

Sinalize, Ferramenta para análise preditiva de falhas em transformadores de frequência e potência do Sistema Interligado Nacional

**Rodrigo Ramos Guimarães,
Orientadores: Sergio Manuel Serra da Cruz e Jorge Zavaleta**

Programa de Pós-Graduação Em informática – Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro – RJ – Brasil

Abstract. This paper presents the implementation of a predictive model for anticipating failures in power grid transformers of the main grid of Sistema Interligado Nacional (SIN), operated by Operador Nacional do Sistema Elétrico (ONS). Using Random Forest algorithms and data balancing techniques, the tool named SINALIZE was able to predict failures across different time horizons based on historical characteristics of maintenance, failures, and equipment usage. The results demonstrate the feasibility of applying machine learning techniques to predictive maintenance in electric power systems, contributing to improved reliability, safety, and reduced operational costs.

Resumo. Este trabalho apresenta a implementação de um modelo preditivo para antecipação de falhas em transformadores da rede básica do Sistema Interligado Nacional (SIN), operado pelo Operador Nacional do Sistema Elétrico (ONS). Utilizando algoritmos de Random Forest e técnicas de balanceamento de dados, a ferramenta, denominada SINALIZE, foi capaz de prever falhas em diferentes horizontes temporais, com base em características históricas de manutenção, falhas e utilização dos equipamentos, demonstrando a viabilidade da aplicação de técnicas de aprendizado de máquina para manutenção preditiva em sistemas elétricos de potência, contribuindo para a melhoria da confiabilidade, segurança e redução de custos operacionais.

Palavras-chave: Manutenção Preditiva, Random Forest, Transformadores, Sistema Interligado Nacional, ONS

1 Introdução

O Sistema Interligado Nacional (SIN) é um sistema elétrico responsável pela produção e transmissão de energia elétrica no Brasil. Trata-se de um sistema híbrido de grande porte, com predominância de usinas hidrelétricas com múltiplos proprietários, com grande participação da geração térmica, fotovoltaica, eólica, estruturas sistêmicas complexas para a transmissão, e com grande carga regulatória, gerenciado pelo Operador Nacional do Sistema Elétrico (ONS), que conta com quatro centros de operação, cobrindo além de todo território nacional, a importação e exportação de energia para países vizinhos (intercambio internacional). Esta operação é coordenada através de normas técnicas chamadas Procedimentos de Rede (PR), que garantem a segurança equidade e qualidade do fornecimento de energia.

A rede básica [ONS Procedimentos de Rede - Submódulo 2.1 2024], é composta por um subconjunto de estruturas, com linhas de transmissão e equipamentos de com tensão a partir de 230kV, que formam a espinha dorsal do SIN, conectando grandes produtores de energia aos distribuidores, esta estrutura é responsável pela ligação das grandes usinas aos centros de consumo.

A operação deste sistema complexo requer monitoramento contínuo e manutenção adequada de seus componentes críticos, especialmente os transformadores, que são elementos fundamentais para a transmissão. As falhas destes equipamentos, podem resultar em interrupções significativas no fornecimento de energia, causando prejuízos econômicos e sociais consideráveis. Conforme os procedimentos de Rede, a manutenção destes equipamentos segue estratégias preventivas baseadas em cronogramas fixos ou corretivas após a ocorrência de falhas. No entanto, a evolução das tecnologias de análise de dados e aprendizado de máquina trás oportunidades para implementação de estratégias de manutenção preditiva mais eficientes.

Este trabalho propõe o desenvolvimento de um modelo preditivo baseado em Random Forest para antecipação de falhas em

transformadores da rede básica do SIN. O modelo utiliza dados históricos de manutenção, falhas e utilização dos equipamentos para prever a probabilidade de falha em diferentes horizontes temporais, permitindo ao ONS otimizar as estratégias de manutenção, ajustando estes tempos melhorando a confiabilidade do sistema e economizando recursos financeiros para os agentes e sociedade, recursos estes que em 2022, pagos como penalização pelos agentes, somaram mais de R\$ 48Mi [], dinheiro poderia ser utilizado em melhorias significativas na rede, sem considerar o custo intangível da sociedade sem energia em diversas ocasiões.

2 Revisão de Literatura

A aplicação de técnicas de aprendizado de máquina em sistemas elétricos de potência tem crescido significativamente nos últimos anos. Diversos estudos demonstram a eficácia de algoritmos como Random Forest, Support Vector Machines e Redes Neurais na predição de falhas em equipamentos elétricos, contudo, os estudos esbarram em limitações relacionadas à cobertura de dados, utilizando primariamente dados com escopo local, com acesso limitado à dados de múltiplos agentes e equipamentos, tornando mais difícil obter respostas assertivas.

