

GRUPO DE ESTUDO DE GERAÇÃO TÉRMICA - GGT

APLICAÇÃO DE MACHINE LEARNING PARA MANUTENÇÃO PREDITIVA EM USINAS TERMELÉTRICAS

**KLEYTON PONTES COTTA(1); ROBSON FELIPE VIANA DA SILVA(1); BRUNO DEON(1); CAMILLA BARROS BATISTA(1); GABRIELA TEIXEIRA JUSTINO(1); GABRIELA CARVALHO FREITAS(1); FLÁVIO LEITE LOUÇÃO JUNIOR(1); ATHOS DOS SANTOS BARBOSA(1); ANGELO MARCELINO CORDEIRO(1); ALEX MORETTI DE MORAIS(2); ISAAC EMMANUEL AZEVEDO DE MEDEIROS(2); RODRIGO JOSÉ SILVA DE ALMEIDA(2); CARLOS ANTONIO ALVES DE ARAÚJO JUNIOR(2)
RADIX ENGENHARIA E DESENVOLVIMENTO DE SOFTWARE S/A(1); EPASA(2)**

RESUMO

A operação em tempo real de sistemas, em especial na geração de energia elétrica, requer agilidade, confiabilidade e assertividade nas tomadas de decisões. Foi implementado um sistema de apoio à decisão em duas usinas termelétricas localizadas na Paraíba, utilizando algoritmos de aprendizado de máquina, para manutenção preditiva para classificação de falhas nas unidades geradoras da usina, além de uma arquitetura escalável capaz de executar algoritmos relacionados a diversos equipamentos simultaneamente. Os ganhos com a predição de eventos abrangem desde questões operacionais até objetivos estratégicos da empresa, devido a redução do número de paradas para manutenção corretiva e consequente redução dos custos de operação e manutenção.

PALAVRAS-CHAVE

Machine Learning, Manutenção Preditiva, Classificação e Predição de Falhas.

1.0 - INTRODUÇÃO

A operação em tempo real de sistemas, em especial na geração de energia elétrica, requer agilidade, confiabilidade e assertividade nas tomadas de decisões. A operação de sistemas de geração termelétricos por motores a combustão é utilizada no Brasil e em outros países, principalmente na forma de complementar a geração do sistema em períodos em que alternativas mais baratas e menos nocivas ao meio ambiente não são capazes de suprir a demanda nacional.

Este tipo de indústria atua com contratos por disponibilidade e apresentam grandes custos com matéria prima, manutenção das máquinas além de apresentar elevadas taxas de emissão de poluentes. Tendo em vista essas características no cenário atual, a busca por soluções de redução nas emissões de gases estufa torna-se importante garantir o máximo desempenho energético e o bom funcionamento das máquinas.

O uso de técnicas da indústria 4.0 como *Machine Learning* (ML) para apoio à tomada de decisão, são artifícios capazes de promover grandes ganhos operacionais, proporcionando um menor custo de manutenção. A abordagem de manutenção preditiva aplicando ML incrementa a disponibilidade, reduz os custos com manutenção e proporciona uma melhoria nas condições de funcionamento das máquinas.

A Manutenção preditiva necessita de sensores coletando dados em tempo real. Mas, além disto, também faz uso de conjuntos de dados históricos e registros de eventos de manutenção dos ativos monitorados. Estes dados, combinados com algoritmos personalizados de aprendizado de máquina, executam diferentes cenários e podem prever quando e qual falha irá ocorrer gerando alertas (1). Este tipo de manutenção está impulsionando a Indústria 4.0, onde as máquinas são conectadas como uma comunidade colaborativa e um sistema ecológico de tomada de decisão orientado a dados.

Os ganhos com a predição de eventos abrangem desde questões operacionais até objetivos estratégicos da empresa no mercado, devido a redução do número de paradas para manutenção corretiva e consequente redução dos custos de O&M (Operação e Manutenção). Essa abordagem exige a utilização de ferramentas avançadas de previsão e análise preditiva, para que os dados possam ser sistematicamente processados em informações e, assim, sejam tomadas decisões mais assertivas (2).

Uma arquitetura de sistema adequada para a manutenção preditiva 4.0 deve apresentar módulos específicos, de modo que a integração total dos módulos leve à implementação de um sistema operante e assertivo. Para isso, o

sistema necessita de componentes básicos. São eles: aquisição de dados, comunicação de dados, armazenamento de dados, análise preditiva e visualização dos resultados (1).

Na literatura, é amplamente discutido a necessidade de melhorar o desempenho dos motores a combustão, sem acarretar o aumento das emissões. Observou-se que métodos utilizados para monitoramento e diagnóstico de falha são divididos em, pelo menos, duas etapas. A primeira etapa consiste na identificação da existência de um distúrbio nos parâmetros operacionais do quadro saudável de operação. A segunda etapa é a classificação dessa falha, podendo então realizar as ações corretivas necessárias.

