**Aluno(a): Lucas Ribeiro Alves**

**Orientador(a): Felipe Pinto da Silva**

**Curso: MBA em Data Science e Analytics**

**Machine Learning Explicativo (XAI) Aplicado ao Diagnóstico de Falhas Intermitentes de Contadores de Eixos**

**Introdução**

A segurança e a eficiência operacional das ferrovias modernas dependem criticamente da confiabilidade dos seus sistemas de sinalização de via (UIC, EN 50126). Entre os componentes vitais desta infraestrutura, encontram-se as balizas de posicionamento, TREs (), os AMV (Aparelhos de mudança de via), e, entre outros, os **Contadores de Eixos (CE)**, que emergiram como a principal tecnologia para determinar a ocupação e a integridade de um trecho de via, superando as limitações dos circuitos de via tradicionais

Os CE, em sistemas de sinalização como o CBTC (Communications Based Train Control), têm a função de realizar a detecção secundária do posicionamento dos trens na via, de forma que o sistema possa ser operado com segurança, ainda que o sistema primário (Comunicação Wireless) esteja inoperante, ou que haja na via um veículo não equipado com o sistema de bordo do CBTC (ALI, 2017). No entanto, a operação em ambientes extremos e constante exposição a choques mecânicos e intempéries fazem com que esses dispositivos estejam sujeitos a falhas, que, por vezes, são intermitentes.

Tais falhas, mesmo quando são de curta duração, são particularmente dispendiosas, pois exigem frequentes intervenções corretivas (reset manual) por parte das equipes de manutenção, que precisam parar suas atividades e se deslocar até o local da falha, resultando em interrupções no tráfego, atrasos e custos operacionais elevados.

A causa-raiz desses eventos é, muitas vezes, difícil de ser isolada, pois resulta da interação complexa de fatores contextuais como flutuações de tensão, variações de temperatura e níveis anormais de vibração na estrutura da via (ROSA, 2022).

Casos assim, demandam um tipo de manutenção mais moderna e voltada para a análise inteligente de dados coletados no momento da ocorrência, ao invés de uma atuação reativa a um problema que não tem uma causa aparente. Uma tecnologia muito útil para esse tipo de abordagem tem se tornado cada vez mais popular, desde o início do século, com o que tem sido chamada de quarta revolução industrial (SCHWAB, 2016), e é o Machine Learning (ML), que é um campo da inteligência artificial que quebra o paradigma tradicional da programação, onde um computador é programado para realizar uma tarefa específica, e passa ao computador a tarefa indireta de aprender os padrões existentes em dados fornecidos (ROSA, 2022).

Em um cenário onde a Manutenção Preditiva (Predictive Maintenance – PdM) se consolida como pilar da indústria 4.0 no setor ferroviário (SCHWAB, 2016; FERREIRA, 2021), este trabalho propõe a aplicação de técnicas de ML para automatizar a Análise de Causa-Raiz (Root Cause Analysis - RCA) das falhas nos contadores de eixos. Enquanto a manutenção preditiva tradicional foca em prever quando a falha ocorrerá, o foco deste estudo é determinar o que causa a falha, oferecendo um diagnóstico explicativo.

**Objetivo**

O objetivo principal dessa pesquisa é desenvolver uma metodologia de classificação baseada em Machine Learning – especificamente utilizando o algoritmo Random Forest (BREIMAN, 2001) – para ranquear a importância dos atributos (Feature Importance). O modelo será treinado para classificar os momentos que antecedem um reset como um estado de “pré-falha”, identificando, assim, quais variáveis coletadas pelo monitoramento (vibração, temperatura, umidade etc.) são as maiores preditoras e, consequentemente, as causas-raiz mais prováveis do evento de falha para um determinado circuito de via da detecção secundária.

Dessa forma, gerando uma redução significativa nas atuações das equipes de manutenção corretiva, reduzindo custos operacionais e, de forma geral, melhorando o serviço prestado ao cliente final.

**Material e Métodos**

Essa pesquisa se caracteriza como um estudo de caso de múltiplos casos, utilizando uma abordagem quantitativa aplicada, com foco na Análise de Causa-Raiz de falhas em um sistema crítico de sinalização de via. O projeto adota uma metodologia Ex-Post-Facto, pois busca identificar as variáveis preditoras e causais após a manifestação do evento de falha (o reset do cartão de ocupação do circuito de via).