O Random Forest, proposto por [Breiman 2001], é um algoritmo versátil, que combina múltiplas árvores de decisão para melhorar a precisão e robustez das predições. Sua capacidade de lidar com dados não-lineares, resistência ao sobreajuste facilidade de implementação e interpretação, o tornam particularmente adequado para aplicações desta natureza.

3 Metodologia

O trabalho consiste na coleta e preparação dos dados, com fonte a partir de dados públicos do portal de dados abertos do ONS, cruzamento com séries de utilização de equipamentos (cargas), e registros de ocorrências de falhas, além de um processo de

descaracterização dos dados, tornando impossível a identificação de registros reais, que não sejam dados públicos.

3.1 Coleta de dados

Os dados foram coletados e tratados em um processo de ETL, que entrega seus datasets finais na pasta RAW do projeto, permitindo a reprodução do estudo, incluindo futuros trabalhos e evoluções. Os seguintes datasets foram utilizados:

- Cadastro de transformadores
O cadastro de transformadores foi extraído do portal de dados abertos do ONS, disponíveis em:
<https://dados.ons.org.br/dataset/capacidade-transformacao>
- Dados de manutenções preventivas
Desligamentos programados para manutenção preventiva e atividades mínimas de manutenção, extraídas do sistema de apuração da transmissão interno do ONS
- Dados de ocorrências de falhas nos equipamentos
Dados de perturbações do sistema, ocorrem quando há uma falha em um equipamento, esta falha é registrada e tratada por um sistema interno do ONS
- Dataset com séries temporais, de utilização (carga) de equipamentos
Dados extraídos do sistema de supervisão e controle, em formato de séries temporais por ampere e minuto a minuto

Com objetivo de garantir maior qualidade dos dados, o período de análise foi limitado a partir de janeiro de 2023. Foram considerados apenas equipamentos que possuíam tanto registros de manutenção quanto de falhas, resultando em um conjunto de dados balanceado e representativo.

3.2 Criação de features

Foi desenvolvido um processo de ETL para extrair datasets em formato de features a partir dos dados brutos, diversas experiencias foram realizadas na engenharia de features, os seguintes itens se mostraram aderentes ao desenho da solução:

- Features de Características do Equipamento
 - idade_dias: Idade do equipamento em dias, média utilizando a data de entrada em operação comercial;
 - limite_potencia: Limite nominal de potência do equipamento;
- Features de Histórico de Manutenção
 - dias_desde_ultima_manut: Dias transcorridos desde a última manutenção preventiva do equipamento;
 - num_anutencoes: Quantidade total de manutenções preventivas realizadas no equipamento;
 - intervalo_medio_manut: Intervalo médio em dias em que as manutenções são realizadas nos equipamentos;
- Features de Falhas
 - num_falhas_historico: Número de falhas históricas do equipamento
 - taxa_falhas_ano: Taxa anualizada de falhas
 - minutos_falha_historico: Total de minutos em que o equipamento esteve em falha
 - taxa_minutos_falha_ano: Taxa anualiza de minutos em falha
- Features de utilização dos equipamentos
 - utilizacao_media: Utilização média do equipamento no mês
 - utilizacao_maxima: Pico de utilização do equipamento no mês
 - utilizacao_minima: Mínimo de utilização do equipamento no mês
 - qtd_sobrecargas: Número de violações de limite do equipamento

imagem com gráfico das features e visão geral

3.3 Definição do objetivo do modelo

O objetivo foi definido como uma variável denominada: **vai_falhar** que indica se o equipamento apresentará falha no período futuro especificado de X dias. Esta abordagem permite ao modelo aprender

padrões que precedem as falhas, possibilitando intervenções preventivas.

3.4 Tratamento para balanceamento de dados

Como as falhas em transformadores são eventos relativamente raros, são produzidas classes desbalanceadas. Para tratar estes casos, foi utilizado a técnica SMOTE [Chawla et al., 2002], que basicamente cria amostras intermediárias através da interpolação de dados vizinhos. Técnica esta, que aumentou consideravelmente a performance geral do modelo

3.5 Escolha do Algoritmo Random Forest

A escolha do Random Forest como algoritmo principal foi baseada em suas vantagens específicas para o problema aqui tratado, pois diferente de modelos de árvore de decisão tradicionais, ele combina múltiplas árvores e seleção aleatória de features, reduzindo o risco de sobreajuste (overlifting), o que na prática, significa que mesmo com dados históricos de exemplo mais raros, ele não irá decorar o dado, entendendo ruídos como casos reais, além de lidar bem com features numéricas sem necessidade de pré-processamento extensivo, na contramão de SVMs, que demandam normalização dos dados de forma extensiva. O Random Forest também demanda menos poder computacional para trabalhar, facilitando a geração de treinamentos para testes e ajuste fino do modelo.

