



ConBRepro

XIII CONGRESSO BRASILEIRO DE ENGENHARIA DE PRODUÇÃO



IA nas Engenharias

29 nov. a 01
de dezembro 2023

Aplicação da Regressão Logística Binária para Manutenção Preditiva em Máquinas de Ressonância Magnética

Carlos Eduardo Barbosa Miranda

Coordenação de Engenharia Eletrônica – UTFPR Campus Toledo

Marcos Roberto Bombacini

Coordenação de Engenharia Eletrônica – UTFPR Campus Toledo

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 1.5 T. Dado o contexto competitivo e os crescentes custos associados a falhas de 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. Para validar a capacidade do modelo de regressão logística no banco de dados utilizado, além da coleta de dados realizada, utilizou-se o Microsoft Excel para a construção do banco de dados e o software Orange para a aplicação do modelo de regressão e para a validação do modelo, com a análise de métricas como a curva ROC e a matriz de confusão. 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 contribui para a eficiência e confiabilidade dos serviços de Ressonância Magnética.

Palavras-chave: Indústria 4.0, Aprendizagem de Máquina Supervisionada, Modelos Preditivos, Manutenção Preditiva.

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

Abstract: This study focuses on the application of logistic regression as a machine learning technique to enhance the maintenance of General Electric 1.5 T Magnetic Resonance Imaging machines. Given the competitive context and the increasing costs 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 work were to identify the tools and theoretical references related to binary logistic regression and its methods, to perform the collection of data from temperature, pressure, and flow sensors through a datalogger installed in the machine. To validate the capability of the logistic regression model in the utilized database, in addition to the data collection conducted, Microsoft Excel was used for constructing the database

and the software Orange for the application of the regression model and for model validation, with the analysis of metrics such as the ROC curve and the confusion matrix. The results demonstrated the effectiveness 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: Industry 4.0, Supervised Machine Learning, Predictive Maintenance, Magnetic Resonance.

1. Introdução

A ressonância magnética é uma técnica avançada de diagnóstico por imagem 2D-3D. 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, 2001).

As primeiras máquinas utilizavam ímãs permanentes ou ímãs resistivos, com o desenvolvimento de ímãs mais potentes como os ímãs supercondutores e juntamente com a evolução de gradientes mais eficientes, proporcionaram uma relação sinal-ruído superiores, níveis de campo mais elevados, contribuindo para a melhoria da resolução e da qualidade das imagens. Isso permitiu a visualização detalhada de estruturas internas do corpo humano.

Para garantir o funcionamento ideal da máquina, a criogenia é fundamental para manter a estabilidade do núcleo do magneto. Devido às suas propriedades supercondutoras, é crucial manter a temperatura em torno de $-269,0^{\circ}\text{C}$ para permitir que correntes elevadas fluam pelas bobinas e gerem o campo magnético necessário para a produção de imagens. O Cold Head desempenha um papel crucial ao garantir a estabilidade do núcleo do magneto a $-269,0^{\circ}\text{C}$, minimizando o consumo de hélio líquido. Dado o funcionamento contínuo da criogenia ao longo do ano, é comum que o Cold Head se desgaste ao longo do tempo, reduzindo sua vida útil e, conseqüentemente, exigindo substituições conforme necessário.

O modelo da máquina em que foi realizada a análise de dados é uma Ressonância Magnética General Eletric LX 1.5 Tesla, conforme a figura 1. 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 1 – Ressonância Magnética General Eletric LX 1.5 Tesla



Fonte: Autor.

Com o intuito de obter um modelo preditivo a partir da análise de um banco de dados de sensores da máquina - dados esses recolhidos dos datalogger - utilizando modelo de regressão logístico binário através do software Orange Canvas, para prever a probabilidade de quebra e assim fornece um modelo preditivo para auxiliar no cronograma de manutenções de uma máquina de ressonância magnética General Eletric GE LX 1.5 T.

