindo apenas duas categorias, 1 ou 0 (sim ou não). Partindo da função do *logit*, define-se a probabilidade do evento conforme a equação (8). Ao expressar a equação de z (9) em termos da equação (8), a equação que descreve a probabilidade do evento acontecer, conforme mostrado em (10).

Os coeficientes derivados para o modelo de regressão logística são denotados como α, β_1, β_2 até β_k , obtidos por meio da aplicação do método de máxima verossimilhança. A Tabela 6 apresenta os coeficientes resultantes com base nos dados coletados por meio do *datalogger*.

Tabela 6 – Coeficientes z regressão logística

nome	sim
alpha	19,7478
x1	-0,169043
x2	-0,732519
x3	0,0362
x4	0,0801501
x5	1,12779
x6	-1,15794
x7	-0,60707

Fonte: Autor.

Aplicando os valores dos coeficientes obtidos no *software* Orange da Tabela 6 na equação (9), obtêm-se os coeficientes através do método de máxima verossimilhança conforme a equação (19), aplicando o *logit(z)* na equação (8) consegue-se a equação do evento acontecer, ou seja, o modelo de regressão logística conforme a equação (20).

$$Z = -19.7478 - 0.1690 * HeL - 0.73254 * Water + 0.0362 * Shield + 0.080 * recon_r + 1.12 * recon_s - 1.15 * ColdH - 0.60 * HeP \quad (19)$$

$$P = \frac{1}{1 + e^{-Z}} \quad (20)$$

A validação e teste final do método de regressão logística consiste em adicionar um pequeno banco de dados “desconhecido” contendo dados de quebra e não quebra conforme a Figura 29, a fim de verificar se o *machine learning* pode prever se a máquina está funcionando ou apresentando falhas. Esse pequeno banco de dados se concentra principalmente nas falhas no sistema de criogenia da máquina.

Na fase de treinamento observou que todos os métodos obtiveram métricas aceitáveis como AUC, acurácia e precisão conforme a Tabela 4, ainda na Tabela 4 observa-se que na fase

de treinamento tanto o *Random Forest* e a *Neural Network* foram melhores que a regressão logística em quase todas métricas, contudo na Figura 26 nota-se que a regressão logística na fase de treinamento acertou todos os dias em que a máquina apresentou erro de criogenia.

Figura 29 – Planilha teste.

Maquina	Date	Ti	He	Water	hield	Reco	ecor	dhe	le_Pr	h	falh
maquina 1	31-Jul-23	03:30	78.29	17,764	39.596	4.173	4.449	4.253	1.043	0	não
maquina 1	31-Jul-24	07:30	78.23	17,764	70.721	4.365	4.449	4.842	0.953	0	sim
maquina 1	31-Jul-25	15:00	75.23	17,764	39.845	4.170	4.415	4.242	1.047	0	não
maquina 1	31-Jul-26	00:30	78.20	17,764	39.471	4.163	4.437	4.237	0.917	0	não
maquina 1	31-Jul-27	02:00	60.20	17,764	39.471	4.173	4.449	4.237	1.038	0	não
maquina 1	31-Jul-28	10:30	78.20	17,764	39.845	4.163	4.437	4.237	0.962	0	não
maquina 1	31-Jul-29	23:30	59.11	17,764	39.721	4.158	4.404	4.237	0.912	0	não
maquina 1	31-Jul-30	08:30	78.23	17,764	39.721	4.165	4.381	4.235	0.957	0	não
maquina 1	31-Jul-31	19:00	78.24	17,764	39.845	4.163	4.415	4.232	0.953	0	não
maquina 1	31-Jul-32	10:30	60.41	17,737	52.419	4.146	3.692	4.656	4.670	0	sim
maquina 1	31-Jul-33	13:00	60.46	17,737	69.691	4.393	4.692	4.853	4.701	0	sim
maquina 1	31-Jul-34	21:00	60.64	17,737	45.537	4.385	3.672	4.353	4.647	0	não
maquina 1	31-Jul-35	02:30	60.62	17,737	39.471	4.382	3.632	4.350	4.562	0	não
maquina 1	31-Jul-36	08:30	60.55	17,737	38.972	4.369	3.622	4.350	4.440	0	não
maquina 1	31-Jul-37	14:00	60.70	17,751	41.338	4.393	3.652	4.350	4.796	0	não
maquina 1	31-Jul-38	13:30	60.53	17,737	49.083	4.387	3.682	4.350	4.692	0	sim

Fonte: Autor.