Dado o contexto apresentado, foi implementado um sistema de apoio à decisão para manutenção preditiva em duas usinas termelétricas localizadas na Paraíba, utilizando algoritmos de inteligência artificial (IA) para classificação de falhas nas unidades geradoras da usina. Para dar suporte ao sistema, foi criada uma arquitetura escalável capaz de executar algoritmos relacionados a diversos equipamentos simultaneamente. Considerando o caráter de execução em tempo real dos modelos, essencial para que a ferramenta atenda à necessidade de tomada de decisão da operação, se faz necessário uma aplicação *on premise*, na qual a hospedagem do sistema desenvolvida é realizada em servidor local. A arquitetura proposta pode ser aplicada a qualquer setor industrial que utilize supervisórios do tipo SCADA, podendo ser adaptada, expandida e evoluída para outras tecnologias de geração como: outras termelétricas, pequenas centrais hidrelétricas, eólicas e term nucleares.

As técnicas de ML empregadas para classificação e predição de falhas em conjunto com a arquitetura desenvolvida no projeto de pesquisa e desenvolvimento (P&D) da Aneel, podem ser replicadas a outros sistemas e, portanto, levando à abrangência de outros setores de energia como distribuição e transmissão, também podendo ser aplicado à indústria de uma maneira geral: química, petroquímica, óleo & gás, entre outras.

2.0 REVISÃO BIBLIOGRÁFICA

Nesta seção, apresentamos de modo sucinto o ambiente a ser analisado e os algoritmos de inferência utilizados na modelagem para manutenção preditiva. Utilizamos os algoritmos de rede neural artificial (MLP), *gradiente boosting* (GB) e *random forest* (RF), devidos seus desempenhos em outros trabalhos.

2.1 – Usina Termelétrica

Neste artigo estudamos o processo de geração de energia de uma termelétrica, no qual envolve a operação de duas usinas idênticas. Juntas possuem capacidade instalada de 342 megawatts, o que representa energia suficiente para atender uma população de cerca de 980 mil habitantes. A estrutura da UTE é composta por motores Diesel como equipamentos principais e subsistemas atrelados a estes, garantindo sua manutenção e funcionamento, contém geradores para conversão da energia cinética gerada pelos motores em energia elétrica.

Cada UTE conta com 19 Motores MAN/STX do tipo 18V32/40 com capacidade de gerar 8,76 MW e 1 Motor MAN/STX do tipo 9L32/40 com capacidade de gerar 4,38 MW. Os motores com a designação V32/40 são motores sobrealimentados, não reversíveis, de 4 tempos, em V, com 18 cilindros e 320 mm de diâmetro de cilindro e 400 mm de curso do pistão e com rotação de 720 rpm. Os motores com a designação 9L32/40 são motores sobrealimentados, não reversíveis, de 4 tempos, em L, com 9 cilindros e 320 mm de diâmetro de cilindro e 400 mm de curso do pistão e com rotação de 720 rpm. O consumo específico é de 212 kg/MWh, contudo este valor varia, principalmente, de acordo com o poder calorífico do combustível utilizado.

De modo simplificado os subsistemas presentes na UTE são:

- **Sistema de Admissão de Ar:** o ar é filtrado, comprimido no turbocompressor e resfriado através da troca térmica com água dos radiadores para, então, ser admitido no motor;
- **Sistema de Admissão de Combustível:** o combustível é aquecido através da troca de calor com o vapor e filtrado para ser alimentado no motor. O óleo combustível-HFO por apresentar alta viscosidade é preciso ser aquecido antes de ser alimentado no motor, porém o óleo diesel-LFO não precisa ser aquecido, pois a sua viscosidade é baixa o suficiente;
- **Sistema de Exaustão:** a queima do combustível com o ar nos motores gera gases de exaustão. Os gases de exaustão passam pelas turbinas dos turbocompressores e trocam calor com a água nas caldeiras de recuperação para geração de vapor. Em seguida, são encaminhados para exaustão nas Chaminés;
- **Sistema de Geração Vapor e Condensado:** o vapor é gerado da troca térmica da água com os gases de exaustão dos motores nas caldeiras de recuperação. Esse vapor gerado é utilizado para aquecimento do combustível. Vale salientar que as caldeiras de recuperação só podem ser acionadas quando há geração. Quando não há geração, são acionadas as caldeiras auxiliares que apresentam queimador de óleo diesel e gás natural;
- **Sistema de Lubrificação:** o óleo de lubrificação armazenado no cárter dos motores irá promover a lubrificação de todas as partes móveis do motor que entram em contato uma com as outras;

- **Sistema de Arrefecimento:** a água utilizada no processo promove o resfriamento do motor e suas partes (bicos injetores e sistema de ar de admissão). Essa água é oriunda dos radiadores e resfriam trocadores de calor dos demais sistemas.

2.1 – Synthetic Minority Oversampling Technique

Os classificadores gerados a partir de bases de treinamento desbalanceadas apresentam altas taxas de falsos negativos para as classes raras, o que é problemático quando a classe de interesse é classe rara. Para evitar esse viés, técnicas de pré-processamento dos dados são utilizadas para alterar as distribuições das classes na base de treinamento, visando reduzir desbalanceamento. Uma abordagem bastante comum é o uso de métodos de amostragem. Tais métodos consistem na eliminação de casos da classe majoritária e replicação (ou geração sintética) de casos da classe minoritária dos dados de treinamento visando obter classificadores melhores do que os obtidos a partir da distribuição original. Um método citado frequentemente na literatura para esse fim é o SMOTE (3).