O estudo envolveu a coleta de dados de sensores de temperatura e pressão por meio de um datalogger da máquina, seguido pela definição de um modelo matemático para o tratamento dos dados e a validação do modelo com o banco de dados completo para prever a probabilidade de falhas. Além disso, o artigo inclui uma seção de análise dos materiais e métodos, abordando os softwares e ferramentas utilizados, seguida por um referencial teórico que explora a aplicação da regressão logística. Por fim, os resultados obtidos após a aplicação do modelo de regressão nos dados coletados da Ressonância Magnética foram apresentados.

2. Materiais e Métodos

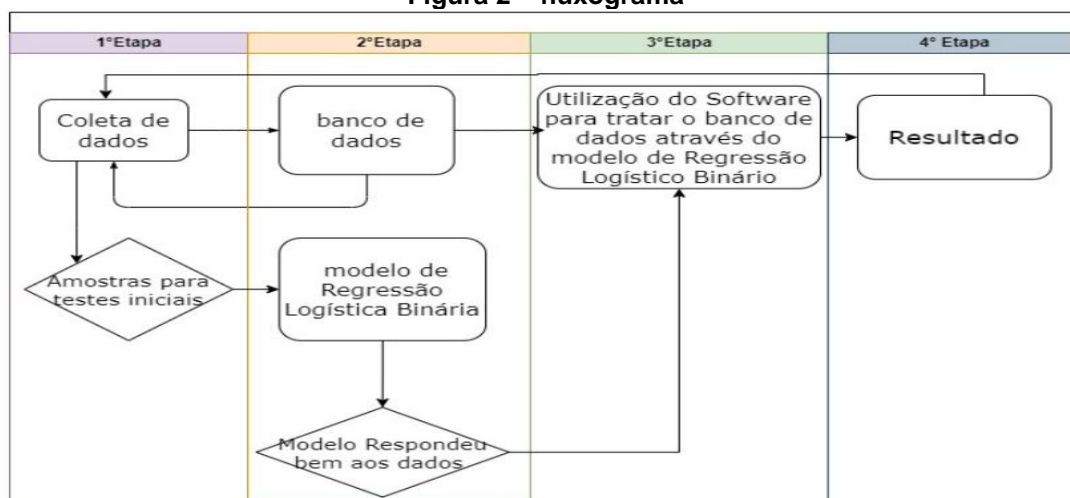
A manutenção preditiva desempenha um papel crucial na eficiência e no funcionamento adequado das máquinas (MARCORIN; LIMA, 2003). A análise dos dados possibilitou a predição de uma possível falha e conseqüentemente uma intervenção de manutenção preditiva mais eficiente e a redução de custos com manutenções preventivas e corretivas (MARCORIN; LIMA, 2003; NEPOMUCENO, 2014).

Para atingir os objetivos durante o trabalho foi utilizado um fluxograma conforme a figura 2, para auxiliar nas atividades e conclusão dos passos a serem realizados.

Os dados coletados da RM são referentes ao datalogger das máquinas, dentro desse datalogger estão todas as medidas dos sensores da parte de criogenia conforme a tabela 1, que ficam armazenados em um mesmo local.

As principais ferramentas para elaborar o trabalho conforme o fluxograma da figura 2, foram a disponibilidade de acesso as máquinas General Eletric LX 1.5 T (três unidades) para coleta de dados, computador com memória disponível para a aquisição e armazenamento de dados, planilha eletrônica (ou software similar) para criação do banco de dados e o software utilizado para a aplicação do método de regressão logística binária foi o Orange Canvas.

Figura 2 – fluxograma



Fonte: Autor

Tabela 1 – Dados coletados do datalogger sistema de ressonância magnética GE LX 1.5T

Item	M1	M2	M3
Nível de Hélio	82.46%	60.66%	78.29%
Pressão de Hélio	2.66	4.35	1.44
Temperatura(k) Cold Head	4.27	4.35	4.28
Temperatura(k) do Shield	41.34	43.69	40.95
Temperatura(°C) água Chiller	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

Fonte: Autor

O computador desempenhou um papel crucial nessa pesquisa, servindo como a principal ferramenta para coletar, armazenar e organizar os dados. Além disso, foi usado para hospedar os softwares essenciais para implementar a regressão logística binária, bem como a planilha eletrônica que desempenhou um papel vital na organização dos dados e na implementação do modelo. As visitas às clínicas para ampliar o banco de dados, a aplicação do modelo, a comparação com resultados de falhas anteriores e a análise das métricas também foram etapas fundamentais para garantir a confiabilidade do modelo de predição. O objetivo final foi desenvolver um modelo de manutenção preditiva eficaz utilizando o método da regressão logística binária.

