Universidade Federal de Campina Grande - UFCG
Centro de Engenharia Elétrica e Informática - CEEI
Coordenação de Pós-Graduação em Ciência da Computação

Predição de Desligamentos de Motores de uma Usina Termoelétrica Baseada no Histórico de Eventos

Bruno Rafael Araújo Vasconcelos

ORIENTADORA

Joseana Macêdo Fechine Régis de Araújo, DSc.

Campina Grande
Fevereiro – 2020

V331p Vasconcelos, Bruno Rafael Araújo.

Predição de desligamento de motores de uma usina termoelétrica baseada no histórico de Eventos / Bruno Rafael Araújo Vasconcelos. - Campina Grande, 2020.

127f. : il. Color.

Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2020.

"Orientação: Profa. Dra. Joseana Macêdo Fechine Régis de Araújo".

Referências.

1. Usina Termoelétrica - Motores. 2. Detecção de Falhas. 3. Sistemas de Alarmes. 4. Predição de Desligamentos. I. Araújo, Joseana Macêdo Fechine de. II. Título.

CDU 321.363(043)

FICHA CATALOGRAFICA ELABORADA PELA BIBLIOTECARIA ITAPUANA SOARES DIAS CRB-15/93

PREDIÇÃO DE DESLIGAMENTOS DE MOTORES DE UMA USINA TERMOELÉTRICA BASEADA NO HISTÓRICO DE EVENTOS

BRUNO RAFAEL ARAUJO VASCONCELOS

DISSERTAÇÃO APROVADA EM 20/02/2020

JOSEANA MACÊDO FECHINE RÉGIS DE ARAÚJO, Dra., UFCG
Orientador(a)

ELMAR UWE KURT MELCHER, Dr., UFCG
Examinador(a)

JOSÉ EUSTAQUIO RANGEL DE QUEIROZ, Dr., UFCG
Examinador(a)

KATIA ELIZABETE GALDINO, Dra., UEPB
Examinador(a)

ADRIANO ARAÚJO SANTOS, Dr., UNIFACISA
Examinador(a)

CAMPINA GRANDE - PB

Universidade Federal de Campina Grande - UFCG
Centro de Engenharia Elétrica e Informática - CEEI
Coordenação de Pós-Graduação em Ciência da Computação

Predição de Desligamentos de Motores de uma Usina Termoelétrica Baseada no Histórico de Eventos

Bruno Rafael Araújo Vasconcelos

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande - Campus I, como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciência da Computação.

ORIENTADORA

Joseana Macêdo Fechine Régis de Araújo, DSc.

Área de Concentração: Ciência da Computação
Linha de Pesquisa: Reconhecimento de Padrões / Aprendizagem de Máquina

Campina Grande
Fevereiro – 2020

“A sabedoria é a coisa principal.
Adquire pois a sabedoria, emprega tudo o que
tiveres na aquisição do entendimento.” (Provérbios 4:7)

Dedico esta pesquisa, primeiramente, ao DEUS criador dos céus e da terra, que me permitiu chegar até aqui. Segundo, a todos aqueles que estiveram ao meu lado, em especial, aos meus pais e a minha esposa que me ajudaram nos momentos difíceis, me impulsionando a prosseguir.

Agradecimentos

Agradeço, inicialmente, a DEUS, por toda a providência durante esta caminhada, por estar comigo nos momentos de dificuldade, dúvida e ansiedade.

A minha família, principalmente aos meus pais Luciano e Lucinéia, com todos os ensinamentos e conselhos que me tornaram uma pessoa melhor e isso tem refletido em toda a minha caminhada profissional e pessoal. Aos meus irmãos, que colaboraram direta ou indiretamente e a minha esposa Bianca, a qual com seu companheirismo me deu força e me ajudou com conselhos desde o início da minha trajetória, ainda na graduação.

Agradeço à professora Joseana, pela oportunidade que me concedeu, por todo o conhecimento científico transmitido, questionamentos e conselhos, que me fizeram evoluir. Nesta etapa, a qual concluo, sigo caminhando para as próximas, sempre me inspirando no exemplo de mestre e profissional da educação que é a professora Joseana.

Aos meus colegas do laboratório LInCE/UFCG, ao professor Adriano, pelos conselhos, pelo acompanhamento da pesquisa durante a ausência da professora e pelas longas conversas para encontrarmos as melhores saídas para os problemas encontrados. A Nicolas, por muitas vezes me auxiliar em outras atividades, para que me dedicasse mais ao mestrado. A Gabriel que, durante seu estágio na usina, pôde compartilhar informações importantes, que me ajudaram a validar os resultados obtidos.

A todos da usina que colaboraram para a realização da pesquisa.

À CAPES, pelo financiamento da pesquisa e à COPIN, por torná-la possível.

Enfim, meu agradecimento a todos que, direta ou indiretamente, contribuíram para minha formação e para a conclusão do meu mestrado.

Resumo

A identificação de falhas em motores de plantas industriais possui grande valor para as empresas, pois tem o objetivo de evitar consequências, tais como: a queima ou danificação de equipamentos, a morte de operários e catástrofes ambientais, dentre outros. Os motores utilizados em usinas termoelétricas são retirados de operação (desligados) sempre que uma falha grave acontece. Para cada mau funcionamento de um motor, no processo de geração de energia, um alarme é emitido para uma central de controle, o qual é mantido em um histórico de eventos. Cada alarme está associado a uma única falha e, portanto, uma sequência de alarmes pode ser vista como uma sequência de falhas. Identificar quando os desligamentos ocorrerão pode ajudar os operadores a evitá-los, através de correções de forma antecipada, para diminuir perdas no processo de produção. Diante deste cenário, a presente pesquisa objetivou, a partir do histórico de alarmes, extrair características e estruturá-las para treinar um modelo de predição e prognóstico. O modelo, construído, consiste em um modelo de aprendizagem de máquina. A abordagem proposta apresentou resultados satisfatórios. Do total de 57 casos válidos em que houve desligamentos, a técnica proporcionou 48 acertos. Dos 507 casos em que não houve desligamentos, o modelo acertou 390 casos, considerando um limiar de predição de 0,5 e técnica de validação cruzada (k-fold, com k = 10). A abordagem para predição de desligamentos possibilitou, portanto, realizar prognósticos nos motores de uma termoelétrica, prevendo com antecedência os desligamentos. O número de falsos negativos mostra que o modelo pode apresentar resultados significativos quando for treinado com uma quantidade maior de exemplos de desligamento. Assim, a abordagem proposta se mostrou viável para a previsão de desligamentos em motores de uma usina termoelétrica.

Palavras-chave: Usina Termoelétrica, Detecção de Falhas, Sistema de Alarmes, Predição de Desligamentos.

Abstract

The identification of failures in industrial plant engines has great value for companies, as it aims to avoid consequences, such as: the burning or damage of equipment, the death of workers and environmental catastrophes, among others. The motors used in thermoelectric plants are taken out of operation (turned off) whenever a serious failure occurs. For each engine malfunction, in the power generation process, an alarm is sent to a control center, which is kept in an event history. Each alarm is associated with a single fault and, therefore, a sequence of alarms can be seen as a sequence of failures. Identifying when shutdowns will occur can help operators avoid them, through corrections in advance, to reduce losses in the production process. Given this scenario, the present research aimed, from the history of alarms, to extract characteristics and structure them to train a prediction and prognosis model. The model, built, consists of a Machine Learning Model. The proposed approach showed satisfactory results. Of the total of 57 valid cases in which there were disconnections, the technique provided a rate for the context, of 48 hits. Of the 507 in which there were no disconnections, the model got 390 cases right, considering a prediction threshold of 0.5 and using the cross-validation technique (k-fold, with k = 10). The approach to predicting shutdowns therefore made it possible to make prognoses on the engines, anticipating shutdowns in advance. The results presented here, the number of false negatives shows that the model can present significant results when trained with a greater number of examples of disconnection. Thus, the proposed approach proved to be viable for predicting engine shutdowns at a thermoelectric plant.

Keywords: Thermoelectric Plant, Fault Detection, Alarm System, Shutdown Prediction.

Lista de Siglas e Abreviaturas

AADA	-	Análise Automática de Alarmes
ACLP	-	Algoritmo para o cálculo do limiar de predição
ADP	-	Algoritmo para a Identificação de Padrões
AEP	-	Algoritmo para a extração de períodos de operação
AGMP	-	Algoritmo para a Geração de Métricas e Predição
AID	-	Alarme Indicador de Desligamento
ASTT	-	Algoritmo para a Separação dos conjuntos de Treino e Teste
ETL	-	Extração, Transformação e Carregamento
FDDC	-	*Fault Detection, Diagnosis and Correction*
IQR	-	Intervalo Interquartil
MCR	-	Matriz de Características
SPM	-	*Sequential Pattern Mining*
UTE	-	Usina Termoelétrica
VCR	-	Vetor de Características
MGMP	-	Módulo de Geração de Métricas e Predição
MIP	-	Módulo de Identificação de Padrões

Lista de Figuras

Lista de Quadros

Lista de Tabelas

Sumário

Capítulo 1

Considerações Iniciais

Neste capítulo, serão apresentadas as considerações iniciais, bem como o problema que motivou o estudo e uma visão geral da abordagem proposta para predição de desligamentos abruptos, baseada no histórico de eventos emitidos por motores de uma usina termoelétrica.

1.1 Introdução

A identificação de problemas em motores de plantas industriais é de grande importância para evitar a ocorrência de distúrbios no sistema, pois um defeito pode causar perda de desempenho no processo de produção e também consequências graves, tais como: a queima ou danificação de equipamentos, a morte de operários, catástrofes ambientais, prejuízos financeiros, dentre outros. A detecção e diagnóstico de falhas com rapidez e antecedência são essenciais para uma operação confiável, segura, eficiente e para a manutenção da qualidade de produção da planta. De acordo com Sartori et al. (2012), a falha representa qualquer desvio em relação aos requisitos de operação confiável do equipamento definidos pelo fabricante ou pelo processo, podendo ou não afetar sua capacidade de desempenhar uma função requerida. As falhas podem ocorrer no processo, nos sensores, nos atuadores e nos instrumentos, de forma independente ou simultânea.

Para minimizar os possíveis erros na detecção, análise e correção de falhas, no contexto de uma termoelétrica, a empresa fabricante dos motores fornece, juntamente com os equipamentos, um sistema útil para o controle, monitoramento e gerenciamento dos equipamentos, através da geração de dados sobre o estado interno das máquinas, e para segurança, através do sistema de alarmes formado por um conjunto de hardware e software. O sistema deve ser capaz de informar, por meio auditivo ou visual, uma condição anormal do processo

industrial ou dos equipamentos (NOYES, 2002; IZADI et al. 2009a; IZADI et al. 2009b; IZADI et al. 2009c), para que seja possível detectar, diagnosticar e corrigir falhas (*Fault Detection, Diagnosis and Correction - FDDC*) o mais rápido possível, a exemplo do que é destacado também em Willsky (1976), Frank (1996), Isermann (1997), Venkatasubramanian (2003a; 2003b; 2003c) e Sartori (2012), dentre outros.