O SMOTE funciona utilizando um algoritmo de k-vizinho mais próximo para criar dados sintéticos. O SMOTE primeiro começa escolhendo dados aleatórios da classe minoritária e, em seguida, são definidos k-vizinhos mais próximos dos dados. Os dados sintéticos seriam então feitos entre os dados aleatórios e o vizinho mais próximo k selecionado aleatoriamente. O procedimento é repetido várias vezes até que a classe minoritária tenha a mesma proporção que a classe majoritária.

2.1 – Gradient Boosting

Muitos algoritmos produzem modelos mais simples, mas a precisão normalmente também tem um desempenho inferior ao desejável (4). Portanto, é possível construir uma combinação linear dos modelos retornados por este algoritmo mais simples. Esta técnica é chamada de *boosting*, no qual tenta aproveitar a união de mais de um algoritmo mais simples para obter respostas diferentes. Assim, é possível gerar um modelo mestre mais preciso, baseado nas respostas combinadas de diferentes modelos mais simples e diferentes amostras de entrada (5).

O algoritmo usado neste trabalho foi *gradiente boosting*, no qual executa uma descida gradiente em uma função potencial. Mas alguns problemas aparecem no algoritmo que usa gradiente descendente, então a função potencial deve ser escolhida com cuidado, dependendo do que você deseja minimizar (5).

A Figura 1 mostra como é a iteração do *gradiente boosting*. Um primeiro modelo é criado com uma aproximação simples. Dessa aproximação, temos o resíduo, que é a distância entre o que foi previsto e o valor real. Um próximo modelo é criado e ajustado em cima do resíduo gerado pelo modelo anterior. Em seguida, um novo resíduo é calculado com esse segundo modelo. Essas iterações são repetidas por um determinado número de vezes, buscando minimizar o resíduo gerado pelos modelos, ou seja, até que a distância entre o previsto e o valor real seja o menor possível. O modelo final é a soma dos ajustes de todos os modelos.

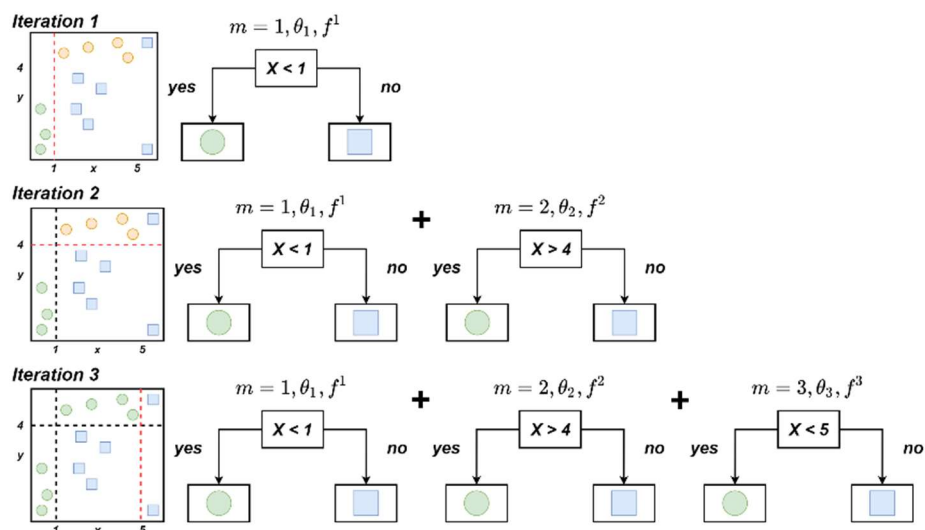


Figura 1 - Um exemplo da iteração do algoritmo de *gradiente boosting* adaptado de (6).

2.1 – Random Forest

O algoritmo de *random forest* é um modelo que aplica as técnicas de aprendizado por agrupamento com as árvores de decisões, foi criado por (7). O algoritmo é a combinação das previsões de diversas árvores de decisões onde cada árvore foi gerada através de um conjunto independente de vetores aleatórios, como demonstrado na Figura 2. Sendo que a seleção do conjunto de atributos para cada árvore é dada utilizando aleatoriedade, *boosting* ou *bagging*. Foi demonstrado que a floresta aleatória é um dos algoritmos mais precisos dentro da área de aprendizado de máquina, oferecendo a possibilidade de paralelização, capacidade de escalar e consegue lidar com dados de alta dimensionalidade (8).

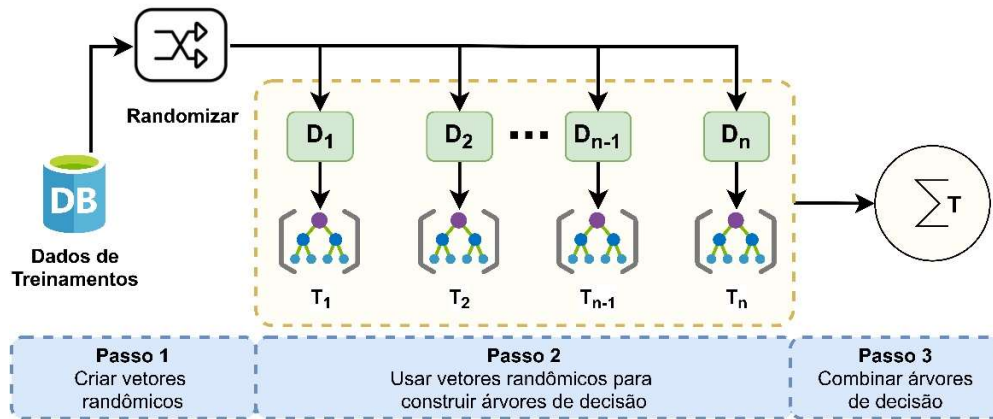


Figura 2 - Representação de uma *Random Forest* adaptado de (9).