2.1 Implantação do sistema

Inicialmente, para aplicar o modelo, é essencial converter o log do datalogger em um formato CSV. Posteriormente, o arquivo filtrado é carregado no software Orange. Durante vários testes e observações, percebeu-se que filtrar a planilha no Excel antes de carregá-la no Orange seria benéfico. Assim, decidiu-se manter apenas as linhas com valores registrados a cada meia hora, reduzindo o tamanho do arquivo de 15.728.640 bytes para 393.216 bytes.

No Orange, a importação da planilha é feita através do widget "file". O software oferece um menu lateral que contém todas as ferramentas necessárias. É crucial definir o tipo de variável para cada coluna, incluindo a variável categórica de interesse, que é especificada como o alvo ("target"): 'sim' para falha e 'não' para operação normal da máquina conforme a figura 3.

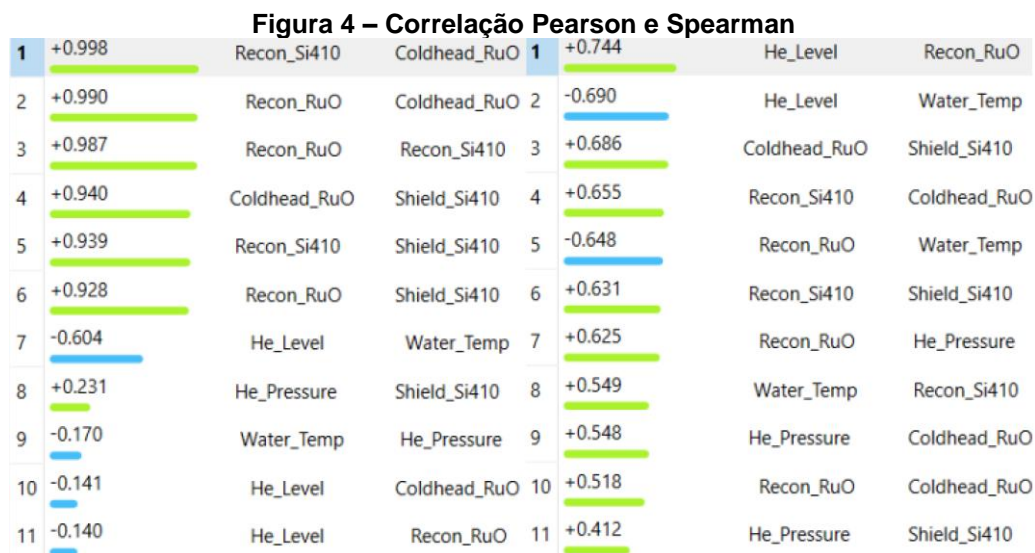
Figura 3 – definição das categorias das variáveis

	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

Com uma análise mais detalhada usando o software Orange, é possível identificar a correlação entre essas medições conforme a figura 4. A correlação de Pearson 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. A correlação de Spearman 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 houve a falha na máquina.

Observando a figura 4, nota-se que o Cold Head 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. 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.



Fonte: Autor

Adicionando o widget da regressão, 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. 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 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 métricas do modelo de regressão logística, uma

das métricas é a acurácia, 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, outras métricas, como precisão, recal (ou sensibilidade) e F1-score podem fornecer uma visão mais completa do desempenho do modelo.

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. 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.

Outro método para validar o desempenho da regressão logística é a matriz de confusão, após o test and score, essa matriz irá permitir observar as taxas de acerto do modelo durante o machine learning, podendo ser avaliado o tanto de previsões erradas o modelo fez, tanto para falha “sim” como para falha “não”.