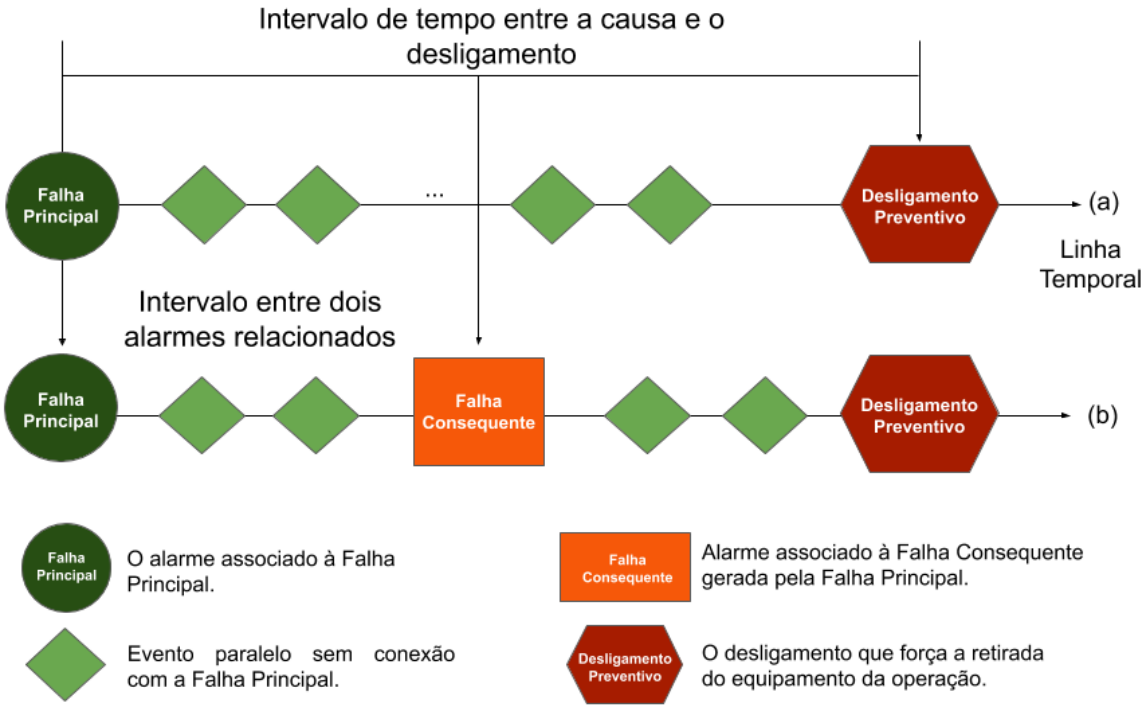
1.2 Motivação e Justificativa para a Pesquisa

Muitos problemas que ocorrem nos sistemas de alarmes industriais foram apontados na literatura ao longo dos anos (SARTORI et al., 2012). Dentre os principais problemas indicados, estão as inundações (HU et al., 2018), as emissões excessivas de falsos alarmes (ROTHENBERG, 2003) e alarmes desnecessários (NOYES, 2002), a falta de manutenção do sistema de alarmes (BRANSBY et al., 1998), má visualização das informações (HSE, 1997), insuficiência de informações no processo de apoio à decisão e até problemas na própria gestão e controle do sistema de segurança (MAKI et al., 1997).

Na usina termoelétrica (UTE), local onde a pesquisa ora descrita foi realizada, os problemas supracitados também podem ocorrer, em especial, a insuficiência de informações no apoio a decisões. Quando uma falha crítica acontece, o sistema de proteção da planta é acionado automaticamente para evitar consequências graves. Dentre as ações de proteção, estão os desligamentos automáticos, que interrompem os motores envolvidos, evitando a propagação da falha.

A título de exemplo, na Figura 1, são apresentadas duas situações que podem ocorrer no sistema de alarmes da UTE. Em (a), depois que a falha acontece, é desencadeada uma sequência de outras falhas com criticidade cada vez maior, até que o próprio sistema, de forma automática, retira os motores de operação. Nesta ocasião, o desligamento automático acontece. Na Figura 1 (b), a falha já foi grave o suficiente para acionar o desligamento logo em seguida.

Figura 1. Sequências de falhas que precedem a ação do sistema de segurança de uma planta industrial.

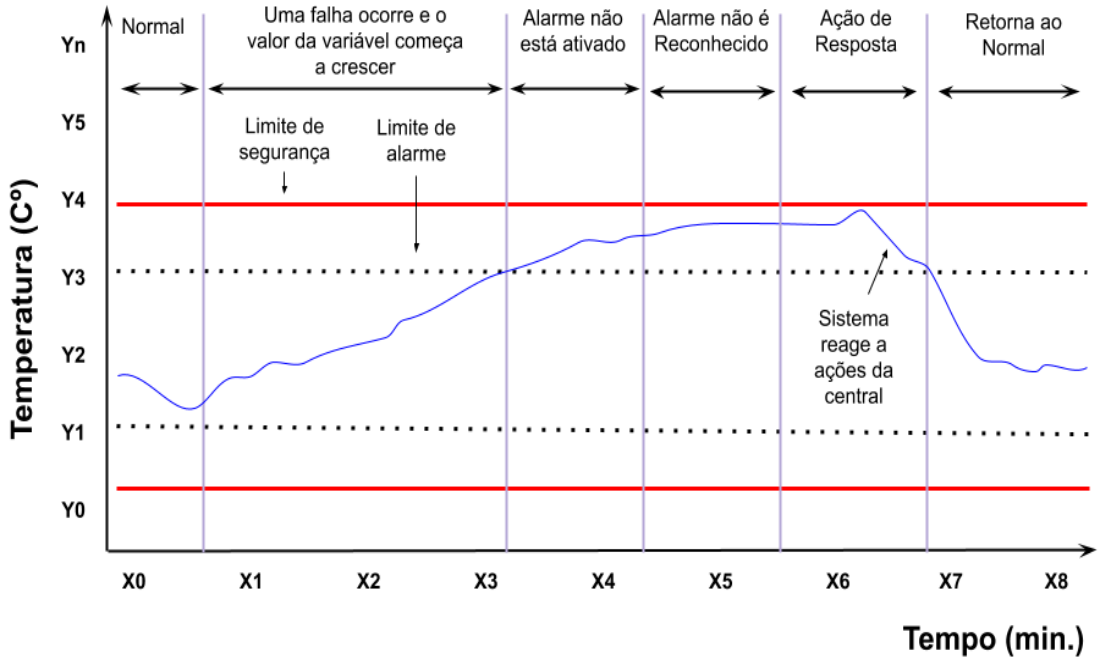


Fonte: Adaptado de Folmer et al. (2014).

Cada registro de alarme emitido está associado a um sensor do sistema interno do equipamento, que gera registros de temperatura, velocidade de rotação, tensão, etc. Essas grandezas físicas são chamadas variáveis de processo.

Quando algumas dessas variáveis ultrapassam os limites, superior ou inferior de falha (linhas tracejadas da Figura 2), um registro de alarme é emitido para a central de controle. Quando a falha ultrapassa o limite de segurança (linhas vermelhas da Figura 2), então, o equipamento, forçadamente, é retirado de operação.

Figura 2. Exemplo de variável do processo ao longo do tempo e seus limites de ativação.



Fonte: Autoria própria.

1.3 Identificação do Problema

Por mais que os desligamentos preventivos sejam necessários para a UTE, estes trazem prejuízos. Um equipamento retirado de operação atrasa o processo de geração, já que é necessário um tempo considerável para sua reativação, aliado aos demais custos de manutenção. No processo de FDDC, os operadores usam as informações dos registros de alarmes, que são emitidos ao longo do tempo, os gráficos das variáveis do processo e o conhecimento técnico para resolver os diversos problemas. Esta pode não ser a melhor forma de lidar com a situação, pois as informações geradas são referentes a diagnóstico e para resolver o problema pode ser necessário utilizar informações de prognóstico.

Os alarmes que ocorrem a partir da partida do motor são usados como informações de apoio na análise do operador. É importante saber quantas vezes os alarmes causaram o desligamento e em quanto tempo ocorrerá o desligamento após a emissão desses para realização de manutenção preventiva. Informações deste tipo, não são disponibilizadas pelo sistema de controle, e com isso, o processo de FDDC fica muito dependente do conhecimento dos operários, que

pode ser limitado ou insuficiente em determinadas situações. Isto demonstra uma limitação do sistema de alarmes da UTE e uma dependência do conhecimento tácito dos operadores no processo.

Embora alguns desligamentos não possam ser evitados por questões técnicas que só podem ser resolvidas se o motor estiver desligado, informações sobre se a sequência de alarmes atual é capaz de levar ao desligamento e quanto tempo isso demorou a acontecer, constituem aprendizado para operadores inexperientes que, em situações semelhantes ou após o desligamento do motor, saberão quais ações executar.

Outra questão importante a ser destacada, reside no fato de que, até certo ponto, os registros de alarmes e outros eventos contribuem como informação útil para identificar o motivo da falha, mas um número elevado pode tornar inviável a percepção dos verdadeiros efeitos, o que pode dificultar a avaliação dos operadores.

Os problemas e limitações supracitados, relacionados ao sistema de alarmes em plantas industriais, indicam a relevância de melhorar o processo de FDDC, principalmente em situações de risco.

1.4 Objeto de Estudo

No contexto da pesquisa, busca-se construir um modelo híbrido de predição, a partir de padrões extraídos das sequências de alarmes geradas por motores de uma termoelétrica, durante o processo de geração de energia, com o objetivo de classificar, considerando determinada probabilidade, se uma situação de falha levará ou não a um desligamento abrupto e quanto tempo será necessário para o fato ocorrer. Sempre que um novo alarme é emitido, o modelo usará todos os registros que ocorreram até aquele momento, para contar o número de vezes que aquela sequência ocorreu, o tempo médio entre esta e os desligamentos, e por último, gerar uma predição, que indicará se haverá ou não desligamento abrupto, baseado em um limiar encontrado a partir do histórico de alarmes.

1.5 Objetivos da Pesquisa

Nesta seção, será discutido o objetivo geral do estudo realizado. Em seguida, serão abordados os objetivos específicos que foram tratados no desenvolvimento da pesquisa.

1.5.1 Objetivo Geral

O principal objetivo desta pesquisa foi construir um modelo capaz de identificar padrões em sequências de eventos e realizar predição de desligamentos abruptos com o intuito de gerar informações de prognóstico sobre os desligamentos de motores, para auxiliar a central de controle de uma termoelétrica no processo de correção de falhas.

1.5.2 Objetivos Específicos

Com a finalidade de atingir o objetivo geral, os seguintes objetivos específicos se fizeram necessários:

1. Estudo do estado da arte sobre a identificação de padrões em sequências temporais de alarmes em plantas industriais e construção de modelos de predição de desligamentos usando algoritmos para reconhecimento de padrões, a partir da aprendizagem de máquina;

2. Estudo de algoritmos de pré-processamento de dados sobre o histórico de registros, para a remoção de dados espúrios;

3. Análise de técnicas de processamento para a extração do conjunto de características relevantes presentes no histórico de alarmes;

4. Estudo de algoritmos de geração de máquinas de estados para identificar associações entre os alarmes e o desligamento; e

5. Definição das técnicas para a identificação de padrões e mecanismo de predição, a partir das sequências de alarmes, conforme objetivo específico 3 e dos padrões descritos no objetivo específico 4.

1.6 Relevância da Proposta

Devido à necessidade de minimização de custos a partir da prevenção de desligamentos que afetam a produção de energia da UTE, torna-se relevante o uso de uma ferramenta, em conjunto com a experiência dos operadores, com a finalidade de prover soluções mais rápidas para resolver distúrbios nos motores de uma UTE. Com isto, os efeitos das falhas podem ser minimizados e os prejuízos reduzidos.

A partir dos dados de alarmes, são extraídos os padrões necessários à construção de modelos computacionais capazes de gerar informações uteis à tomada de decisão e predição de desligamentos dos motores para, assim, reduzir os impactos causados devido a falhas (LEVITT, 2011; MOBLEY, 2002).

1.7 Questão de Pesquisa

A partir do objetivo formulado, foi possível formular a questão de pesquisa a seguir.

QP1: É possível prever o desligamento de um motor com tempo suficiente para que sejam tomadas ações preventivas e corretivas?

1.7.1 Hipótese

QP1 - H0: Não é possível prever o desligamento de um motor com tempo suficiente para que sejam tomadas ações preventivas.

1.8 Organização do Documento

Este documento está dividido em seis capítulos. O primeiro capítulo, em questão, contém uma introdução sobre o contexto e problemática da pesquisa, alguns problemas encontrados na UTE, bem como os objetivos geral e específicos da pesquisa.

No segundo capítulo, são apresentadas pesquisas relacionadas ao tema deste estudo, as quais tratam da análise de sequências de alarmes emitidos pelos

motores, em especial em usinas termoelétricas. Além disto, os trabalhos apresentados no capítulo tratam também do problema de encontrar a causa raiz para geração das sequências de alarmes.

No terceiro capítulo, é descrita a fundamentação teórica sobre métodos de predição e reconhecimento de padrões aplicados à pesquisa.