3.6 Configurações e Parâmetros do Modelo Random Forest.

Foi utilizado o RandomForest do pacote sklearn, configurado com os seguintes parâmetros:

- `n_estimators`: 100,1000
 - Define o número de árvores na floresta, mais árvores tendem a melhorar a estabilidade e a generalização do modelo.

- `max_depth: 10,20,30`
 - Controla a profundidade máxima das árvores. Árvores mais profundas capturam padrões mais complexos, mas podem gerar problemas de overfitting
- `min_samples_split: 2,5`
 - Número mínimo de amostras necessário para dividir um nó, valores maiores tornam o modelo mais conservador ajudando a evitar divisões baseadas em poucos dados (ruído), mas valores pequenos podem não pegar casos reais diminuindo a performance do modelo
- `class_weight: balanced`
 - Ajusta automaticamente os pesos das classes de forma inversamente proporcional à sua frequência, é essencial em problemas com classes desbalanceadas, como falhas raras em equipamentos. Evita que o modelo favoreça apenas a classe majoritária.

***Gráfico com o mapa de calor das features

3.7 Otimização com Grid Search e validação cruzada

O Grid Search testa todas as combinações possíveis desses parâmetros e seleciona aquela que apresenta o melhor desempenho segundo a métrica definida (ex.: F1-score, ROC-AUC), e na validação cruzada, os dados são divididos em 5 subconjuntos (folds), e em cada iteração 4 são usados para treino e um para validação, forçando que a proporção de classes seja mantida em todos os folds, como benefícios desta abordagem temos:

- Avaliação mais confiável do desempenho do modelo;
- Redução do risco de overfitting;
- Melhor generalização para dados.

3.8 Métricas de avaliação

O desempenho do modelo foi avaliado utilizando as seguintes métricas:

- ROC-AUC: Área sob a curva ROC, medindo a capacidade de discriminação;
- Precision: Proporção de predições positivas corretas;
- Recall: Proporção de casos positivos corretamente identificados;
- F1-Score: Média harmônica entre precision e recall;
- Accuracy: Proporção total de predições corretas.

4 Resultados

4.1 Características do dataset final

O dataset final, gerado na pasta gold, compreendeu equipamentos válidos (com histórico de manutenção e falhas), com distribuição variável de falhas dependendo do horizonte temporal analisado. A Tabela 1 apresenta a distribuição de falhas para cada período contemplado no estudo:

Tabela 1. Distribuição de falhas por período de análise em dias

Período (Dias)	Equipamentos	Falhas	% Falhas
30	1168	41	3,51%
35	1167	44	3,77%
40	1167	53	4,54%
45	1167	60	5,14%
50	1165	73	6,27%
55	1163	86	7,39%
60	1160	94	8,10%
65	1159	109	9,40%
70	1156	120	10,38%
75	1154	126	10,92%
80	1153	131	11,36%
85	1152	141	12,24%
90	1152	154	13,37%

Fonte: Produzido pelo autor

4.2 Importância das Features

A análise de importância das features confirma a intuição técnica sobre os fatores que influenciam falhas em transformadores. A predominância da taxa histórica de falhas como preditor principal sugere que equipamentos com histórico problemático tendem a continuar apresentando falhas, validando a abordagem baseada em dados históricos.

O tempo desde a última manutenção emerge como segundo fator mais importante, reforçando a importância de estratégias de manutenção preventiva adequadas. A idade do equipamento, de maneira contraintuitiva, aparece em quarto lugar, indicando que os equipamentos mais novos tendem a apresentar defeitos em número maior que equipamentos mais antigos, e indicando que o histórico operacional acaba sendo mais relevante que o tempo de uso do equipamento, o que é um possível indicativo de obsolescência programada

**** [[FOTOS DOS GRÁFICOS de IMPORTANCIA]]

4.3 Análise final sobre a escolha do modelo

A capacidade de fornecer dados interpretáveis de importância das features confirmou a escolha do Random Forest sobre alternativas como redes neurais profundas. Esta facilidade de interpretar os resultados, permitiu os especialistas da empresa, analisarem os resultados e validarem maneira preliminar, os aspectos técnicos e precisão dos resultados, em detrimento a outros algoritmos que

trabalham no formato de “caixa-preta”, que não oferecem esta visão, que facilita muito a interpretação e confiança no modelo.

4.4 Performance

Os resultados demonstram performance consistente e crescente com o aumento do horizonte temporal, indicando conforme o esperado, que o modelo é mais eficaz na predição de falhas em períodos mais longos.

*** foto da performance do modelo para achar o goldspot

4.5 Análise de casos de exemplo

- ** CASO 1 – Transformador com baixa probabilidade
- ** CASO 2 – Transformador com média probabilidade
- ** CASO 3 – Transformador com uma puta probabilidade.

4.6 Limitações e oportunidades de melhoria

O modelo apresenta algumas limitações a serem consideradas.