3.0 - METODOLOGIA

A abordagem utilizada na solução do projeto é separada em três etapas, sendo o primeiro o pré-processamento, seguido da análise de dados, modelagem e a validação dos experimentos. Em outras palavras, a primeira etapa trata os dados brutos para que possam ser realizadas as análises das variáveis no momento de operação, desta forma facilita a identificação de *outlier* para remoção. A segunda etapa faz as análises mais refinadas, identificando padrões e simplificando a massa de dados, formando então a base de dados com as condições anteriores as falhas/alarmes e condições saudáveis de operação. Por fim, a última apresenta a abordagem de apoio à decisão para manutenção preditiva com diversos experimentos usando modelos de *machine learning*. A Figura 3 apresenta o esquema da abordagem.

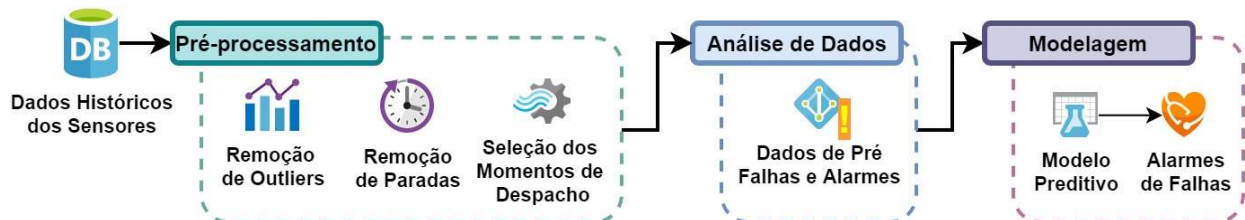


FIGURA 3 - Diagrama das Etapas de Desenvolvimento dos Modelos de *Machine Learning*.

Utilizando a abordagem citada anteriormente, foram desenvolvidos 4 modelos de classificação para cada unidade geradora, os quais buscam classificar as falhas relacionadas aos componentes do sistema de injeção, aos componentes do sistema de lubrificação, ao turbocompressor e ao trocador de calor do ar de admissão (*intercooler*). Totalizando 160 modelos a serem executados em tempo real.

3.1 – Modelagem manutenção preditiva

A manutenção preditiva é uma metodologia baseada no monitoramento e inspeção de máquinas e equipamentos para avaliar seus indicadores de qualidade. Utiliza os dados para auxiliar o gestor para tomar decisões rápidas e de forma certa, antecipando a possíveis desgastes de peças e a improdutividade operacional.

O modelo preditivo é responsável pela identificação das características das falhas, assim como simular novas, e classificar a condição da unidade geradora. Como pode ser visto pela Figura 4, os dados que antecedem falhas e alarmes são separados e é realizada uma análise estatística aplicando o desvio padrão e a mediana nas variáveis alvo. As entradas escolhidas alimentam um algoritmo, o qual simula pequenas variações e gera novas falhas. Com esses dados, o modelo é separado em treinamento e teste. O resultado alimenta um segundo algoritmo que utiliza

do método de classificação. Quando a falha é identificada pelo modelo, um alarme é gerado mostrando a probabilidade da ocorrência.

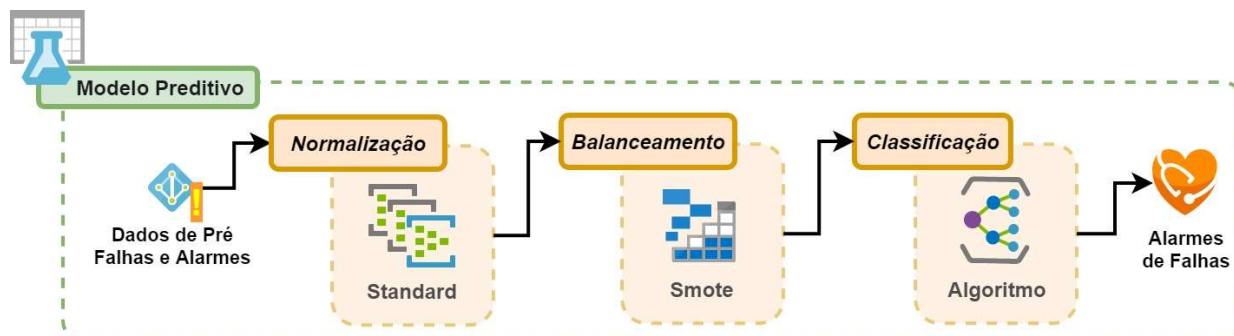


FIGURA 4 - Diagrama da Abordagem para o Modelo Preditivo.