O quarto capítulo trata da abordagem proposta, com a descrição da metodologia utilizada para treinamento, predição e testes, a ser aplicada no prognóstico de falhas que levam ao desligamento dos motores.

No quinto capítulo, são apresentadas a metodologia e detalhes sobre os algoritmos e técnicas utilizados.

No sexto capítulo, os resultados obtidos são apresentados e discutidos.

Ao final, no sétimo capítulo, são enunciadas as considerações finais, as contribuições e as sugestões para pesquisas futuras.

No Apêndice A, estão as análises estatísticas descritivas dos dados e a análise de correlação entre os alarmes.

Por último, no Apêndice B, são detalhadas informações relevantes da modelagem da pesquisa.

Capítulo 2

Pesquisas Correlatas

Neste capítulo, é apresentado o estudo do estado da arte no que diz respeito à predição de alarmes. Serão apresentadas técnicas e abordagens semelhantes às utilizadas no estudo ora descrito.

Uma parte das pesquisas apresentadas compreende estudos de técnicas e algoritmos de reconhecimento de padrões para identificar a causa raiz de uma sequência de eventos ordenados em função do instante de tempo em que ocorreram.

Outros estudos apresentados estão relacionados a uma área denominada *Sequential Pattern Mining (SPM)*, que consiste no estudo de técnicas e algoritmos para identificar padrões em sequências de eventos.

Serão apresentadas pesquisas que abordaram problemas clássicos dos sistemas de alarmes, a exemplo das inundações e do número elevado de alarmes falsos.

Por fim, serão apresentadas as principais diferenças entre as pesquisas selecionadas mais relevantes e o estudo realizado (Quadro 1).

2.1 Padrões em Sequências Temporais

Uma sequência temporal é um conjunto de eventos ordenado pelo instante de tempo. Cada evento contém informações sobre a falha correspondente, como a identificação do equipamento onde a falha ocorreu, o instante de tempo da ocorrência, o valor lido pelo sensor correspondente à variável de processo (temperatura, tensão, rotação do motor, pressão, etc) no momento em que o distúrbio ocorreu e etc.

Todos os eventos são mostrados em um sistema de visualização, denominado *Process Control System (PCS)*, que é o sistema de gerenciamento e visualização usado para auxiliar operadores no controle da produção em plantas

industriais (FOLMER et al., 2014; VOGEL-HEUSER et al., 2012). Cada registro é guardado em um banco de dados pelo PCS para manter o histórico de ocorrências. Além dos eventos, o PCS é o sistema responsável por apresentar ao operador as informações mais importantes sobre a planta (FAN YANG et al., 2014) como, por exemplo, a tensão, temperatura e pressão dos motores, em tempo real.

Além da observação dos dados do processo, o operador deve reconhecer situações anormais da planta. As situações anormais podem ser divididas em dois tipos, como alarme e alerta (NOYES, 2002). Uma situação anormal da planta é indicada por visões que mostram o funcionamento do processo, por gráficos que mostram os valores dos sensores nos equipamentos e também por alarmes que identificam a falha (FOLMER et al., 2014). Neste estudo, os eventos são considerados ocorrências comuns que indicam o funcionamento normal dos equipamentos enquanto que os alarmes são os eventos considerados críticos.

Muitas técnicas de aprendizagem de máquina têm sido utilizadas para descobrir informações importantes em grandes aglomerados de dados. Em muitos estudos na literatura, autores como Folmer et al. (2014), Zhao et al. (2003) e Boghey et al. (2013) apresentaram abordagens utilizando técnicas e algoritmos de SPM para detectar padrões importantes em sequências de eventos que ocorreram ao longo do tempo. Um padrão pode ser um evento isolado ou uma subsequência de eventos que aconteceram com frequência antes de outras sequências. Uma parte de uma sequência de eventos, delimitada por um instante de tempo inicial e final, restringe um período de tempo de interesse para a análise, formando uma janela de tempo. Para este estudo, a detecção e análise das associações entre os eventos dentro das janelas de tempo, conduzem a informações estruturadas sobre as ocorrências de alarmes.

Em Yang et al. (2014), os autores utilizaram técnicas de aprendizagem de máquina (Redes bayesianas) e algoritmos de otimização, baseados em funções matemáticas, em uma base de dados de variáveis de processo e registros de alarmes para verificar se a correlação entre as variáveis de processo correspondem à correlação dos alarmes. O objetivo é auxiliar na configuração correta dos limites de alarme. O problema abordado no estudo de Yang et al

(2014) está voltado ao estudo do número elevado de falsos positivos gerados pelo sistema de alarmes.

Leemans e Vander (2015) identificaram características importantes em episódios (coleção de eventos parcialmente ordenados) definidos nos registros de eventos e de informações sobre processo, com o intuito de prever e descobrir comportamentos correlacionados entre os processos. Para isto, desenvolveram uma ferramenta, a qual chamaram de plug-in (ProM), que explora os dados para descobrir alarmes frequentes e padrões dentro dos episódios.

Folmer et al. (2014) propuseram algoritmos para encontrar alarmes com dependência causal através de análises em sequências temporais. Os autores utilizaram estatística para identificar se dois eventos possuem relação de causa e efeito considerando o tempo. Os autores encontraram os parâmetros de uma função densidade de probabilidade, a partir da média de tempo entre as ocorrências de alarmes causa e efeito, que ocorreram no histórico. Posteriormente, utilizaram a função para identificar alarmes relacionados. Os resultados encontrados foram aplicados para auxiliar na construção de um novo sistema de alarmes.

SCHLUTER et al. (2011) descreveram várias abordagens para encontrar regras de associação em sequências temporais. Uma relação temporal é descoberta analisando a ordem e o intervalo em que os eventos aconteceram. Eventos associados podem ocorrer periodicamente.

THUAN et al. (2012) e LI et al. (2001) se concentraram em estudar características que são sempre verdadeiras em determinados intervalos de tempo.

SCHLUTER et al. (2011) implementaram algoritmos de mineração para encontrar regras de associação temporal entre os eventos da mesma sequência ou entre eventos de sequências diferentes.

2.2 Supressão de Alarmes Indesejados

Os alarmes indesejados acontecem em diversas situações na UTE. Em diversos momentos, fora dos períodos de operação, podem ser realizados testes nos

motores. Ao ocorrer alarmes desta natureza, os operadores simplesmente os ignoram. A técnica de identificação de padrões proposta neste estudo, por outro lado, leva em consideração se os alarmes ocorreram com a mesma frequência em períodos em que houve desligamentos abruptos e em períodos que não houve. Caso seja frequente, em ambas situações, este alarme é considerado indesejado pois acontece de forma desorganizada sem um propósito específico. Desta forma, são considerados apenas os alarmes de interesse dos operadores, considerando como característica fundamental para um alarme relevante que este não ocorra com maior frequência em períodos em que houve desligamento abrupto.

DORGO et al. (2018) propuseram uma metodologia que, inicialmente, encontra padrões temporais frequentes no histórico de alarmes através de transformações de sequências multi-temporais em classificadores bayesianos. As regras de associação obtidas podem ser usadas para definir as regras de remoção de alarmes desnecessários. O conjunto de dados utilizado foi coletado de uma plataforma de experimentação de tratamento de água em escala de laboratório para ilustrar que sequências multi-temporais são aplicáveis para a descrição de padrões de operação.

2.3 Otimização de Algoritmos

A otimização dos algoritmos, no contexto desta pesquisa, representa uma proposta para estudos futuros. A identificação dos períodos de operação pode ser de uma forma mais precisa, considerando exatamente o momento que o motor iniciou a operação. Otimizações, na utilização das variáveis e na forma como os dados que são salvos no banco, podem ser realizadas para que o desempenho do treinamento e execução do modelo de predição seja aumentado.

Em HU et al. (2016), os autores propuseram uma melhoria para o algoritmo Smith-Waterman (SW), que possui alta complexidade computacional, impedindo que as aplicações que o utilizam alcancem um tempo de resposta tolerável. Como proposta, apresentou um novo algoritmo baseado na ferramenta de busca de alinhamento local (BLAST). Para tanto, foram utilizados os dados de registros de alarme de uma usina de transformação de substâncias através de processos químicos.

2.4 Predição de Alarmes

A predição de alarmes, em geral, representa a generalização do objetivo deste estudo.

Em LANGONE et al. (2015), os autores desenvolveram um modelo não-linear autoregressivo (NAR), que consiste em um procedimento sistemático de seleção de modelos, que permite ajustar cuidadosamente os parâmetros do modelo. Posteriormente, o NAR foi usado online para prever a tendência futura da temperatura. Finalmente, um classificador que usa como entrada os resultados do modelo NAR permite prever os alarmes futuros.

Em ZHU et al. (2016), os autores criaram um algoritmo dinâmico de predição de alarmes que utiliza registros de alarmes de um sistema de controle distribuído para calcular a probabilidade de ocorrência dos que são mais críticos, com o intuito de prevê-los com antecedência e evitar a consumação de falhas.

2.5 Identificação de Causa Raiz em Sequências de Alarmes

Muitos estudos documentados na literatura da área consideram o problema de identificação da causa raiz em sequências de alarmes. É importante descobrir essa causa raiz para identificar o motivo que causou a falha. Este estudo não tem como objetivo principal identificar a causa raiz do problema, mas quantificar as chances de cada alarme, ativado ao longo do tempo, levar a um desligamento. O foco é evitar a consumação da falha ao invés de descobrir o que a causou.

ISERMANN et al. (1997) utilizaram modelos de redes bayesianas treinadas a partir de Informações de processo e padrões gráficos para capturar a relação de causa e efeito entre padrões gráficos e informações de processo para encontrar causas de falhas.

DAHLSTRAND (2002) construiu um sistema utilizando técnicas de modelagens de sistemas, com o objetivo de melhorar a identificação das causas raiz de uma sequência de alarmes para o setor de operação. Ele utilizou documentação de processos industriais e manuais de equipamentos para modelar

os fluxos e os papéis de cada equipamento. Os modelos utilizados foram do tipo fluxo multinível (MFM).

ALAEDDINI et al. (2011) utilizaram redes bayesianas treinadas a partir de informações geradas por um algoritmo que captura relação de causa e efeito entre padrões gráficos, informações de processo e possíveis causas raiz de sequências de alarmes com o objetivo de identificar, em tempo real, as causas atribuíveis das falhas.

2.6 Inundação de Alarmes

Um problema muito comum encontrado na literatura é o problema da inundação de alarmes que ocorre principalmente por causa de variáveis de processo correlacionadas, que acabam afetando umas às outras quando um distúrbio acontece. Os trabalhos na literatura consideraram várias técnicas, ao longo dos anos, as quais propõem algoritmos para correção das inundações de alarmes. As inundações ocorrem no sistema de controle quando uma situação de falha acontece. A quantidade de alarmes emitidos em algumas situações pode obscurecer os alarmes importantes com informações sobre o problema.

Em WANG et al. (2015), os autores apresentam uma proposta para o problema da inundação dos alarmes. Eles construíram algoritmos que encontram os alarmes causa e consequência e identificam os seus caminhos de evolução. Para isto, os autores utilizaram algoritmos de análise de similaridade e análise de causalidade no histórico de alarmes e também levaram em consideração os atrasos entre os eventos de processo. Além disto, os autores utilizaram o método de causalidade de Granger para esclarecer os impactos gerados pelas duas vertentes: processo e alarmes.

HU et al. (2018), propuseram um método para encontrar e remover grupos de alarmes irrelevantes após a falha, fazendo com que o operador receba apenas os que possuem relação com a falha. Os autores apresentam uma técnica própria adaptada, de mineração de dados históricos de alarmes para solucionar o problema.

FOLMER et al. (2012), apresentaram uma proposta com o intuito de melhorar e redesenhar um sistema de alarmes, para reduzir o número de alarmes apresentados ao operador e propuseram um algoritmo para análise automática de alarmes (AADA), que identifica os estágios assumidos pelo alarme ao longo do tempo usando autômatos finitos. O algoritmo agrupa as sequências em conjuntos, conforme uma métrica de similaridade. Cada elemento da sequência corresponde a um estado no autômato. O autômato reconhece a subsequência encontrada se esta corresponder às mudanças de estado do alarme ao longo do tempo. Desta forma, o algoritmo encapsula a estrutura geral dos dados sequenciais identificando os diferentes estados em que o alarme esteve. As transições refletem as mudanças no equipamento e informa se a central de controle reconheceu ou não o alarme.