O projeto é conduzido em uma Ferrovia de Carga de Passageiros localizada no estado de São Paulo, com uma malha extensa e alta circulação de trens. A instituição opera com uma frota superior a 100 trens, tendo a segurança operacional como prioridade crítica para a gestão de ativos.

O estudo será realizado em dois trechos de via contíguos (ou geograficamente próximos e com características operacionas semelhantes) pertencentes à ferrovia:

1. **Caso Teste (Sensor Anômalo)**: Um trecho de via onde a taxa de falhas intermitentes nos CE é alta, resultando em frequentes intervenções corretivas (resets) e impactos operacionais.
2. **Caso Controle (Sensor Normal)**: Um trecho de via com características de tráfego, geometria e ambiente semelhantes, mas que apresenta um histórico estável e baixa incidência de falhas.

O objeto de estudo é o conjunto de dados primário (sinais elétricos, ambientais e mecânicos) gerados pelos sensores instalados para monitorar o sensor indutivo e seus sistemas de alimentação/processamento periférico (o equipamento da sala técnica).

A coleta de dados é realizada por um período de uma semana (7 dias) ininterruptos de operação normal, sem interferir na operação ou manutenção do sistema de sinalização. A técnica de obtenção de informações é o levantamento de Dados Primários através de instrumentação instalada.

A instrumentação e os dados coletados (com timestamp para cada registro) foram obtidos como descrito na Tabela 1.

Tabela 1. Descrição dos Dados

|  |  |  |
| --- | --- | --- |
| Instrumento (Local) | Variáveis de Campo | Finalidade |
| Acelerômetro (Caixa de Via) | Níveis de Vibração nos Eixos Gx, Gy e Gz | Detecção de impactos, trepidação ou fixação inadequada |
| Sensores Ambientais (Caixa de Via) | Temperatura e umidade interna | Identificação de estresse térmico, vedação ineficaz ou condensação |
| Voltímetro (Caixa de Via) | Tensão de alimentação do sensor | Detecção de picos, quedas ou flutuações na fonte de energia |
| Coletor de Sinais (Sala Técnica) | Sinais indutivos, de Ocupação e de Reset | Definição da variável alvo (o reset é a ocorrência de falha) e monitoramento do comportamento elétrico do sinal |

Fonte: Dados originais da Pesquisa

O modelo de análise é baseado em ML e segue as seguintes etapas:

1. **Unificação dos Dados (JOIN)**: As três bases de dados coletadas (sensor anômalo, sensor normal e sala técnica) serão unificadas em um único dataset macro usando o timestamp como chave primária, garantindo a coerência temporal entre a condição da via e o estado da sala técnica.
2. **Criação da Variável Alvo (Target)**: Será criada uma variável binária para o aprendizado supervisionado. Os dados coletados em uma janela de tempo de 15 minutos (a ser ajustada se necessário) imediatamente anterior a cada evento de reset (manual ou automático) no **Caso Teste** serão rotulados como TARGET = 1 (“pré-falha”). Todos os demais dados serão rotulados como TARGET = 0 (“Normal”).
3. **Algoritmo**: O algoritmo utilizado para o treinamento será o Random Forest. A escolha é fundamentada na capacidade da RF de lidar com a não-linearidade e heterogeneidade dos dados de sensores e, principalmente, por sua característica de oferecer um ranking de Importância dos Atributos robusto e interpretável.
4. **Análise Explicativa (eXplainable AI – XAI)**: O modelo será treinado com os dados unificados. O foco não é a previsão (predictive), mas sim a explicação (descriptive/explicative).
5. **Resultado Primário**: O resultado da modelagem será o ranking da importância dos atributos, que indicará quais variáveis foram mais relevantes na classificação do estado de “Pré-Falha”.

Dado que a coleta de dados é realizada exclusivamente em equipamentos e na infraestrutura física, sem envolver dados pessoais, entrevistas ou participação direta de indivíduos, este projeto não requer sumissão e aprovação pelo Comitê de Ética em Pesquisa (CEP).