3. Referencial Teórico

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 usada 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).

3.1 Regressão Logística

Entre os modelos de regressão, existem duas categorias principais. A primeira, composta por modelos de logit, inclui a regressão logística, onde a variável dependente Y é binária e segue a distribuição de Bernoulli, associada a uma variável independente X. A segunda categoria abrange os modelos loglineares, onde a variável de resposta é vinculada a uma variável aleatória de Poisson. (FIGUEIRA, 2006; FAVERO; FÁVERO, 2015).

A regressão logística é amplamente utilizada por pesquisadores como (HARRELL et al., 2001; GONZALEZ, 2018; FIGUEIRA, 2006) para classificar padrões úteis em bancos de dados, padrões que poderiam ser ignorados por outros modelos. Essa técnica visa prever a probabilidade de um evento específico ocorrer, como "sim" ou "não", "quebrado" ou "não quebrado", em situações binárias (dicotômicas), utilizando apenas uma variável independente. O modelo estatístico é preferido na análise de respostas binárias e permite determinar a probabilidade de ocorrência de um evento com base em variáveis qualitativas, revelando a influência independente de cada variável sobre o evento de "quebra" ou "não quebra". O modelo de regressão logística viabiliza várias 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 trabalhou com a probabilidade do evento, de maneira descrita nas equações (1) e (2):

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

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

3.1.1 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 de o 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 21 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.

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

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):

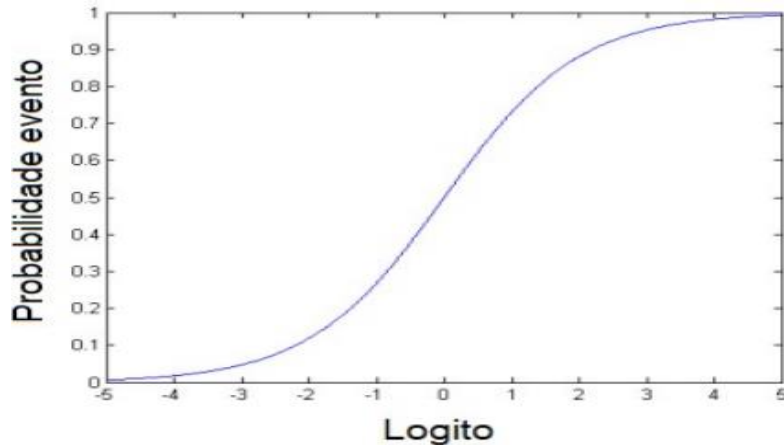
$$\text{logit}(z) = \frac{e^z}{1 - e^z} \quad 5$$

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):

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

Pois desse modo a inversa da função logit retornará à probabilidade da variável dependente ser 1 ou 0. A curva da figura 5 pode ser chamada de curva de sigmoide, e em resumo e apenas a inversão dos eixos x e y .

Figura 5 – Exemplo de curva de logit x probabilidade



Fonte: Autor

3.1.2 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 à equação (8):

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

colocando a equação de z (9) em função da equação (8) conforme a função descrita na (8), é gerado a equação de probabilidade de o 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 de o evento acontecer é conforme a equação (10), as variáveis independentes são X_1, X_2, \dots, X_k e os parâmetros definidos foram α, β_1, β_2 até β_k através do método de máxima verossimilhança.

3.1.3 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):

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

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):

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

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):

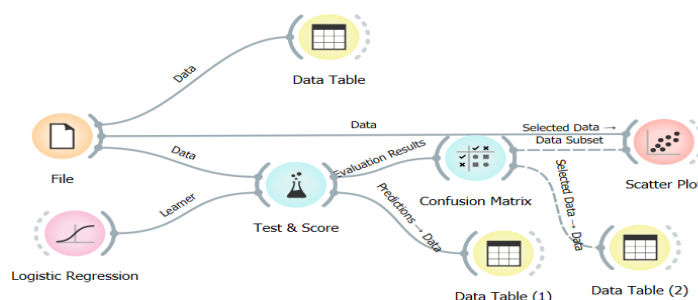
$$\frac{\partial \ln(L(\beta))}{\beta_1} = \sum_{i=1}^n y_i x_i - \pi(x_i) x_i \quad 14$$

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.