2.7 Melhoraria dos Processos Industriais ou do Sistema de Controle de Alarmes e Processos

Outras pesquisas na literatura da área abordaram o problema de melhorar os processos industriais ou o sistema de controle de alarmes, como descrito em PARIYANI et al. (2011). O autor desenvolveu na primeira parte do estudo uma metodologia dinâmica de análise de risco baseado em três grandes etapas: Rastreamento de eventos, Formulações de árvores de eventos e Análise bayesiana.

Na segunda parte do trabalho, em PARIYANI et al. (2012), os autores se concentraram na terceira etapa da análise de riscos baseada na análise bayesiana. Os autores desenvolveram um novo método de análise bayesiana que utiliza uma base de dados de alarmes. O objetivo dos dois trabalhos foi melhorar a segurança do processo e a qualidade do produto no contexto da indústria onde a pesquisa foi realizada. Para chegar aos objetivos, o autor utilizou dados do sistema de controle distribuído e do sistema de desligamento de emergência da indústria.

2.8 Mineração de Dados

Todas as operações de busca realizadas neste estudo consideram a mineração de dados. São realizadas tarefas de agrupamento, pesquisa, identificação, separação, *matching* e filtragem para obter os conjuntos de dados para as fases de treinamento e teste do modelo.

Em LI et al. (2017), os autores melhoraram a validade e a precisão do agrupamento de alarmes, mediante um algoritmo melhorado do algoritmo tradicional de agrupamento hierárquico aglomerativo usado para previsão e prevenção de riscos. Os autores propuseram uma abordagem usando os conceitos de mineração de dados e correlação vetorial juntamente com probabilidade condicional.

2.9 Resumo das diferenças entre este estudo e as pesquisas apresentadas

No Quadro 1, é apresentado um resumo das principais diferenças entre o estudo realizado e as pesquisas apresentadas. O principal diferencial deste estudo, em relação aos demais, reside no uso de um modelo de predição para prognóstico de desligamentos em motores de uma UTE. Um modelo híbrido gera informações de prognóstico para predição de desligamentos abruptos.

Quadro 1. Resumo comparativo entre a pesquisa realizada e as principais pesquisas correlatas (Continua).

Pesquisa	Título	Objetivo	Técnica usada	Base de dados
Identificação de Padrões em Sequências Temporais				
Folmer et al. (2010)	*Criteria-based alarm flood pattern recognition using historical data from automated production systems (aPS)*	Apresentaram uma abordagem com o objetivo de reduzir a carga de informações apresentadas a um operador, substituindo sequências de alarmes causalmente dependentes por uma informação única, eliminando múltiplas informações desnecessárias.	Análise estatística e Máquinas de estados	Registros de notificação de oito aPS industriais existentes, bem como a avaliação de especialistas industriais são levados em consideração.
Yang et al. (2014)	*Correlation Analysis of Alarm Data and Alarm Limit Design for Industrial Processes*	Verificaram se a correlação das variáveis de processo é refletida na correlação dos alarmes correspondentes para garantir que os limites de alarme sejam configurados corretamente.	Técnicas de aprendizagem de máquina (Redes bayesianas) e algoritmos de otimização baseados em funções matemáticas.	Valores coletados das variáveis de processo e registros de alarmes

Fonte: Autoria Própria.

Quadro 1. Resumo comparativo entre a pesquisa realizada e as principais pesquisas correlatas (Continuação).

Leemans e Vander (2015)	*Discovery of Frequent Episodes in Event Logs*	Identificaram regras de episódio (coleção de eventos parcialmente ordenados) para prever e descobrir comportamentos correlacionados em processos.	Desenvolveram um plug-in (ProM) que explora algoritmos eficientes para a descoberta de episódios frequentes e regras de episódios.	Registros de eventos e de processo
Schluter e Conrad (2011)	*About the analysis of time series with temporal association rule mining*	Mineraram regras de associação temporal, para encontrar relações entre os eventos em uma ou entre pares de sequências de eventos com o objetivo de encontrar alarmes relacionados.	Técnicas de mineração de regras de associação temporal.	Dados contendo informações sobre o aumento dos níveis dos rios.
Li et al. (2001)	*Discovering calendar-based temporal association rules*	O foco foi encontrar algoritmos eficientes para resolver o problema de minerar regras de associação e encontrar intervalos de tempo relacionados.	Encontrar regras que são sempre verdadeiras nos intervalos de tempo estendendo o bem conhecido algoritmo Apriori com técnicas de remoção usando componentes de tempo.	Supõe-se que os bancos de dados mantenham informações sobre transações do usuário, em que cada transação é uma coleção de itens qualquer (pedidos de restaurantes por exemplo).

Fonte: Autoria Própria.

Quadro 1. Resumo comparativo entre a pesquisa realizada e as principais pesquisas correlatas (Continuação).

Levantamento de técnicas, metodologias, algoritmos e modelos para identificação de padrões em sequências.				
Zhao e Bhowmick (2003)	*Sequential Pattern Mining: A Survey*	Listaram algoritmos e técnicas relacionados a área de SPM para documentar os principais conceitos sobre a área.	Identificou as subáreas de SPM e descreveu os conceitos e técnicas mais utilizadas.	Documentações na literatura, livros, artigos etc.
Boghey e Singh (2013)	*Sequential Pattern Mining: A Survey on Approaches*	Analisaram o progresso atual dos métodos de identificação de padrões sequenciais em bases de dados.	Ele descreve uma variedade de modelos para tarefa de mineração de padrões sequenciais, classificando em categorias.	Artigos, livros e documentos da literatura.
Inundações de Alarmes				
Wang et al. (2015)	*A data similarity based analysis to consequential alarms of industrial processes*	Encontraram os alarmes causa/ consequência que podem ser efetivamente identificados juntamente com seus caminhos de evolução com o objetivo de amenizar as inundações de alarmes.	Combina a análise de similaridade e análise de causalidade de alarmes levando em consideração os atrasos entre os eventos de processo e o método de causalidade de Granger para esclarecer impactos mútuos.	Dados de alarme e dados de processo (informações relacionadas ao processo como auditoria do produto, mecanismos de verificação, registros de operações, etc).

Fonte: Autoria Própria.

Quadro 1. Resumo comparativo entre a pesquisa realizada e as principais pesquisas correlatas (Continuação).

Folmer e Vogel-Heuser (2012)	*Computing Dependent Industrial Alarms for Alarm Flood Reduction*	Identificaram alarmes frequentes e os que possuem maior ligação com a falha nas sequências de alarmes, para melhorar e redesenhar um sistema de alarmes com o intuito de reduzir o número de alarmes apresentados ao operador.	Apresentaram um algoritmo para análise automática de alarmes (AADA), que identifica os estágios assumidos pelo alarme ao longo do tempo usando autômatos finitos.	Registros de alarmes
Hu et al.(2018)	*Detection of Frequent Alarm Patterns in Industrial Alarm Floods Using Itemset Mining Methods*	Propuseram um método para encontrar e remover grupos de alarmes irrelevantes após a falha.	Apresenta uma técnica adaptada de mineração de conjunto de itens	Dados históricos de alarmes
Otimização de Algoritmos				
Hu et al. (2016)	*A local alignment approach to similarity analysis of industrial alarm flood sequences*	Propuseram um algoritmo alternativo com melhorias ao algoritmo Smith-Waterman como proposta para solução das inundações de alarmes.	Propôs um novo algoritmo de baseado na ferramenta de busca de alinhamento local (BLAST)	Registros de alarme de uma usina de conversão de óleo

Fonte: Autoria Própria.

Quadro 1. Resumo comparativo entre a pesquisa realizada e as principais pesquisas correlatas (Continuação).

Supressão de Alarmes Indesejados				
Dorgo e Abonyi (2018)	*Sequence Mining Based Alarm Suppression*	Encontraram regras de associação para definir regras de remoção de alarmes desnecessários e diminuir o número de alarmes emitidos indevidamente.	Propôs uma metodologia que identifica padrões temporais frequentes dos sinais de alarme, através de transformações de sequências multi-temporais em classificadores de Bayes com o objetivo de remover alarmes que não tem relação com o inicial emitido.	Conjunto de dados de uma plataforma de experimentação de tratamento de água em escala de laboratório.
Identificação de Causa Raiz em sequências de alarmes				
Dahlstrand (2002)	*Consequence analysis theory for alarm analysis*	Utilizaram modelagens de sistemas, com o objetivo de melhorar a identificação das causas raízes de uma sequência de alarmes.	Modelos de fluxo multinível (MFM)	Documentação de processos industriais, Manuais de equipamentos.

Fonte: Autoria Própria.

Quadro 1. Resumo comparativo entre a pesquisa realizada e as principais pesquisas correlatas (Continuação).

Alaeddini et al.(2011)	*Using Bayesian networks for root cause analysis in statistical process control*	Implementaram um método para identificação em tempo real de causas atribuíveis únicas e múltiplas de falhas	Treinaram redes bayesianas usando dados gerados por um algoritmo que captura relação de causa e efeito entre padrões gráficos, informações de processo e possíveis causas raiz de sequências de alarmes.	Registros de alarmes, gráficos de controle e de processo.
Isermann (1997)	*SUPERVISION, FAULT-DETECTION AND FAULT-DIAGNOSIS METHODS - AN INTRODUCTION*	Construíram uma rede bayesiana para capturar a relação de causa e efeito entre padrões gráficos, informações de processo para encontrar possíveis causas raiz.	Técnicas de aprendizagem de máquina	Informações de processo e padrões gráficos
Melhorar os processos industriais ou o sistema de controle de alarmes e processos				
Pariyani et al. (2011)	*Dynamic Risk Analysis Using Alarm Databases to Improve Process Safety and Product Quality: Part I — Data Compaction*	Melhoraram a segurança do processo e a qualidade do produto no contexto da indústria onde a pesquisa foi realizada.	Desenvolveu uma metodologia dinâmica de análise de risco baseado em três etapas: ● Rastreamento de eventos; ● Formulações de árvores de eventos; Análise bayesiana.	Bancos de dados do sistema de controle distribuído (DCS) e de desligamento de emergência (ESD)

Fonte: Autoria Própria.

Quadro 1. Resumo comparativo entre a pesquisa realizada e as principais pesquisas correlatas (Continuação).

Pariyani et al. (2012)	*Dynamic Risk Analysis Using Alarm Databases to Improve Process Safety and Product Quality: Part II — Data Compaction*	Melhoraram a segurança do processo e a qualidade do produto no contexto da indústria onde a pesquisa foi realizada.	Desenvolveu um novo método de análise bayesiano que utiliza alarmes do sistema de controle distribuído (DCS) e de desligamento de emergência (ESD)	Bancos de dados do sistema de controle distribuído (DCS) e de desligamento de emergência (ESD)
Folmer et al. (2014)	*Detection of temporal dependencies in alarm time series of industrial plants*	Desenvolveram uma abordagem para auxiliar especialistas para redesenhar o sistema de alarmes e / ou projetar um sistema de predição.	Técnicas de análise estatística em séries temporais	Registros de alarmes de processos contínuos (indústria de controle de processos), discretos (indústria de manufatura) e híbridos.
Predição de alarmes				
Zhu et al. (2016)	*Dynamic alarm prediction for critical alarms using a probabilistic model*	Criaram um algoritmo baseado em um modelo probabilístico para prever alarmes críticos com antecedência para evitar a consumação de falhas.	Apresentou um algoritmo dinâmico de previsão de alarmes, um modelo probabilístico que utiliza dados de alarme do sistema de controle distribuído, para calcular a probabilidade de ocorrência de alarmes críticos.	Registros de alarmes

Fonte: Autoria Própria.

Quadro 1. Resumo comparativo entre a pesquisa realizada e as principais pesquisas correlatas (Continuação).