Os modelos de classificação foram definidos com o propósito de identificar falha no comportamento de componentes dos subsistemas dos motores. Assim, foram desenvolvidos 4 modelos de classificação, os quais buscam classificar as falhas relacionadas aos componentes a seguir:

- **Sistema de Injeção:** análise de eventos em bombas injetoras e vazamento de combustível;
- **Sistema de Resfriamento:** análise de ocorrências no sistema de resfriamento do ar de admissão;
- **Sistema de Lubrificação:** análise de eventos de cárter, trocador de calor de óleo lubrificante e filtros de óleo lubrificante;
- **Sistema de Exaustão:** análise de eventos relacionados à manutenção do turbocompressor e vazamento de ar de admissão e gases de exaustão.

4.0 - RESULTADOS E DISCUSSÕES

Nesta seção vamos apresentar os resultados obtidos e suas devidas análises voltadas para atender na operação de geração de energia em relação à manutenção preditiva.

Foram realizados alguns experimentos para classificar falhas e eventos ocorridos, os algoritmos de *Gradiente Boosting* (GB), *Random Forest* (RF) e *Multilayer Perceptron* (MLP) foram avaliados. Devido ao desbalanceamento das classes utilizou-se o SMOTE como alternativa de melhorar a classificação. Também foi utilizado o normalizador *StandardScaler* (SC) em algumas situações. Todos os experimentos utilizam a aplicação de *StratifiedKFold* para *cross validation*, no qual garante elementos de todas as classes para treino e teste em cada *fold*. Foi definido 5 *fold* para a análise. A abordagem calcula a mediana, o desvio padrão e relações entre os sensores por campanha e utiliza na entrada dos modelos.

Como métricas de avaliação foi usado o *F1-score*, *Precision*, *Recall* e *Accuracy*. Sendo que, devido a característica do problema ser de classes desbalanceadas, foi usado a versão Macro das métricas, no qual calcula as métricas para cada rótulo e encontra sua média não ponderada, não levando em consideração o desequilíbrio do rótulo. Sendo assim, os resultados são mostrados a seguir.

Para classificação de eventos e falhas, foram analisadas 8 falhas, distribuídas entre os modelos de acordo com o subsistema relacionado. O número total de campanhas ao longo de 2 anos para 40 motores é de 3869, onde 2444 são campanhas saudáveis. Das restantes, encontram-se eventos de diversas naturezas, das quais foram filtradas para cada modelo. Todos os modelos recebem as campanhas de classe "saudável".

Modelo para Sistema de Injeção: As classes de falhas para este modelo são de manutenção do sistema de injeção (presença de vazamento de combustível ao longo do processo ou reparo de bomba injetora), com 110 ocorrências, e inspeções no sistema, com 9 ocorrências. As variáveis utilizadas na classificação destas falhas são a potência gerada, a temperatura e a pressão do combustível, a temperatura de exaustão e a diferença de temperatura do óleo lubrificante na entrada e no interior dos cilindros. Os resultados podem ser vistos na Tabela 01.

TABELA 1 - Resultados da Classificação para o Sistema de Injeção.

Modelos	F1	Recall	Precision	Acurácia
GB & GB/SC	34.45 ± 2.72	34.14 ± 1.77	37.01 ± 7.22	94.47 ± 2.39
GB/SMOTE	33.47 ± 1.57	36.88 ± 3.30	34.35 ± 0.89	82.82 ± 3.96

GB/SC/SMOTE	33.23 ± 1.66	36.20 ± 4.25	34.15 ± 1.05	82.90 ± 3.33
RF & RF/SC	35.38 ± 2.71	34.80 ± 1.52	49.91 ± 1.68	96.36 ± 0.10
RF/SMOTE	36.17 ± 3.58	35.57 ± 3.00	37.71 ± 4.79	94.47 ± 1.12
RF/SC/SMOTE	35.53 ± 2.53	34.80 ± 1.58	42.23 ± 1.33	95.34 ± 0.98
MLP/SC/SMOTE	36.32 ± 3.63	35.51 ± 2.94	43.24 ± 13.14	94.31 ± 2.33

Modelo para Sistema de Lubrificação: As classes de falhas para este modelo são de manutenção no cárter (vazamento de óleo lubrificante e reparos), com 67 ocorrências, e manutenção no filtro de lubrificante (vazamento de lubrificante no filtro e substituição de filtro), com 308 ocorrências. As variáveis utilizadas na classificação destas falhas foram a temperatura e pressão do óleo lubrificante após o filtro, no interior dos cilindros, na entrada e na saída do turbocompressor e na entrada e saída do trocador de calor. Os resultados podem ser vistos na Tabela 02.

TABELA 2 - Resultados da Classificação para o Sistema de Lubrificação.