Para realizar as previsões, o primeiro passo é criar uma planilha de teste, conforme exemplificado na Figura 29. Essa planilha deve conter dados aleatórios que representem possíveis quebras e dados de funcionamento normal da máquina, com uma coluna indicando “falha” no canto lateral direito, tal coluna é retirada antes do *upload* no Orange. Na ilustração 29, essa coluna foi mantida para permitir que o leitor possa comparar e validar os modelos de *machine learning*.

Após a inserção da Tabela 29 no *software*, é realizada uma comparação entre os três modelos para avaliar suas capacidades de predição conforme a Figura 22, onde é possível observar o *widget predictions* e os 3 modelos ligados a ele. Os resultados obtidos podem ser visualizados na Figura 30.

Comparando a Figura 30 com a coluna “falha” da Figura 29 é evidente que o modelo de regressão logística alcançou um desempenho superior, acertando todas as previsões, referentes a planilha de teste 29.

Figura 30 – Resultado da predição.

Logistic Regression (1)	Random Forest (1)	Neural Network	aqui	Date	Tir	falha	He_L	ater	ield_	lecon	con_	jhea	e_Pre
1.00 : 0.00 → não	1.00 : 0.00 → não	1.00 : 0.00 → não	m...	31-Jul-23	03...	?	78...	17...	39...	4.1...	4.4...	4.2...	1.043
0.00 : 1.00 → sim	0.82 : 0.18 → não	0.64 : 0.36 → não	m...	31-Jul-24	07...	?	78...	17...	70...	4.3...	4.4...	4.8...	0.953
0.99 : 0.01 → não	1.00 : 0.00 → não	1.00 : 0.00 → não	m...	31-Jul-25	15...	?	75...	17...	39...	4.1...	4.4...	4.2...	1.047
1.00 : 0.00 → não	1.00 : 0.00 → não	1.00 : 0.00 → não	m...	31-Jul-26	00...	?	78...	17...	39...	4.1...	4.4...	4.2...	0.917
0.94 : 0.06 → não	1.00 : 0.00 → não	0.99 : 0.01 → não	m...	31-Jul-27	02...	?	60...	17...	39...	4.1...	4.4...	4.2...	1.038
1.00 : 0.00 → não	1.00 : 0.00 → não	1.00 : 0.00 → não	m...	31-Jul-28	10...	?	78...	17...	39...	4.1...	4.4...	4.2...	0.962
0.93 : 0.07 → não	1.00 : 0.00 → não	0.99 : 0.01 → não	m...	31-Jul-29	23...	?	59...	17...	39...	4.1...	4.4...	4.2...	0.912
1.00 : 0.00 → não	1.00 : 0.00 → não	1.00 : 0.00 → não	m...	31-Jul-30	08...	?	78...	17...	39...	4.1...	4.3...	4.2...	0.957
1.00 : 0.00 → não	1.00 : 0.00 → não	1.00 : 0.00 → não	m...	31-Jul-31	19...	?	78...	17...	39...	4.1...	4.4...	4.2...	0.953
0.00 : 1.00 → sim	0.00 : 1.00 → sim	0.64 : 0.36 → não	m...	31-Jul-32	10...	?	60...	17...	52...	4.1...	3.6...	4.6...	4.670
0.00 : 1.00 → sim	0.00 : 1.00 → sim	0.54 : 0.46 → não	m...	31-Jul-33	13...	?	60...	17...	69...	4.3...	4.6...	4.8...	4.701
0.99 : 0.01 → não	1.00 : 0.00 → não	0.80 : 0.20 → não	m...	31-Jul-34	21...	?	60...	17...	45...	4.3...	3.6...	4.3...	4.647
0.99 : 0.01 → não	1.00 : 0.00 → não	0.94 : 0.06 → não	m...	31-Jul-35	02...	?	60...	17...	39...	4.3...	3.6...	4.3...	4.562
0.99 : 0.01 → não	1.00 : 0.00 → não	0.95 : 0.05 → não	m...	31-Jul-36	08...	?	60...	17...	38...	4.3...	3.6...	4.3...	4.440
1.00 : 0.00 → não	1.00 : 0.00 → não	0.89 : 0.11 → não	m...	31-Jul-37	14...	?	60...	17...	41...	4.3...	3.6...	4.3...	4.796
0.00 : 1.00 → sim	0.00 : 1.00 → sim	0.74 : 0.26 → não	m...	31-Jul-38	13...	?	60...	17...	49...	4.3...	3.6...	4.3...	4.692

Fonte: Autor.