Langone et al. (2015)	*Alarm prediction in industrial machines using autoregressive LS-SVM models*	Propuseram um algoritmo de predição de alarmes que baseado na predição, ajusta parâmetros de configuração da máquina de produção dinamicamente.	Desenvolveram um modelo não-linear autoregressivo, Máquinas de Vetores de Suporte de Mínimos Quadrados (LS-SVMs), em que um procedimento sistemático permite ajustar cuidadosamente os parâmetros do modelo.	Dados coletados de uma máquina de produção de aço usada para trefilação de arame.
Mineração de dados				
Thuan et al. (2012)	*An Approach Mining Cyclic Association Rules in E-Commerce*	Encontraram regras de associação que ocorrem com frequência no banco de dados de organizações comerciais para identificar produtos que são frequentemente comprados pelos clientes.	Mineração de dados como uma técnica para dar suporte às vendas no comércio eletrônico usando a mineração de padrões cíclicos para a varredura de conjuntos de itens.	Bancos de dados de vendas de produtos na internet (E-Commerce).
Li et al.(2017)	*Data mining algorithm for correlation analysis of industrial alarms*	Melhorou a validade e a precisão do agrupamento de alarmes através de um algoritmo combinado com o algoritmo tradicional de agrupamento hierárquico aglomerativo para previsão e prevenção de riscos	Propõe uma abordagem usando os conceitos de mineração de dado e correlação vetorial juntamente com probabilidade condicional.	Registros de alarmes

Fonte: Autoria Própria.

Quadro 1. Resumo comparativo entre a pesquisa realizada e as principais pesquisas correlatas (Continuação).

Soares (2016)	Detecção e diagnóstico de falhas em plantas industriais com base em padrões de alarmes.		Análise estatística	
Pesquisa Proposta				
Vasconcelos, B. R. A. et al.	Geração de Prognóstico e Predição de Desligamentos a partir do Histórico de Alarmes Emitidos por Motores de uma Usina Termoelétrica	Construíram um modelo de aprendizagem de máquina híbrido, com o intuito de gerar informações de prognóstico sobre os desligamentos de motores de uma UTE, com antecedência suficiente para auxiliar a central de controle de uma termoelétrica no processo de correção de falhas.	Construção de um modelo de aprendizagem de máquina usando árvores de decisão e geração de informações sobre as sequências de alarmes usando máquinas de estados.	Registros de alarmes gerados pelos motores De uma UTE.

Fonte: Autoria Própria.

2.10 Discussão

Em nenhuma das pesquisas revisadas os autores estudaram os alarmes no contexto de usinas termoelétricas e não utilizaram técnicas para identificação de padrões, baseadas em máquinas de estados e aprendizagem, com o intuito de construir um modelo de predição. A pesquisa ora apresentada, objetiva utilizar as técnicas de SPM no contexto de usinas termoelétricas, com o objetivo de verificar se há associação entre os alarmes e, assim, construir um modelo capaz de classificar, previamente, se haverá ou não desligamento abrupto.

Por fim, pode-se concluir, a partir da revisão bibliográfica, que a pesquisa realizada apresenta um diferencial, pois considera o uso de um modelo de predição, construído a partir de técnicas de identificação de padrões, levando em consideração a frequência com que os alarmes ocorreram e a geração de informações de prognóstico para tomada de decisão. Em nenhum dos outros estudos revisados, esta abordagem foi utilizada.

Neste estudo, serão identificadas sequências de alarmes frequentes e que não foram corrigidos antes do desligamento. As máquinas de estado serão usadas para acompanhar os estados dos alarmes e os algoritmos de aprendizagem de máquina serão utilizados para construção de modelos de predição. O objetivo de utilizar estas técnicas é identificar alarmes críticos, a partir das mudanças de estado ao longo do tempo e utilizar as informações resultantes para predizer se ocorrerá um desligamento.

A forma como o intervalo de tempo entre a sequência de alarmes e o desligamento foram abordados neste estudo, é semelhante àquela apresentada em FOLMER et al. (2014). Para serem consideradas características de relação causal, as subsequências devem acontecer uma após a outra, frequentemente, em uma janela de tempo, definida entre o momento da inicialização do motor e o alarme de desligamento, seja este abrupto ou não.

No capítulo seguinte, será descrita a fundamentação teórica da proposta de pesquisa.

Capítulo 3

Fundamentação Teórica

3.1 O Problema Técnico Científico

O problema técnico científico da pesquisa reside no uso de um modelo de predição construído usando técnicas de aprendizagem de máquina, para identificação de padrões, com o objetivo de prever desligamentos, devido a falhas, de motores utilizados na geração de energia em uma usina termoelétrica.

O sistema de controle instalado na UTE possui, em linhas gerais, todas as informações que os operadores precisam saber em termos de diagnóstico do processo de geração de energia. Este sistema possui um conjunto de sensores ajustados de forma específica, por meio de calibração *in loco* ou na fabricação. Os sensores monitoram o comportamento de variáveis do processo, como temperatura, tensão elétrica, número de rotações do motor, pressão e etc, que indicam o estado interno de uma máquina (motor). Sempre que uma variável estiver acima ou abaixo de limites pré-definidos pelos operadores ou em calibração, uma informação de alarme é gerada.

Os dados de alarmes são importantes para os operadores, porém, o sistema de controle não gera nenhuma informação acerca de prognósticos, como por exemplo, indicação, se vai ocorrer, com determinada probabilidade, um desligamento do motor, a probabilidade de determinado alarme ou sequência de alarmes levar a um desligamento, o tempo médio entre um alarme e o desligamento, dentre outras. Por isto, a pesquisa ora apresentada, consiste no desenvolvimento de um modelo supervisionado de predição, a partir das sequências de alarmes, que informa se haverá o desligamento de um motor, com determinada probabilidade.

3.2 Funcionamento do Sistema de Controle de Alarmes

A ativação de um alarme no sistema de controle é feita automaticamente, sem a necessidade de um agente humano. Para cada falha que acontece no tempo, sem

exceção, um alarme é gerado. Quando os valores coletados pelos sensores ultrapassam os limites aceitáveis de operação, por um determinado período de tempo para evitar a geração de falsos alarmes, caracteriza-se uma falha e um alarme descrevendo o que aconteceu é emitido.

A partir do histórico de alarmes, vários tipos de análises podem ser realizados (DAHLSTRAND, 2002). Como exemplo, é mostrado na Figura 2, na Seção 1.2 um gráfico da variável temperatura (T) *versus* o tempo. Existem dois tipos de alarmes que merecem destaque (SOARES, 2016), os alarmes que causam de fato o desligamento (ultrapassam os limites de segurança) e os alarmes que causam apenas um alerta (ultrapassam os limites de falha). Se um determinado sensor registra um valor e esse é maior do que os limites de falha (linha pontilhada da Figura 2), então, um alarme é emitido na central de controle, com informações sobre o problema. Se o valor ultrapassar o limite de segurança superior (linha vermelha da Figura 2), então o equipamento é retirado de operação. Nem todos os alarmes podem levar ao desligamento diretamente. Alguns representam problemas que podem se agravar ao longo do tempo, gerando novas falhas, que de fato podem levar ao desligamento. Estes alarmes serão chamados de *alarmes intermediários*, enquanto que os alarmes que levam ao desligamento diretamente serão chamados *alarmes diretos.*

Basicamente, uma falha pode ser vista como um desvio do funcionamento normal do equipamento. Por exemplo, a pressão do óleo que circula no motor pode baixar de repente e se tornar inferior aos valores normais de operação. KONDAVEETI (2013) destacam que um registro de alarme genérico, para apresentar informações suficientes sobre o problema, deve conter alguns campos como o instante de tempo em que foi ativado, um rótulo que o identifique de forma única na planta e a descrição do distúrbio. Um registro de alarme emitido pelo sistema de controle, além destas, gera outras informações, conforme exibidas no Quadro 2.

Todo registro possui o atributo *EventStamp*, que indica o instante de tempo em que o alarme ocorreu. No estudo ora descrito, o histórico de alarmes foi organizado em uma sequência temporal, em que todos os alarmes são ordenados

pelo atributo *EventStamp*, pois a dimensão tempo é muito importante para descobrir a existência de *associação de causalidade*.

Quadro 2. Atributos de um registro de alarme.

EventStamp	Instante de tempo em que o alarme ocorreu. Ex: "01/02/2018 00:23:12.320"
Description	Descrição do problema ao qual o alarme se refere.
TagName	Identificação do alarme na planta. Ex: "NHA051G001CLO"
Area	Área/setor da planta em que o alarme foi ativado. Ex: “Genset_1”
AlarmState	Estado do alarme (registros com AlarmState igual à NA são eventos). Ex: “UNACK_ALM”
Value	Valor da variável de processo à qual o alarme está relacionado no momento em que foi disparado (O “Value” também pode assumir valores reais). Ex: “3.8”
Operator	Tipo de usuário logado no sistema no momento em que o alarme foi ativado. Ex: “Operator”

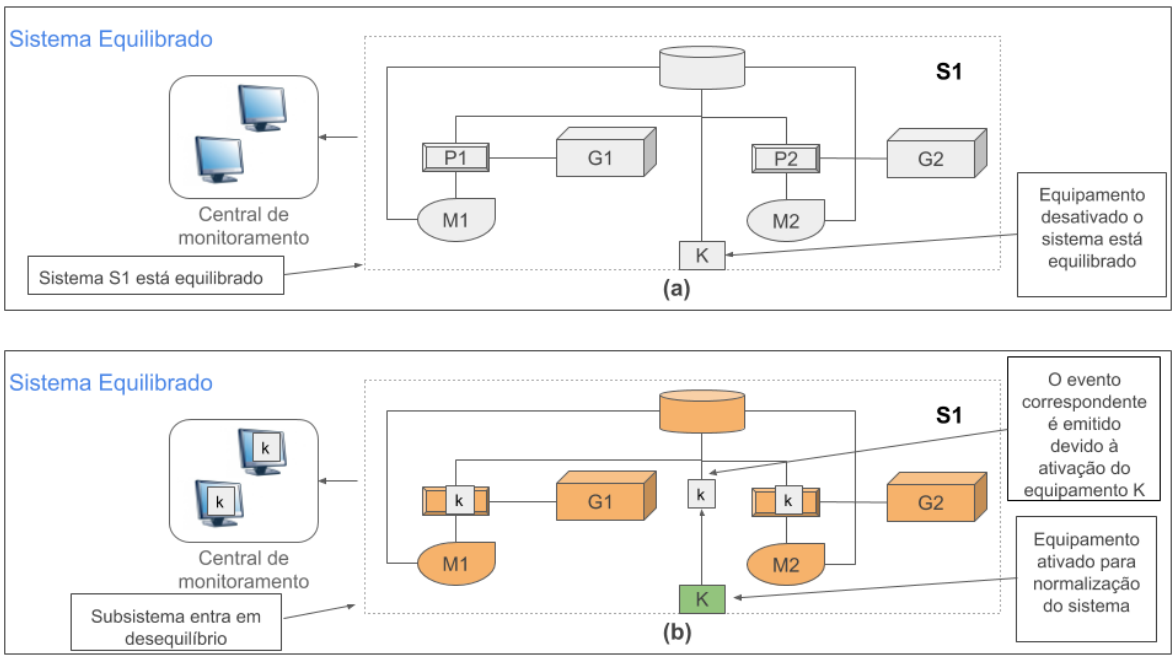
Fonte: Autoria Própria.

Alguns equipamentos também notificam a central de controle sobre o funcionamento normal do sistema. Como exemplo, o sistema exibido na Figura 3 está normalizado, funcionando corretamente, com todos os parâmetros dentro dos limites. Porém, em um determinado momento, o sistema começa a entrar em desequilíbrio (por exemplo, a pressão do óleo no motor começa a aumentar). Assim, é necessário que, automaticamente, sejam tomadas ações para normalizar os componentes. Quando o distúrbio acontece de forma natural, o próprio sistema

é capaz de se recuperar, pois ao ser projetado, seus idealizadores sabiam que de tempos em tempos haveria quedas de pressão.

A bomba Jockey, na Figura 3 (b), é representada como o equipamento K, a qual é ativada para regularização do sistema. A bomba entra em ação sempre que necessário para compensar perdas de pressão e, ao ativá-la, o sistema gera um evento nos painéis P1 e P2 e também na central de controle. Logo após a ação da bomba, o sistema de controle volta ao normal, conforme a situação (a). Todos os eventos envolvidos são salvos pelo sistema em uma base histórica.