Modelos	F1	Recall	Precision	Acurácia
GB & GB/SC	40.22 ± 3.96	39.05 ± 3.03	44.61 ± 6.12	89.23 ± 1.90
GB/SMOTE	41.19 ± 2.91	50.43 ± 4.50	40.64 ± 1.97	73.58 ± 2.34
GB/SC/SMOTE	42.82 ± 3.09	53.25 ± 5.32	41.44 ± 1.95	74.87 ± 2.37
RF & RF/SC	37.66 ± 2.85	36.80 ± 1.84	46.32 ± 9.13	90.09 ± 1.20
RF/SMOTE	44.36 ± 3.65	45.24 ± 3.34	44.44 ± 4.69	85.82 ± 2.52
RF/SC/SMOTE	42.95 ± 1.27	43.62 ± 2.91	43.10 ± 3.02	85.71 ± 1.27
MLP/SC/SMOTE	38.81 ± 3.83	45.75 ± 3.56	38.58 ± 2.41	76.25 ± 7.54

Modelo para Sistema de Exaustão: A classe de falha analisada neste modelo é de manutenção nos turbocompressores dos motores. Foram registradas 192 ocorrências de reparos e substituições. As variáveis utilizadas na classificação deste modelo foram a potência gerada, a velocidade de rotação dos turbocompressores, a temperatura e pressão do ar de admissão após a compressão e a diferença de temperatura do gás de exaustão antes e depois da expansão. Os resultados podem ser vistos na Tabela 03.

TABELA 3 - Resultados da Classificação para o Sistema de Exaustão.

Modelos	F1	Recall	Precision	Acurácia
GB & GB/SC	52.87 ± 3.06	52.33 ± 1.87	59.51 ± 7.82	93.89 ± 0.17
GB/SMOTE	58.61 ± 1.75	70.26 ± 5.04	57.34 ± 1.39	83.05 ± 0.92
GB/SC/SMOTE	59.00 ± 1.51	70.73 ± 4.85	57.64 ± 1.15	83.32 ± 2.04
RF & RF/SC	50.46 ± 1.74	50.95 ± 0.87	60.58 ± 12.68	94.35 ± 0.09
RF/SMOTE	61.25 ± 2.64	63.05 ± 5.27	60.35 ± 1.72	90.87 ± 0.92
RF/SC/SMOTE	60.86 ± 2.42	61.72 ± 4.63	60.98 ± 1.83	91.41 ± 1.57
MLP/SC/SMOTE	60.01 ± 1.50	66.25 ± 2.65	58.12 ± 1.17	87.11 ± 0.93

Modelo para Sistema de Resfriamento: As classes de falha para este modelo são relacionadas à manutenção do *intercooler*, com 54 ocorrências, vazamento de ar de admissão, com 30 ocorrências e vazamento de água de refrigeração, com 19 ocorrências. As variáveis utilizadas na classificação deste modelo foram a potência gerada, as temperaturas na entrada e saída do *intercooler* do ar de admissão e das águas de baixa temperatura (LT) e de alta temperatura (HT). Os resultados podem ser vistos na Tabela 04.

TABELA 4 - Resultados da Classificação para o Sistema de Resfriamento.

Modelos	F1	Recall	Precision	Acurácia
GB & GB/SC	24.39 ± 0.15	24.65 ± 0.28	24.13 ± 0.03	95.21 ± 1.14
GB/SMOTE	27.19 ± 5.18	32.52 ± 13.37	27.02 ± 3.24	85.11 ± 3.44
GB/SC/SMOTE	27.66 ± 3.12	33.32 ± 8.61	27.33 ± 1.97	84.27 ± 4.52
RF & RF/SC	24.66 ± 0.01	25.00 ± 0.00	24.14 ± 0.03	96.55 ± 0.11
RF/SMOTE	28.66 ± 6.98	28.07 ± 5.60	29.79 ± 9.44	95.05 ± 1.00
RF/SC/SMOTE	28.59 ± 5.80	28.21 ± 5.02	29.19 ± 6.89	94.89 ± 0.70
MLP/SC/SMOTE	24.47 ± 0.09	24.81 ± 0.19	24.14 ± 0.01	95.84 ± 0.06

A maior dificuldade encontrada nos modelos de classificação foi quanto a confiabilidade nos registros de eventos, dos quais não faziam distinção clara de manutenções corretivas e planejadas, além de datas de eventos imprecisas. Sendo assim, foi realizado uma padronização e um estudo de categorização dos eventos e alarmes, desta forma, foi possível dar seguimento para classificação de falhas.

Em modo geral os modelos obtiveram bons resultados para os subsistemas e equipamentos selecionados para a análise. Os modelos que utilizam o SMOTE para balanceamento obtiveram os melhores resultados, tendo como destaque os algoritmos de RF que sobressaíram dentre os demais. Entretanto, os algoritmos testados podem ser melhorados aplicando regras de negócio para definir os limiares de acionamento dos alarmes, reduzindo assim os falsos positivos. Vale ressaltar a possibilidade de realizar estudos mais aprofundados para determinados subsistemas, como de resfriamento, para aumentar a eficiência da predição.

Os modelos foram implementados em ambiente de produção e testados. A Figura 5 apresenta um exemplo de exceção dos modelos durante a operação em tempo real. Percebe-se que os modelos reagiram aproximadamente em 7 dias antes do registro da identificação da falha, sendo assim, é possível antecipar e/ou disparar algum processo para mitigar o risco de quebra do equipamento e até mesmo uma manutenção programada.