Por outro lado, o modelo de *neural network* apresentou o pior desempenho, falhando em todas as previsões de 'sim' para falha. Surpreendentemente, o modelo *Random Forest*, apesar de suas métricas estatísticas promissoras, errou uma previsão 'sim' para falha. Com base nos resultados da previsão na Figura 30, pode-se concluir que, para os dados deste estudo e as configurações adotadas, o modelo de regressão logística se destacou como a melhor escolha, mesmo não apresentando as métricas mais elevadas no *test and score*.

5 CONCLUSÃO

O trabalho abordou a aplicação de um modelo de regressão logística binária para prever a probabilidade de quebra em máquinas de ressonância magnética RM com base em dados de sensores e histórico de quebras. O objetivo geral foi prever possíveis falhas e adquirir um modelo de manutenção preditiva para reduzir paradas indesejadas nas máquinas GE LX 1.5 T. Para alcançar esse objetivo, foram estabelecidos objetivos específicos, incluindo a coleta de dados de sensores e histórico de quebras, tratamento de erros e formatação do banco de dados e amostra, escolha do modelo ideal conforme a amostra de dados coletados e a validação do modelo com o banco de dados completo.

Durante o desenvolvimento do trabalho, foram destacadas as vantagens da regressão logística binária como uma poderosa ferramenta estatística para lidar com dados complexos e variáveis múltiplas. A importância do uso de materiais adequados, como os banco de dados, planilhas eletrônicas e *software* estatístico, foi enfatizada na viabilização da pesquisa.

As dificuldades encontradas incluíram a coleta de dados de sensores e histórico de quebras, que exigiram visitas às clínicas com as máquinas, bem como a necessidade de ampliar o banco de dados, a fim de aumentar a confiabilidade do modelo.

Em relação à contribuição para a comunidade acadêmica e profissional, este trabalho propõe um método eficaz de manutenção preditiva para máquinas de RM, que pode levar à otimização dos recursos e à maximização do tempo de operação das máquinas. A aplicação da regressão logística binária neste contexto oferece uma abordagem para prevenir paradas inesperadas e reduzir custos. O modelo foi validado em comparação com outros modelos e por meio de experimentos utilizando dados simulados, porém condizentes com a realidade.

Para estudos futuros, recomenda-se a expansão da base de dados, coletando informações mais abrangentes sobre falhas em máquinas do mesmo modelo GE em uma variedade de condições. Isso contribuiria significativamente para fortalecer a confiabilidade do método. Além disso, para pesquisas posteriores, seria valioso ampliar as medições dos sensores para englobar todo o equipamento, incluindo variáveis como a tensão de entrada e a corrente de entrada, além do escopo atual da análise, que se limita à criogenia.