Figura 3. Desequilíbrio e normalização do sistema da planta.



Fonte: Autoria própria.

Os alarmes são tipos específicos de eventos, a diferença é que são relacionados a uma falha, enquanto os outros eventos são comuns ao funcionamento normal do sistema. Por isto, os operadores precisam “reconhecê-lo”, a partir do sistema de controle, para garantir que estão cientes do problema e que a central de controle irá intervir para controlá-lo o mais rápido possível, por meio de ações corretivas, para evitar que ocorra algo mais sério ou que o sistema de segurança da UTE seja acionado a partir de desligamentos abruptos.

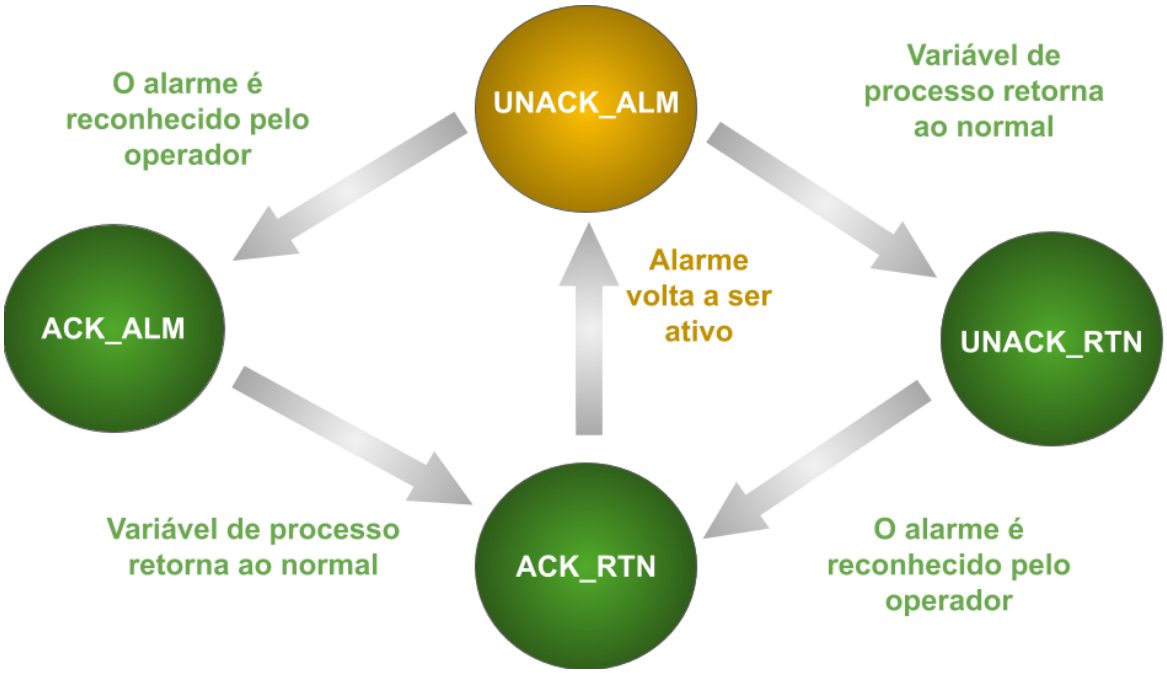
Ao ligar um motor, um evento de partida chamado *Sync On* é lançado indicando o momento em que o operador ligou o motor. Ao ser desligado, é

lançado um evento de desligamento. Caso seja um desligamento natural de fim de produção, o evento lançado será o ENGSTO, caso seja um desligamento forçado, será o S011SDI. O par formado pelo evento de partida do motor e o evento de desligamento, seja esse qual for, define um período de operação do motor que contém uma sequência de alarmes (FOURNIER-VIGER, 2017) inclusive o de desligamento. No histórico de alarmes, sempre que houver um evento de partida, deverá haver um par correspondente de desligamento forçado ou não. Caso não haja um par correspondente, constituem-se ruídos nos dados e o período é descartado. Os dados relevantes ao estudo foram constituídos dos eventos que ocorreram somente em períodos de operação do motor, pois não é de interesse da UTE avaliar as sequências de alarmes em momentos em que o motor está desligado.

Um alarme pode estar em um dentre quatro estados possíveis ao longo do tempo. Em um dos estados, o risco de desligamento é considerável, porque indica que a falha está ativa, conforme apresentado no diagrama da Figura 4. Um estado indica a evolução da falha, se esta foi reconhecida ou não pela central ou se o sistema voltou ao normal sem nenhuma intervenção. Quando o sensor detecta um distúrbio no equipamento, um alarme no estado *unacknowledge alarm* (UNACK_ALM*)* é emitido*,* do inglês “alarme desconhecido” e significa que os operadores ainda não o marcaram no sistema de controle como reconhecido. Após a central de controle reconhecer o alarme, um novo registro é emitido com o estado *acknowledge alarm* (ACK_ALM), ou “alarme reconhecido”, ou seja, as ações corretivas foram devidamente tomadas e, possivelmente, o problema foi corrigido. Se a falha não ocorrer mais, então o alarme retorna ao estado inicial de normalidade *acknowledge return to normal* (ACK_RTN).

Outra possibilidade para o estado do alarme é a alternância do estado ACK_RTN para UNACK_ALM (situação (1) da Figura 5) e voltando depois do estado UNACK_ALM para ACK_RTN (situação (a) da Figura 5), passando pelo estado *unacklowedge return to normal* (UNACK_RTN), que significa que a variável “voltou ao normal” sem nenhuma intervenção de correção dos operadores. Porém, o alarme não retorna ao normal se a central de controle não o reconhecer no sistema de controle.

Figura 4. Os estados possíveis em que um alarme pode ser encontrado.

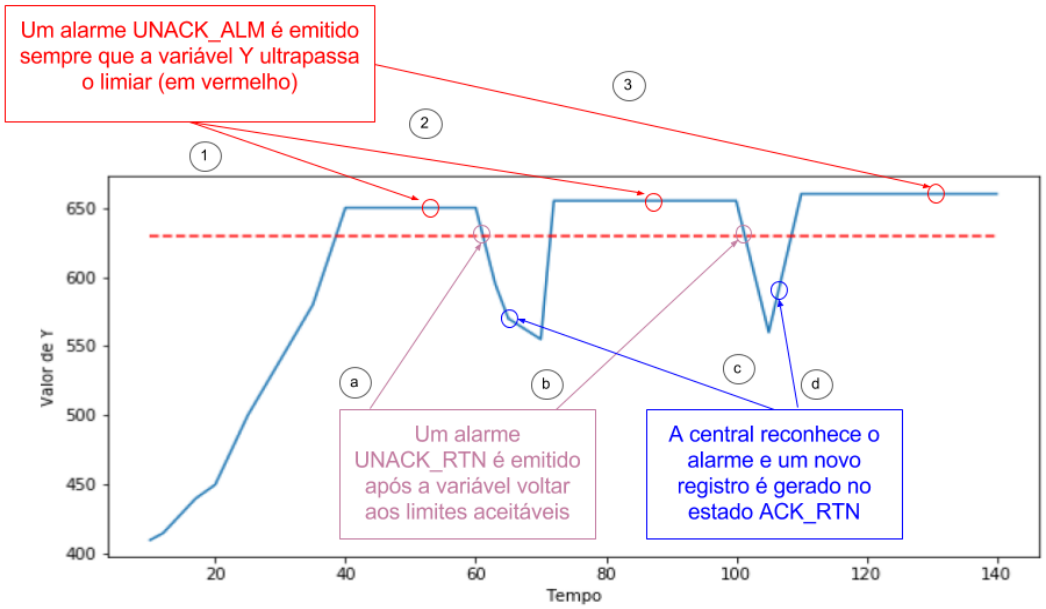


Fonte: http://global.wonderware.com. Acesso em 08/02/2018.

Na Figura 5, a variável Y ultrapassa o limiar superior (situação 1), o que faz com que o alarme mude do estado ACK_RTN para UNACK_ALM. Após certo tempo, a variável volta ao normal (situação a) e o alarme retorna para o estado UNACK_RTN. Somente quando a central de controle o reconhece (situação c), que significa que o problema está sob controle, o alarme volta para o estado ACK_RTN. Para cada mudança de estado, um novo registro é emitido. Isto se repete nas situações 2 e 3 e o alarme oscila seu estado várias vezes ao longo do tempo. Este tipo de situação pode ser causada por diversos motivos, como por exemplo, um defeito no sensor ou algum fenômeno externo repetitivo que influencia a variável Y a oscilar por um curto intervalo de tempo.

A segunda situação que ocorre, é o alarme ser reconhecido pela central de controle antes da variável voltar ao normal. Após o alarme ser ativado e passar do estado ACK_RTN para UNACK_ALM, a central pode tomar ações de correção e reconhecê-lo. Um novo alarme é emitido com estado *acklowedge alarm* (ACK_ALM) para só então a variável voltar aos limites aceitáveis, conforme os estados mostrados na Figura 5. Se não houver retorno aos limites aceitáveis, significa que o problema não foi devidamente corrigido.

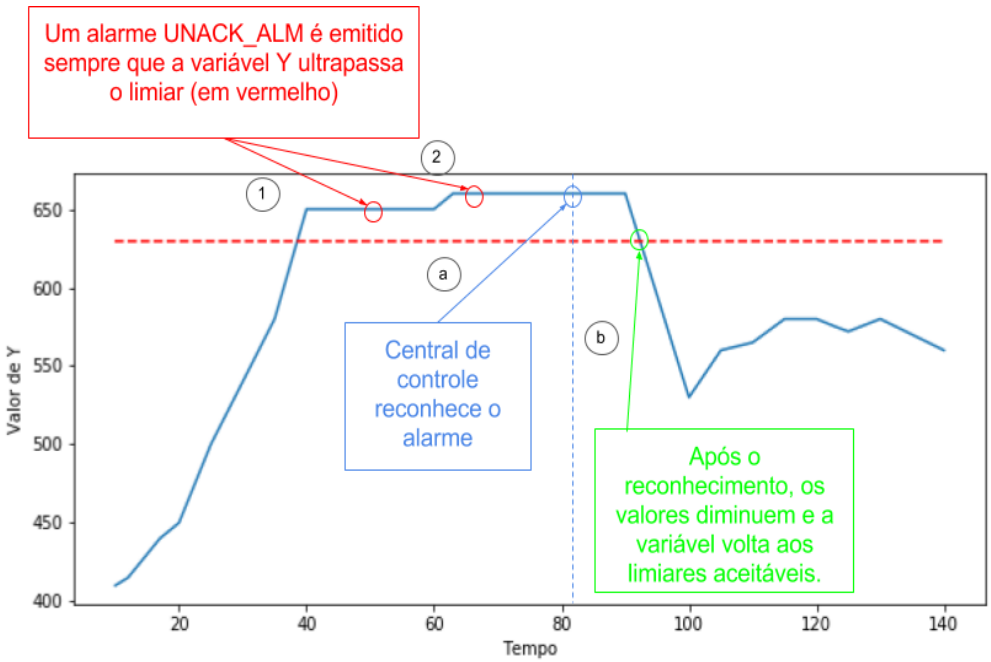
Figura 5. Oscilações dos valores registrados pelo sensor da variável Y.



Fonte: Autoria própria.

Na Figura 6, o alarme é ativado duas vezes (situações 1 e 2) e depois que as ações devidas foram tomadas, o alarme é reconhecido na situação a. Após isto, o sensor da Variável Y passa a registrar valores dentro dos limites e o alarme retorna ao normal (ACK_RTN).

Figura 6. Ativação e reconhecimento do alarme pela central de controle.



Fonte: Autoria própria.

Os eventos gerados são armazenados em um formato específico, em arquivos, formando uma base histórica de operação. Uma sequência de alarmes está relacionada univocamente a uma sequência de sensores, em que cada alarme foi gerado devido a um distúrbio detectado por um sensor.