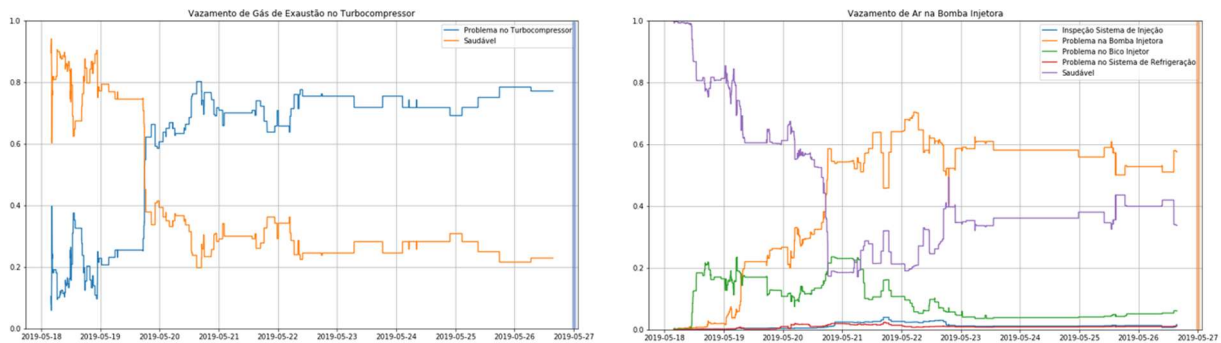


Figura 5 – Resultado dos modelos do sistema de exaustão e de injeção para UGD 13 e 26 respectivamente.

5.0 - CONCLUSÕES

Este artigo apresentou o resultado de uma pesquisa para classificação de falhas e eventos baseada em modelos de *machine learning* para usinas termelétricas, que corresponde a uma proposta de arquitetura em tempo real para manutenção preditiva aplicada na Indústria 4.0. A arquitetura tenta lidar com modelos de ML, como algoritmos para classificação de falhas em tempo de operação.

Uma das contribuições deste artigo é a implementação dos modelos de classificação utilizando a arquitetura presente para execução dos modelos de ML em tempo real, que pode ser aplicada a qualquer setor industrial que utilize supervisórios do tipo SCADA, podendo também ser adaptada, expandida e evoluída para outras tecnologias de geração como: outras Termelétricas, pequenas centrais hidrelétricas, eólicas e term nucleares.

Os próximos passos deste trabalho se referem a melhoria na fase de extração de variáveis latentes, utilizando outras abordagens como por exemplo, análises de sinais para classificação de falhas, aplicar novos experimentos com outros algoritmos e estudar novos métodos para redução de falsos positivos e negativos em geral.

Os autores agradecem à Centrais Elétricas da Paraíba (EPASA) pelo apoio financeiro ao Projeto PD-07236-0010-2020 - Digital Twin de Apoio à Tomada de Decisão Operacional de Usina Termelétrica a Motores de Combustão, desenvolvido no âmbito do programa de Pesquisa e Desenvolvimento da Agência Nacional de Energia Elétrica (P&D ANEEL), que foi executado pela empresa de engenharia Radix Engenharia e Software S/A.

6.0 - REFERÊNCIAS BIBLIOGRÁFICAS

- (1) SANG, G., XU, L., DE VRIEZE, P. T., BAI, Y., AND PAN, F. Predictive maintenance in industry 4.0. International Conference on Information Society and Technology, 2020.
- (2) MOTAGHARE, O., PILLAI, A. S., AND RAMACHANDRAN, K. Predictive maintenance architecture. In 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), pp. 1–4. IEEE, 2018.
- (3) CHAWLA, N. V., BOWYER, K. W., HALL, L. O., AND KEGELMEYER, W. P. Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16:321–357, 2002.
- (4) DUFFY, N. AND HELMBOLD, D. Boosting methods for regression. Machine Learning, 47(2):153–200, 2002.
- (5) ZEMEL, R. S. AND PITASSI, T. A gradiente-based boosting algorithm for regression problems. Advances in neural information processing systems, pp. 696–702, 2001.

- (6) ZHANG, Z., MAYER, G., DAUVILLIERS, Y., PLAZZI, G., PIZZA, F., FRONCZEK, R., SANTAMARIA, J., PARTINEN, M., OVEREEM, S., PERAITA-ADRADOS, R., ET AL. Exploring the clinical features of narcolepsy type 1 versus narcolepsy type 2 from european narcolepsy network database with machine learning. Scientific reports, 8(1):1–11, 2018.
 - (7) HO, T., 1995, "Random decision forests". In: Analysis and Recognition, 1995., Proceedings of the.
 - (8) CARUANA, R., KARAMPATZIAKIS, N., 2008, "An empirical evaluation of supervised learning in high dimensions". In: on Machine learning. pp. 96–103.
 - (9) TAN, P., STEINBACH, M., KUMAR, V., 2005, Introduction to data mining. S.I., AddisonWesley Longman Publishing Co., Inc.
- DADOS BIOGRÁFICOS**



Mestre em Engenharia de Sistemas e Computação pela UFRJ (2018), graduado em Ciências da Computação pela UFRJ (2014) e Design Gráfico pela UNESA (2008). Atuou como designer gráfico e desenvolvedor full stack desde 2009. Foi pesquisador bolsista no IPEA, atuando diretamente no design, modelagem e nas ferramentas de análises de dados. Professor de algoritmos, Machine Learning e Data Mining na pós-graduação no Instituto Infnet. Atualmente é líder técnico de ciência de dados no grupo de P&D da Radix, desenvolvendo atividades e projetos de inovação com inteligência artificial, criando soluções para diversos setores da indústria.