**Resultados Esperados**

O presente projeto de pesquisa, por seu caráter aplicado e experimental, visa produzir resultados que transcenderão a análise descritiva, fornecendo insights concretos para a gestão de ativos e otimização da manutenção na infraestrutura ferroviária.

Espera-se que o modelo de RF, treinado com os dados primários de condição do Caso Teste e do Caso Controle, seja capaz de:

1. **Classificação e Determinação da Causa-Raiz (RCA)**
   1. **Geração do Ranking de Importância dos Atributos**: O resultado central esperado é um ranking quantitativo, gerado pelo modelo RF, que classifique as variáveis mais importantes para a ocorrência do estado de “pré-falha”. Este ranking permitirá a identificação e o isolamento dos fatores causais mais prováveis para o sensor anômalo.
   2. **Identificação de Limites de Anomalia**: Espera-se que a análise explicativa revele limites de operação não documentados. Por exemplo, que o modelo aponte que a falha tem alta probabilidade de ocorrer quando a temperatura interna da caixa está alta em combinação com uma flutuação na fonte de energia.
2. **Validação da Metodologia e Otimização da Manutenção**
   1. **Validação do Desenho Experimental**: Comprovar a eficácia da metodologia de comparação dual (Teste vs. Controle). Os resultados devem demonstrar que as variáveis críticas identificadas no Caso Teste não se manifestam com a mesma relevância ou intensidade no Caso Controle, isolando o problema.
   2. **Transição Para a Manutenção Prescritiva**: A informação gerada pelo ranking de importância dos atributos permitirá à equipe de manutenção ir além do reset manual. Os resultados esperados serão recomendações específicas de manutenção prescritiva, como focar a inspeção na fixação mecânica da caixa de via (se a vibração for a principal feature), ou melhorar a vedação da caixa (se a umidade for a principal feature).
3. **Contribuição Científica e Aplicada**
   1. **Contribuição para o conhecimento aplicado**: O projeto visa contribuir para a Engenharia Ferroviária, demonstrando a viabilidade e o valor do uso de técnicas de XAI para solucionar problemas complexos e intermitentes em sistemas de sinalização de vias, um campo tradicionalmente dominado por métodos determinísticos.

**Cronograma de Atividades**

O Cronograma de Atividades é o planejamento e organização da pesquisa e da escrita do TCC, o qual deve ser elaborado considerando as entregas das etapas do TCC estipuladas pela coordenação do curso.

**Atenção:** antes de enviar o arquivo para o Sistema de TCCs, remova todas as instruções originais que estão abaixo do conteúdo dos tópicos.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Atividades planejadas** | **Mês** | | | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

**Referências**

ALI, Naeem. Key CBTC Functions Transit Operators Must Understand. 2017.

UNIÃO INTERNACIONAL DE FERROVIAS (UIC) / NORMAS EUROPEIAS (EN 50126, EN 50128, EN 50129).

**ROSA, Ciro Magalhães da. Aplicação de Machine Learning na identificação das causas de falhas em máquinas para auxílio na gestão da manutenção. 2022. 64 f. Trabalho de Conclusão de Curso (Graduação em Engenharia de Produção) - Universidade Federal Fluminense, Niterói, 2022**

**SCHWAB, Klaus.** **A Quarta Revolução Industrial.** São Paulo: Edipro, 2016

**FERREIRA, Macilio da Silva.** **Aplicação de Técnicas de Aprendizado Profundo para a Detecção e Diagnóstico do Estado de Operação de Ativos Ferroviários.** Tese de Doutorado ou artigos derivados, 2021.

**BREIMAN, Leo.** **Random Forests.** *Machine Learning*, vol. 45, n. 1, p. 5–32, 2001

Nas Referências deve ser elencado todas as obras utilizadas para a elaboração do Projeto de Pesquisa e normatizadas conforme o que se pede no “[Manual de Instruções e Normas para Trabalhos de Conclusão de Curso](https://www.dropbox.com/scl/fi/m1t43csoioys8x54pz60z/Manual-de-Instru-es-e-Normas-TCC_PT.pdf?rlkey=7isvkxss2lcl9f6gj9tkx9ipw&st=81xbs4gh&dl=0)”.

**Atenção:** antes de enviar o arquivo para o Sistema de TCCs, remova todas as instruções originais que estão abaixo do conteúdo dos tópicos.