REFERÊNCIAS

- ALMEIDA, P. S.; ROCCA, J. E. **Manutenção mecânica industrial princípios técnicos e operações**. São Paulo: Saraiva Educação SA, 2015.
- ALMEIDA, P. S. D. **Manutenção Mecânica Industrial Conceitos Básicos e Tecnologia Aplicada**. São Paulo: Saraiva Educação SA, 2018.
- ARCURI, S. M.; MCGUIRE, P. K. Ressonância magnética funcional e sua contribuição para o estudo da cognição em esquizofrenia. **Brazilian Journal of Psychiatry**, São Paulo, v. 23, p. 38–41, maio. 2001.
- ARMSTRONG, P.; KEEVIL, S. F. Magnetic resonance imaging–1: Basic principles of image production. **BMJ: British Medical Journal**, Londres, v. 303, n. 6793, p. 35, jul. 1991.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 5462**: Confiabilidade e manutenibilidade. Rio de Janeiro, 1994.
- BATHISTA, A. Princípios básicos de ressonância magnética nuclear do estado sólido. **Instituto de Física de São Carlos - USP**, São Paulo, v. 1, p. 49, jul. 2005.
- BORLIDO, D. J. A. **Indústria 4.0: aplicação a sistemas de manutenção**. 2017. Dissertação (Mestrado) — Faculdade de Engenharia Universidade do Porto, Portugal, 2017.
- CERQUEIRA, E. O. d. *et al.* Redes neurais e suas aplicações em calibração multivariada. **Química Nova**, Franca, v. 24, p. 864–873, mar. 2001.
- CHIU, H.-N.; HUANG, B.-S. The economic design of x-control charts under a preventive maintenance policy. **International Journal of Quality & Reliability Management**, Taiwan, Republic of China, v. 13, n. 1, p. 61–71, feb. 1996.
- DAWSON, M. J. **Paul Lauterbur and the Invention of MRI**. Massachusetts, EUA: MIT Press, 2013.
- FAVERO, L.; FÁVERO, P. **Estatística aplicada: Para cursos de Administração, Contabilidade e Economia com Excel e SPSS**. São Paulo: Elsevier Brasil, 2015.
- FIGUEIRA, C. V. **Modelos de regressão logística**. 2006. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, 2006.
- GONZALEZ, L. d. A. **Regressão logística e suas aplicações**. 2018. Dissertação (Monografia Graduação em Engenharia de Ciência da Computação) — Universidade Federal do Maranhão, São Luís, 2018.
- HAGE, M.; IWASAKI, M. Imagem por ressonância magnética: princípios básicos. **Ciência Rural**, São Paulo, v. 39, p. 1275–1283, dez. 2009. Disponível em: <https://doi.org/10.1590/S0103-84782009005000041>. Acesso em: 16 nov. 2023.
- HARRELL, F. E. *et al.* **Regression modeling strategies**: with applications to linear models, logistic regression, and survival analysis. Berlim: Springer, 2001.
- HIVAREKAR, N. *et al.* Preventive and predictive maintenance modeling. **2020 Annual Reliability and Maintainability Symposium**, IEEE, Palm Springs, CA, p. 1–6, jul. 2020.

- JARDINE, A. K.; LIN, D.; BANJEVIC, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. **Mechanical Systems and Signal Processing**, Toronto, v. 20, n. 7, p. 1483–1510, 2006.
- MARCORIN, W. R.; LIMA, C. R. C. Análise dos custos de manutenção e de não-manutenção de equipamentos produtivos. **Revista de ciência & tecnologia**, Santa Bárbara d' Oeste, v. 11, n. 22, p. 35–42, 2003.
- MARINS, M. A. **Classificação de Falhas em Máquinas Rotativas Utilizando Métodos de Similaridade e Random Forest**. 2016. Dissertação (Monografia Graduação em Engenharia Eletrônica e de Computação) — Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2016.
- MASUYAMA, S.; NUMAZAWA, T. Characteristics of a 1.6 w gifford-mcmahon cryocooler with a double pipe regenerator. **Conf. Ser.: Mater. Sci. Eng.**, IOP Publishing Ltd, Madison USA, v. 278, n. 1, p. 12041, 2017.
- MAULAZ, C. M. *et al.* Estudo comparativo do desempenho de imagens por ressonância magnética, mamografia e ecografia na avaliação de lesões mamárias benignas e malignas. **Revista brasileira de física médica**, Núcleo de Pesquisa em Imagens Médicas PUCRS, Porto Alegre RS, n. 1, p. 1–7, 2018.
- MAZZOLA, A. A. Ressonância magnética: princípios de formação da imagem e aplicações em imagem funcional. **Revista brasileira de física médica**, Núcleo de Pesquisa em Imagens Médicas PUCRS, Porto Alegre RS, v. 3, n. 1, p. 117–129, 2009.
- MENDONÇA, M.; NÉIA, J. M.; BRINHOLE, E. R. Redes neurais aplicadas em simulação de circuito eletrônicos. **UNOPAR Científica Ciências Exatas e Tecnológicas**, Paraná, v. 1, n. 1, p. 1–4, 2002.
- MINGUETTI, G. Ressonância magnética na esclerose múltipla: análise de 270 casos. **Arquivos de Neuro-Psiquiatria**, São Paulo, v. 59, p. 563–569, 2001. Disponível em: <https://doi.org/10.1590/S0004-282X2001000400015>. Acesso em: 16 nov. 2023.
- MIRSHAWKA, V.; OLMEDO, N. L. **Manutenção-combate aos custos da não eficácia: A vez do Brasil**. São Paulo: Makron Books, 1993.
- MOBLEY, R. **An introduction to predictive maintenance - 2.ed.** Washington, D.C: Butterworth-Heinemann, 2002.
- MURTY, A.; NAIKAN, V. Availability and maintenance cost optimization of a production plant. **International Journal of Quality & Reliability Management**, MCB UP Ltd, Kharagpur, Índia, v. 12, n. 2, 1995. Disponível em: <https://doi.org/10.1108/02656719510080596>. Acesso em: 16 nov. 2023.
- NASCIMENTO, J. S. d. **Comparação dos algoritmos Random Forest, Random Tree e J48 para detectar ataques DDoS**. 2022. Dissertação (B.S. thesis) — Universidade Tecnológica Federal do Paraná, Ponta Grossa, 2022.
- NATUKUNDA, F. *et al.* Approaches in cooling of resistive coil-based low-field magnetic resonance imaging (mri) systems for application in low resource settings. **BMC Biomedical Engineering**, Springer, Pensilvânia, EUA, v. 3, p. 1–11, 2021. Disponível em: <https://doi.org/10.1186/s42490-021-00048-6>. Acesso em: 16 nov. 2023.
- NEPOMUCENO, L. X. **Técnicas de manutenção preditiva**. São Paulo: Editora Blucher, 2014. v. 1.