- (2) ROBSON FELIPE VIANA DA SILVA, graduou-se em Engenharia Mecânica na UFF (2019) onde está cursando mestrado em Engenharia Mecânica com ênfase em transferência de massa e energia (2022). Seus interesses são estudos de manutenção preditiva e otimização de sistemas. Atua em projetos de P&D no setor de energia elétrica.
- (3) BRUNO DEON, é atualmente aluno de mestrado do Programa de Pós-Graduação em Engenharia Química da UFSC, onde também recebeu bacharel em Engenharia Química (2019). Nos últimos dois anos, sua experiência profissional foi focada na ciência de dados, onde a partir da utilização de algoritmos de inteligência artificial desenvolveu ferramentas para o auxílio à tomada de decisão e predição de falhas no setor industrial.
- (4) CAMILLA BARROS BATISTA, graduou-se em Engenharia Química na UERJ (2017) onde está cursando o mestrado em Otimização de Processos desde 2021. Tem experiência na área de projetos de engenharia no setor de óleo e gás e em projetos de P&D da ANEEL. Seus principais interesses são estudos de novas tecnologias na área de geração, transmissão e distribuição de energia.
- (5) GABRIELA TEIXEIRA JUSTINO, engenheira química com mestrado em Tecnologia de Processos Químicos e Bioquímicos. Possui experiência em otimização, análise de dados, modelagem matemática.
- (6) GABRIELA CARVALHO FREITAS, graduada em Engenharia Química pela UFRJ (2020). Atua como cientista de dados na Radix Engenharia e Desenvolvimento de Software, onde desenvolve projetos utilizando inteligência artificial aplicada em soluções para problemas de engenharia.
- (7) FLÁVIO LEITE LOUÇÃO JUNIOR, graduou-se em Engenharia Química pela UFRJ (2015) onde finalizou mestrado em Otimização de Processos no ano de 2016. Atua em projetos de P&D para o setor elétrico desde 2017 e atualmente gerencia um grupo voltado à formatação e execução de novos P&Ds no contexto da ANEEL. Seus principais interesses consistem no estudo da aplicação novas tecnologias e metodologias de gestão e execução de pesquisas para o desenvolvimento de melhorias e solução de problemas no setor elétrico como um todo.
- (8) ATHOS DOS SANTOS BARBOS, graduou-se em Engenharia de Controle e Automação no CEFET/RJ (2019). Tem experiência na área de desenvolvimento de software voltado para projetos de engenharia ligados ao setor de energia e sistemas PIMS, principalmente da OSIsoft.
- (9) ANGELO MARCELINO CORDEIRO, graduou-se em Ciências e Tecnologia na UFRN (2019) onde está cursando o bacharelado em Engenharia de Computação. Seus interesses são nas áreas de ciência de dados, com ênfase em inteligência artificial e aprendizagem profunda. Atualmente, atua em projetos de P&D no setor de energia elétrica.
- (10) ALEX MORETTI DE MORAIS, pós-graduando em Ciências de Dados e BigData pela PUC Minas e graduação pela USF (2008) e UFSCAR (1996). Possui 25 anos de experiência e especialização em tecnologia da informação, com ênfase em infraestrutura e datacenter T3, desses, 12 anos no setor elétrico. Atuou como Gestor Latam na Stefanini IT Solutions, Gestor de Serviços de TI na CPFL Energia e atualmente como Gestor de Infraestrutura Corporativa na EPASA.

(11) ISAAC EMMANUEL AZEVEDO DE MEDEIROS, graduado em Engenharia Elétrica pela UFPB. Atualmente, cursa Pós-Graduação em Ciência de Dados e Big Data pela PUC Minas e mestrado em Engenharia Elétrica pela UFPB. Desde 2019 atua na indústria, ocupou o cargo de Analista no ramo de Automação Industrial. Atualmente é Engenheiro Trainee da EPASA, atuando no departamento de engenharia e em projetos P&D nas áreas de Inovação.

(12) RODRIGO JOSÉ SILVA DE ALMEIDA, possui graduação em Automação Industrial pelo IFPB (2011) e graduação em Engenharia Elétrica pela UNINASSAU (2019). Atualmente está cursando Ciência de Dados e Big Data na PUC Minas. Trabalha na EPASA, desde 2010 e atualmente tem o cargo de especialista em usina termoeletrica II.

(13) CARLOS ANTONIO ALVES DE ARAÚJO JUNIOR, graduado em Engenharia Elétrica pela UFCG e Mestre em Sistemas Eletrônicos e Automação pela UFPB. Tem experiência como Engenheiro de Controle e Automação, atuando em projetos de manutenção da automação e melhorias em plantas industriais. Atualmente trabalha na EPASA como Especialista em Termoeletrica II, onde realiza melhorias no processo de geração de energia elétrica, implantação de novos projetos (CAPEX), gestor técnico de projetos P&D Aneel, avaliação de parâmetros de proteção e contato com ONS referente a SGR, SIPER, RANOR e OTRS.