3.3 Períodos de Operação

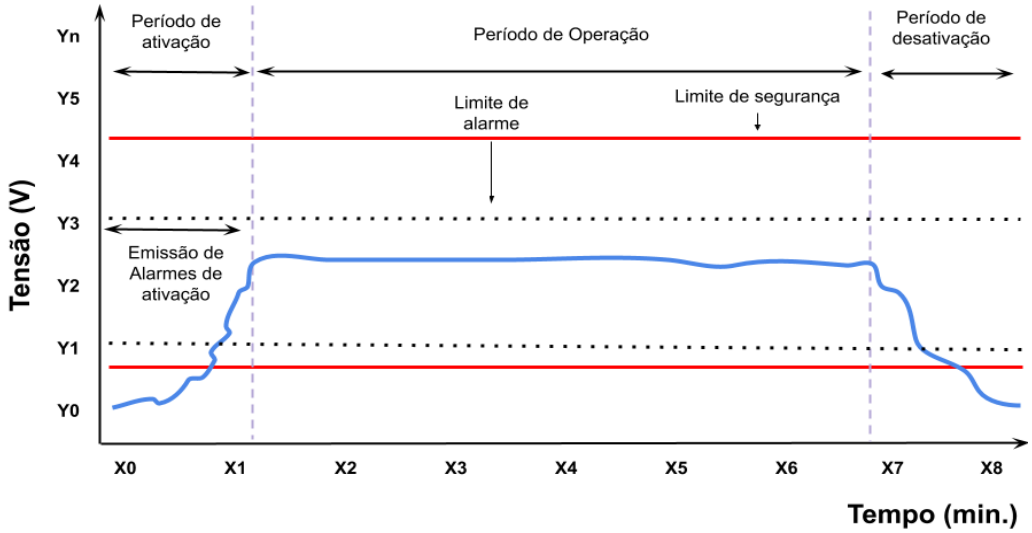
A produção de energia da UTE é definida e tem hora marcada para começar e terminar (modo *stand by*). Os motores ficam ligados por um determinado período de tempo, com o objetivo de produzir certa quantidade de energia elétrica. Quem define isso é um órgão chamado Operador Nacional do Sistema Elétrico (ONS). O instante de tempo em que o motor é iniciado e posteriormente desligado constitui um período de produção. Se o desligamento for natural, realizado pelo operador, o desligamento é considerado normal. Caso haja uma falha que force o sistema de segurança a ser ativado, então, o desligamento é considerado abrupto. Os alarmes são ordenados em sequência no tempo para tornar possível a análise de consequência (ESLING, 2012). No caso deste estudo, a consequência avaliada é o desligamento.

O início da operação do motor não ocorre de imediato, é necessário um espaço de tempo entre o momento em que o operário deu partida e o instante em que o motor começou de fato a operar, conforme mostrado na Figura 7. Durante o período de ativação, muitos alarmes são emitidos porque as variáveis de processo saem dos limites aceitáveis de operação. Por isto, o período de inicialização deve ser descartado durante as análises, o que também pode atrapalhar a busca automática por relações entre as falhas (alarmes) e os desligamentos.

Da mesma forma que a inicialização, o desligamento não ocorre de imediato, nem se for um desligamento abrupto. Nos dois tipos de desligamento, o motor demora certo tempo para desligar, o abastecimento é cortado imediatamente, porém o motor continua operando, enquanto ainda houver combustível internamente. O fato de o desligamento não ser imediato, diferentemente da ativação, não atrapalha a avaliação da relação da falha com o desligamento, pois o alarme selecionado será o primeiro alarme que foi lançado

antes do comando de desativação, evitando assim, a sequência de alarmes que será emitida no período de desativação do motor.

Figura 7. Processo de inicialização do motor da UTE.



Fonte: Autoria Própria.

3.4 Critérios para Identificação de Causalidade

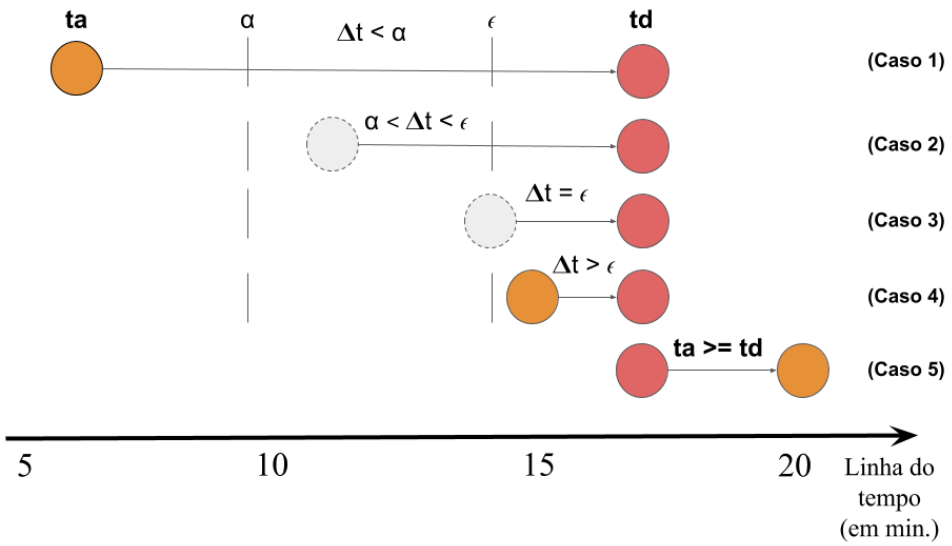
Para que um alarme tenha relação de causalidade com o desligamento abrupto, alguns critérios devem ser satisfeitos.

Primeiro Critério

Para que os alarmes possuam relação de causalidade com o desligamento abrupto, a distância temporal entre esses e o desligamento não deve ser muito pequena nem muito grande. Seja a distância temporal **Δ**t entre um alarme que ocorreu no instante *ta* e um desligamento que ocorreu no instante *td*, seja *α e* ε limitadores do *período de causalidade*, que é o período de tempo em que o alarme pode ser considerado relacionado ao desligamento, então, para que faça sentido um alarme ter causado um desligamento, necessariamente α <= **Δ**t <= ε, como na Figura 8. Além disto, também é necessário que o alarme não tenha ocorrido depois do desligamento, ou seja, *ta* < *td* (caso 3). A relação de causalidade está ligada à posição no tempo e o evento que ocorreu antes influencia o que ocorreu depois e não o contrário.

Se o alarme ocorrer depois (ta >= td), muito perto ($\Delta t > \varepsilon$) ou muito antes ($\Delta t > \alpha$) do desligamento, a possibilidade de causalidade é muito baixa ou inexistente.

Figura 8. Limites temporais para que um alarme seja considerado relacionado ao desligamento.

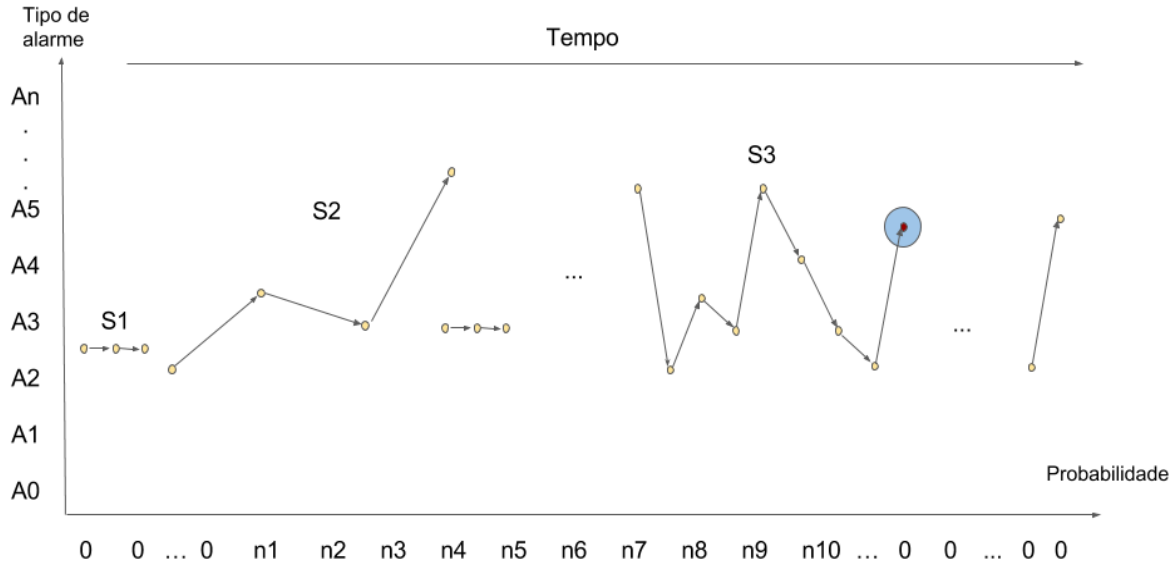


Fonte: Autoria própria.

Conforme a Figura 9, os alarmes da sequência S1 tem menos chance de serem a causa do desligamento do que a sequência S2 que, por sua vez, tem menos chance do que a sequência S3. Quanto mais distante, maior a chance de a sequência de alarmes não ter relação com o desligamento. O mesmo vale para um intervalo de tempo bem próximo, pois a central de controle não poderá fazer mais nada a respeito da situação.

Ao final da operação, o sistema terá produzido uma sequência de alarmes que, na verdade, se trata de uma sequência de falhas. É importante destacar que, em todas as análises realizadas pelas técnicas utilizadas (apresentadas nos Capítulos 4 e 5), um único alarme será sempre considerado uma sequência.

Figura 9. Distribuição dos alarmes em sequência temporal.



Fonte: Autoria própria.

Segundo Critério

Para que um alarme tenha relação de causa com um desligamento, esse deve ocorrer com maior frequência antes de desligamentos abruptos do que antes de desligamentos normais. Um alarme que ocorre frequentemente antes de ambos os desligamentos é um alarme aleatório, não sendo possível identificar com o que este tem relação.

Terceiro Critério

Um alarme que possui relação com o desligamento não foi tratado pela central de controle. Quando há tratamento por parte dos operadores, o alarme é considerado potencialmente sem relação com o desligamento, dado que a central de controle ao reconhecer o alarme indica que a situação está sob controle. Portanto, todos os alarmes considerados causais do desligamento, não foram reconhecidos pela central.

Quarto Critério

Um alarme emitido por um determinado motor não terá relação de causalidade com desligamentos ou quaisquer outros eventos ocorridos em outros motores.

3.5 Definição da Aprendizagem de Máquina e Máquina de estados

Na abordagem proposta, a aprendizagem é levada a efeito a partir da identificação dos padrões. Quanto mais representativos os padrões encontrados, mais o modelo será capaz de gerar informações precisas de prognóstico. Sem padrões, não há aprendizado, o modelo não é capaz de predizer nenhum caso subsequente. O modelo final construído pode ser considerado híbrido no sentido de que a aprendizagem é obtida a partir de técnicas de aprendizagem de máquina e reconhecimento de padrões. O modelo foi treinado usando uma abordagem supervisionada, cada característica (*feature*) inserida no modelo tem seu resultado correspondente identificado.

A máquina de estados apresentada na Figura 16 e o analisador de estados são responsáveis pela identificação dos padrões. A combinação de ambos é o que gera a aprendizagem do modelo. Além disto, melhora o seu desempenho com relação às informações de prognóstico. Analisando a Figura 4, novamente, a máquina de estados é uma técnica diretamente ligada às transições de estado dos alarmes. Ao longo do tempo, os alarmes mudam de estado conforme a falha evolui. Para mudar o estado correspondente, a máquina avalia o estado atual e o valor da entrada do atributo *AlarmState.* Na Tabela 1, é apresentada a função de transição da máquina de estado para as sequências de alarmes. Como exemplo é impossível que um alarme vá do estado ACK_ALM para o UNACK_ALM. O sistema não contempla este tipo de transição para os alarmes.

Tabela 1. Função de transição da máquina de estados.

Estado Atual	Entrada	Novo Estado
ACK_ALM	ACK_RTN	ACK_RTN
UNACK_ALM	UNACK_RTN	UNACK_RTN
UNACK_ALM	ACK_ALM	ACK_ALM
ACK_RTN	UNACK_ALM	UNACK_ALM
UNACK_RTN	ACK_RTN	ACK_RTN

Fonte: Autoria Própria.