PAOLANTI, M. *et al.* Machine learning approach for predictive maintenance in industry 4.0. **ASME International Conference on Mechatronic and Embedded Systems and Applications**, IEEE, Uppsala, Suécia, p. 1–6, 2018. Disponível em: [10.1109/MESA.2018.8449150](https://doi.org/10.1109/MESA.2018.8449150). Acesso em: 16 nov. 2023.

PINTO, R. I. P. **Segmentação Automática de Lesões de Esclerose Múltipla em Imagens de Ressonância Magnética**. 2017. Dissertação (Mestrado) — Faculdade de Engenharia Universidade do Porto, Porto, Portugal, 2017. Disponível em: <https://hdl.handle.net/10216/105286>. Acesso em: 16 nov. 2023.

PIRES, C. A.; OKADA, R. H. Manutenção preditiva: estratégia de produção e redução de custos. **Revista Interface Tecnológica**, Taquaritinga, SP, v. 17, n. 1, p. 635–647, 2020. Disponível em: <https://doi.org/10.31510/infa.v17i1.781>. Acesso em: 16 nov. 2023.

POSSES, F. P. das. Ressonância magnética (rmn): o que é, tipos e como funciona. STAR DIAGNOSTICOS LTDA. 2023. Disponível em: <https://star.med.br/o-que-e-ressonancia-magnetica-rmn/>. Acesso em: 16 nov. 2023.

RIBEIRO, H. S. **Ressonância Magnética e seu Funcionamento Mecânico para Gerar Imagens**. 2018. Dissertação (Monografia Graduação em Engenharia Mecânica) — Fundação de Ensino e Pesquisa do Sul de Minas, Varginha, MG, 2018. Disponível em: <http://repositorio.unis.edu.br/handle/prefix/588>. Acesso em: 16 nov. 2023.

SANTIAGO, J. S.; SILVA, C. A. T. Análise da influência tecnológica sobre a formação do custo do serviço de ressonância magnética em entidades hospitalares: Um estudo comparativo. **Anais Do Congresso Brasileiro De Custos - ABC**, João Pessoa, PB, 2007. Disponível em: <https://anaiscbc.abcustos.org.br/anais/article/view/1636>. Acesso em: 16 nov. 2023.

SCHETTINI, L. **Estudo dos artefatos da imagem diagnóstica gerada em aparelhos ressonância magnética**. 2016. Dissertação (Monografia Graduação em Engenharia Elétrica) — Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, MG, 2016.

SILVA, L. F. D. *et al.* Estudos sobre a manutenção preventiva e preditiva: História e perspectivas para indústria brasileira. **Anais do SIMPÓSIO NACIONAL DE CIÊNCIAS E ENGENHARIAS (SINACEN)**, Anápolis, GO, v. 5, n. 1, p. 94–111, 2020.

SOUZA, R. d. Q. **Metodologia e desenvolvimento de um sistema de manutenção preditiva visando a melhoria da confiabilidade de ativos de usinas hidrelétricas**. 2006. Dissertação (Mestrado) — Universidade DE Brasília, Distrito Federal, 2006. Disponível em: <https://repositorio.unb.br/handle/10482/3635>. Acesso em: 16 nov. 2023.