4. Resultados

Os registros foram coletados ao longo de cinco meses, de 11 de abril de 2023 a 31 de agosto de 2023, totalizando 202 mil linhas que descrevem o funcionamento da criogenia de três máquinas do modelo RM General Electric GE LX 1.5 T. O estudo analisa a eficácia da regressão logística na detecção de falhas nas Ressonâncias Magnéticas, utilizando dados coletados por sensores e armazenados nos datalogger das máquinas. Destaca-se a necessidade de ajustes no modelo e na base de dados ao aplicá-lo em outros equipamentos, visto que a regressão logística foi adaptada especificamente para os dados das ressonâncias GE 1.5 T. O sistema completo para a obtenção do modelo foi implementado no software Orange Canvas, conforme ilustrado na figura 6

Figura 6 – Modelo completo no software Orange



Fonte: Autor

Para validar o modelo foi utilizado métricas estatísticas, já descritas na metodologia do trabalho. 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'. Precisão e acurácia estão próximos 1, o que demonstra que o modelo se adaptou bem ao banco de dados.

Tabela 2 – Test and Score regressão logística

Modelo	AUC	CA	F1	Prec	Recall
Regressão Logística	0.833	0.933	0.905	0.934	0.933

Fonte: Autor

Ao comparar ambas as regularizações no Orange 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
Regressão Logística Lasso	0.833	0.933	0.905	0.934	0.933
Regressão Logística Ridge	0.833	0.932	0.905	0.934	0.932

Fonte: Autor

O modelo preditivo obtido através do método de aprendizagem supervisionada de máquina da regressão logística binária foi:

$$P(\text{falha} = 1) = \frac{1}{1 + e^{-z}} \quad 15$$

Onde:

$$z = 19,747 - 0,1690 * He_{level} - 0,7325 * Water_{temp} + 0,0362 * Shield_{si} + 0,0801 * Recon_{Ru} + 1,1277 * Recon_{si} - 1,1579 * ColdHead \quad 16$$

Para avaliar o desempenho do método supervisionado de regressão logística foram utilizadas duas medidas fornecidas pela Matriz de Confusão e pela curva ROC.

Observando-se a matriz de confusão da figura 7, foi possível obter as medidas de desempenho do modelo preditivo.

Observando-se a matriz de confusão foi possível obter as métricas de desempenho do modelo, conforme descritos nas equações 17,18,19,20:

Figura 7 – Matriz de Confusão para o Modelo de Regressão Logística

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

Fonte: Autor

$$Acurácia = \frac{93620 + 610}{101050} = 0,9325 \quad 17$$

$$\text{Sensitividade} = \frac{610}{610 + 6790} = 0,0824 \quad 18$$

$$\text{Especialidade} = \frac{93620}{93620 + 30} = 0,9997 \quad 19$$

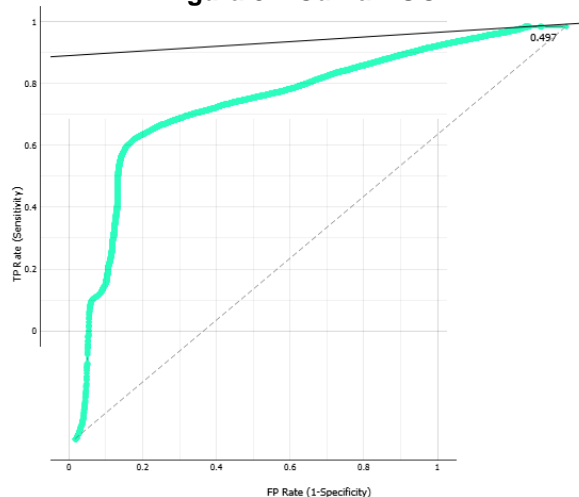
$$\text{Precisão} = \frac{610}{610 + 30} = 0,9531 \quad 20$$

A acurácia do modelo, ou seja, a porcentagem de falhas previstas corretamente foi de 93,25%. A sensibilidade, ou seja, a capacidade do modelo em identificar corretamente as falhas foi de 8,24%. A especificidade, ou seja, a capacidade do modelo em identificar corretamente as não falhas foi de 99,97%. A precisão, ou seja, a proporção de previsões corretas pelo modelo em relação ao total de previsões feitas foi de 95,31%.