- Dados de fatores ambientais e climáticos
Não foram considerados dados de fatores ambientais, como temperatura média, humidade, índice UV, vento e fatores de queimadas, estes dados podem ser adicionados para estudos futuros;
- Dados de monitoramento do equipamento
Existem outros dados de monitoramento dos equipamentos, que não foram considerados no estudo, como temperatura e características internas, marca do óleo, modelo de placa de comunicação, fabricante do equipamento etc.;

- Detalhamento das falhas
Existem em sistemas internos, dados de detalhamento das falhas, indicando qual foi o componente que causou a falha, estes dados não foram considerados.

5 Conclusões

Este trabalho demonstrou que é viável a aplicação de técnicas de aprendizado de máquina para prever falhas em equipamentos, mais especificamente transformadores. Mesmo não apresentando performance satisfatória em horizontes temporais de curto prazo, obteve em geral ROC-AUC variando de 0.847 a 0.879, em uma validação empírica com dados reais do sistema elétrico descaracterizados.

Dentre as contribuições do trabalho, pode-se destacar pontos como uma metodologia sólida, pois o pipeline de dados trata dados brutos, extraídos diretamente dos sistemas de origem, gerando os arquivos de features, adaptáveis a outros tipos de equipamentos e cenários com pouco desenvolvimento, confirmando a possibilidade de expandir o estudo para outros tipos de equipamentos, como linhas de transmissão, bancos de capacitores, equipamentos de controle reativo, e até mesmo na parte de geração com usinas e unidades geradores.

E como contribuição à sociedade, o modelo já contempla de partida, uma API para simulação de casos especiais, habilitando a utilização produtiva, e integração com sistemas de gestão internos.

6 Trabalhos futuros

Dentre as principais evoluções do modelo, destacam-se:

- **Dados de meteorologia**
A agregação de dados de meteorologia ao modelo, agrega novas possibilidades de fatures a serem analisadas, como temperatura média, vento, índice UV e até incidência de raios na região do equipamento. Os dados estão disponíveis em dataset já formatado e serão tratados em demandas futuras;
- **Mais metadados de equipamentos**
Adicionar mais propriedades de equipamentos, como fabricante das placas controladoras, modelo detalhado, marca, fabricante, temperatura média de utilização, existe em alguns casos até o

datasheet do equipamento, que pode ser indexado via web scrapping no futuro;

- **Categorização do detalhamento das falhas via LLM**
As falhas possuem um detalhamento textual, criado pelos engenheiros da apuração, com a descrição da falha, indicação da causa e impactos em outros equipamentos, este texto pode ser tratado utilizando uma LLM e categorizando para novas features do modelo a serem analisadas;
- **Agrupamento via cluster para identificações de padrões comportamentais de falhas**
O modelo pode prover base para estudos mais aprofundados, identificando padrões desconhecidos de características ofensoras ao sistema, respondendo perguntas por exemplo: Existe algum agente negligenciando a manutenção e causando problemas? Existe algum tipo de equipamento que está sendo utilizado incorretamente? E etc.;

Referencias

[Glossário ONS 2025] OPERADOR NACIONAL DO SISTEMA ELÉTRICO (ONS). Glossário de termos do setor elétrico. Disponível em: <https://www.ons.org.br/paginas/conhecimento/glossario>. Acesso em: 10 nov. 2025.

[ONS - que é o SIN 2025] OPERADOR NACIONAL DO SISTEMA ELÉTRICO (ONS). O que é o SIN? Disponível em: <https://www.ons.org.br/paginas/sobre-o-sin/o-que-e-o-sin>. Acesso em: 10 nov. 2025.

[ONS Procedimentos de Rede - Submódulo 2.1 2024] OPERADOR NACIONAL DO SISTEMA ELÉTRICO (ONS). Procedimentos de rede – submódulo 2.1 – Definição das redes do Sistema Interligado Nacional Disponível em <https://www.ons.org.br/paginas/sobre-o-ons/procedimentos-de-rede/vigentes> Acesso em 13 nov. 2025.

[Breiman 2001] Breiman, L. Random Forests. *Machine Learning* 45, 5–32 (2001).
<https://doi.org/10.1023/A:1010933404324>

[B. C. Matheus et al 2024] Fault Detection and Prediction for Power Transformers
<https://www.mdpi.com/1996-1073/17/2/296>

[Syeda T. Zahra et al., 2025] Power Transformer Health Index and Life Span Assessment: A Comprehensive Review of Conventional and Machine Learning based Approaches — Syeda T. Zahra et al., 2025

[Chawla *et al.*, 2002] Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P
SMOTE: Synthetic Minority Over-sampling Technique.
Journal of Artificial Intelligence Research 16, 321–357 (2002).
<https://doi.org/10.1613/jair.953>