Vale destacar que o valor de 8,24% de sensibilidade se deve pela reduzida quantidade de observações onde não houve falha presentes no banco de dados.

A curva roc ou a area sob 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. Conforme os valores demonstrados na tabela 1 e na figura 8, mostra que o modelo está com um valor 0.833, valor aceitável que o modelo está distinguindo bem as classes de falha "sim" e "não".

Figura 8 – Curva ROC



Fonte: Autor

5. Conclusões

Este estudo explorou o uso da regressão logística em dados de criogenia de uma máquina de ressonância magnética General Electric 1.5 T, com o objetivo de desenvolver um modelo preditivo para otimizar o cronograma de manutenção preditiva e minimizar interrupções não programadas. A metodologia empregou o software Orange e processos de banco de dados, com avaliação dos resultados por meio de métricas como a curva ROC e a matriz de confusão.

Os resultados revelaram a capacidade do modelo em identificar potenciais falhas, mesmo com dados limitados e poucas máquinas para a coleta de logs, validando sua eficácia em antecipar problemas e apoiar estratégias de manutenção proativa. A implementação desse modelo pode reduzir significativamente os custos de manutenção, prevenindo

intervenções desnecessárias e prevendo falhas, especialmente na área crítica de criogenia, onde problemas podem resultar em prejuízos consideráveis.

No entanto, a pesquisa enfrentou desafios de coleta de dados devido à necessidade de acessar fisicamente as máquinas em horários específicos durante o expediente clínico, limitando a disponibilidade de informações. Assim, para futuros estudos, é essencial ampliar a coleta de dados, tanto em relação ao funcionamento regular das máquinas quanto às falhas de criogenia, a fim de fortalecer as métricas do modelo. Além disso, a inclusão de mais máquinas para a extração de logs ajudará a validar a robustez do modelo em diferentes cenários operacionais, garantindo sua eficácia prática na implementação de estratégias de manutenção preventiva em grande escala.

Referências

FAVERO, L.; FÁVERO, P. **Estatística aplicada: Para cursos de Administração, Contabilidade e Economia com Excel e SPSS**. [S.l.]: Elsevier Brasil, 2015.

FIGUEIRA, C. V. Modelos de regressão logística. Dissertação (Mestrado em Matemática) -Programa de Pós-graduação em matemática . . . , 2006.

GONZALEZ, L. d. A. Regressão logística e suas aplicações. Universidade Federal do Maranhão, 2018.

HARRELL, F. E. et al. **Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis**. [S.l.]: Springer, 2001. v. 608.

MARCORIN, W. R.; LIMA, C. R. C. Análise dos custos de manutenção e de não-manutenção de equipamentos produtivos. **Revista de ciência & tecnologia**, v. 11, n. 22, p. 35–42, 2003.

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 (ONLINE)**, 2019.

MINGUETTI, G. Ressonância magnética na esclerose múltipla: análise de 270 casos. **Arquivos de Neuro-Psiquiatria**, SciELO Brasil, v. 59, p. 563–569, 2001.

NEPOMUCENO, L. X. **Técnicas de manutenção preditiva-vol. 1**. [S.l.]: Editora Blucher, 2014. v. 1.

PAOLANTI, M. et al. Machine learning approach for predictive maintenance in industry 4.0. In: IEEE. **2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)**. [S.l.], 2018. p. 1–6.

PIRES, C. A.; OKADA, R. H. Manutenção preditiva: estratégia de produção e redução de custos. **Revista Interface Tecnológica**, v. 17, n. 1, p. 635–647, 2020.

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)**, v. 5, n. 1, p. 94–111, 2020.