A máquina de estados será aplicada na identificação de alarmes potencialmente causadores do desligamento, pois as mudanças de estado do alarme ao longo do tempo refletem a evolução da falha correspondente. A máquina não é capaz de realizar transições diferentes da destacada na **Tabela 1**.

O analisador de estados é um algoritmo cuja técnica consiste em avaliar o último estado em que o alarme esteve durante o período de produção, para verificar em qual situação ele se encontrou no momento em que a produção foi finalizada.

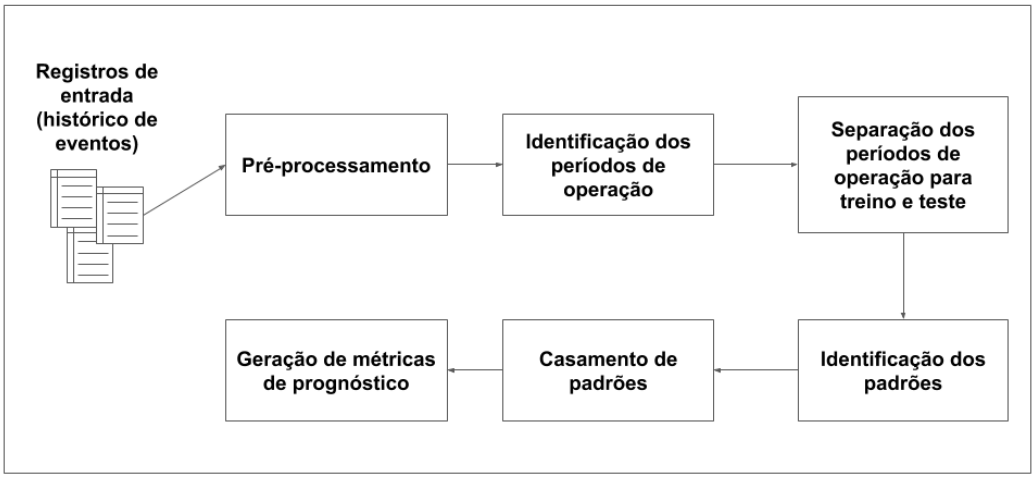
A aprendizagem consiste em identificar os períodos de operação e gerar padrões, formados por alarmes em sequência, que podem ter causado o desligamento. A base de dados gerada consiste no conjunto de conhecimento que o modelo tem para gerar informações sobre novas sequências de alarmes emitidas. Um padrão sempre será formado por uma sequência de alarmes identificados no estado UNACK_ALM ao término do período de operação.

3.5.1 Modelo de Aprendizagem

O modelo de aprendizagem é construído e avaliado segundo, respectivamente, as etapas de treino e de teste. O tratamento de instâncias “problemáticas” (com falhas) de operação, a separação do conjunto em treino e teste para a validação cruzada e a identificação dos períodos de operação fazem parte da fase de treinamento. A fase de treinamento busca encontrar padrões nas sequências de alarmes e salvá-los no banco de dados para análise posterior. Para isto, são executadas em sequência, as etapas definidas na Figura 10 de pré-processamento dos dados, identificação dos períodos de operação, separação dos períodos de operação para identificação de padrões e teste e a identificação dos padrões. Na segunda parte, em que o modelo já é capaz de gerar informações de prognóstico para os operadores, podem ser realizados os casamentos de novas sequências de alarmes com os padrões e com o número de casamentos encontrados e geradas as métricas, o que representa o objetivo principal do modelo de predição.

Além das etapas definidas na Figura 10, há outra executada em separado, quando há necessidade e após a execução da etapa de treinamento e de teste, que é a geração do limiar de predição. O limiar poderá ser recalculado sempre que houver novos registros de alarmes.

Figura 10. Etapas necessárias para treinar e testar o modelo de predição.



Fonte: Autoria Própria.

Treinamento

Pré-Processamento: os dados são lidos dos arquivos gerados pelo sistema de controle da usina. Os registros problemáticos (inválidos) são tratados ou removidos. Esta fase é composta pelas etapas de ETL, definidas na Seção 4.1.

Identificação dos períodos de operação: são delimitados os intervalos de interesse correspondentes à produção de energia da UTE, nos quais ocorreram os alarmes e onde serão analisadas as correspondências entre alarmes e desligamentos.

Separação dos períodos de operação para treino e teste: são separados os períodos de operação, uns usados no treinamento e outros no teste. De qualquer forma, todos os períodos passarão pela etapa de identificação de padrões, porém alguns serão usados para gerar uma base de conhecimento formada pelos padrões, que servirão para os cálculos das métricas associadas aos alarmes e da relação destes com o desligamento. O restante dos períodos será usado na avaliação de desempenho do modelo.

Identificação dos padrões: usando a máquina de estados, a relação entre os alarmes de cada período e os desligamentos abruptos ou com os desligamentos normais, é identificada a partir dos estados assumidos pelos alarmes ao longo do tempo.

Teste

Casamento de padrões: novas sequências são “casadas” com os padrões encontrados na fase de treinamento, na tentativa de identificar a sequência dentro dos padrões.

Geração de métricas de prognóstico: o número de casamentos encontrados é a base para o cálculo das métricas de probabilidade, tempo médio e predição, que serão os valores contidos no conjunto de métricas de prognóstico.

3.5.2 Métodos para Geração de Prognóstico

O método para a geração de prognóstico leva em consideração alguns aspectos relevantes acerca dos padrões gerados. Inicialmente, não há disponibilidade de padrões extraídos. Porém, a partir dos períodos de operação, as máquinas de estados são usadas para acompanhar os alarmes e identificar quais possuem fortes indícios de que causaram o desligamento abrupto.

Pode ser utilizada uma abordagem supervisionada para construção do modelo de predição, em que os padrões são gerados com base na ocorrência do desligamento abrupto presente nos períodos de operação. Se não houver um desligamento, então não há padrão, dado que neste estudo, este conceito está ligado ao desligamento.

Sempre que uma nova sequência de alarmes ocorrer, ela será casada com os padrões disponíveis. A parte da nova sequência de alarmes casada com algum padrão disponível consiste em um episódio. Para encontrar os episódios de uma forma precisa, a parte de casamento de sequências de alarmes foi combinada com as máquinas de estados. A Máquina tem um papel fundamental na identificação dos padrões, pois o casamento acontece entre a nova sequência gerada pelo sistema de controle e os padrões identificados pelas máquinas de

estados, dando a possibilidade de tornar o modelo mais promissor quanto à predição de desligamentos abruptos.

3.6 Discussão

O objetivo deste capítulo foi apresentar os conceitos relevantes à modelagem da abordagem desenvolvida.

Foram apresentadas questões importantes referentes ao sistema de alarmes, aos motores e ao sistema de controle da UTE. Neste contexto, são indispensáveis o entendimento sobre os estados dos alarmes, os períodos de operação, a existência e o funcionamento dos eventos, como ocorre a ativação de um alarme, características das variáveis de processo no que consiste a aprendizagem, as funções de transição das máquinas de estado, quais etapas constituem o treinamento e quais os principais componentes envolvidos na etapa de aprendizagem. O conhecimento desses conceitos é de suma importância para o entendimento da pesquisa ora descrita.

No capítulo seguinte, será descrita a abordagem proposta.

Capítulo 4

Abordagem para a Predição de Desligamentos Baseada no Histórico de Eventos Emitidos por Motores de uma Usina Termoelétrica

Neste capítulo, será apresentada a abordagem de solução para predição de desligamentos. A proposta consiste em um modelo, capaz de predizer se irá ocorrer um desligamento e, além disto, gerar informações de prognóstico a partir das sequências de alarmes.

Devido a questões de confidencialidade relacionadas ao projeto de pesquisa no qual este estudo está inserido, algumas informações não serão apresentadas. Entretanto, a omissão dessas no documento, não compromete a descrição e validação da abordagem. Serão omitidas informações sigilosas da UTE, decisões e princípios que norteiam a empresa como nomes de funcionários não envolvidos neste estudo, valores monetários de qualquer natureza, telas do sistema de controle e de outros sistemas internos da UTE, descrição dos equipamentos e detalhes do processo de produção de energia. Os dados dos eventos não serão inteiramente disponibilizados, será apresentado somente o essencial que mantenha a integridade e validação dos resultados da pesquisa.

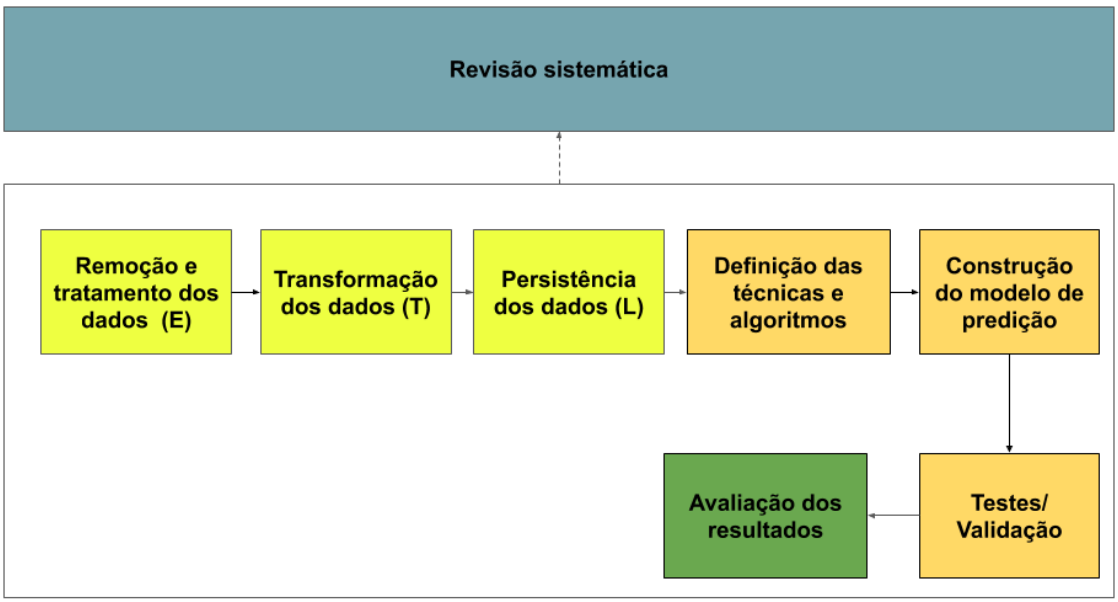
Com base nos dados de prognóstico e na indicação de que haverá ou não desligamento, cada operador poderá utilizar seu próprio conhecimento sobre a situação, juntamente com as informações extras geradas pelo modelo de predição, para tirar melhores conclusões e diagnosticar as falhas de forma mais precisa. O modelo de predição poderá ser usado pela central de controle e servirá como ferramenta de *apoio à tomada de decisão*, uma área que vem crescendo muito nos últimos anos (SOARES, 2016). O sistema será útil para a usina no processo de FDDC, melhorando a precisão dos diagnósticos por parte dos operadores.

4.1 Abordagem para Execução da Pesquisa

Na Figura 11, estão apresentadas as etapas definidas para execução deste estudo. Inicialmente, tornou-se necessário extrair e tratar os dados (E), transformá-los (T), fazendo com que novos dados sejam gerados a partir dos originais, e persisti-los (L) em uma base de dados para uso posterior. Após estas fases, foram executadas as etapas de definição das técnicas e algoritmos, criação do modelo de predição, testes/validação e análise dos resultados. Concomitantemente à evolução destas etapas, foi realizada uma revisão bibliográfica em busca do estado da arte sobre aprendizagem de máquina, reconhecimento de padrões e predição de eventos.

As etapas da pesquisa estão definidas na Figura 11, as quais serão descritas a seguir.

Figura 11. Etapas para execução da pesquisa.



Fonte: Autoria Própria.

Extração e tratamento dos dados (E): Os dados são lidos dos arquivos gerados pelo sistema de controle. Os registros problemáticos (inválidos) são tratados ou removidos.

Transformação dos dados (T): Novas informações são geradas a partir do conjunto gerado na etapa de Extração e tratamento dos dados. Por exemplo, o campo identificador de alarme foi gerado a partir do campo “Tag do alarme” e o número do equipamento a partir do campo “Area”.

Persistência dos dados (L): Os dados gerados nas fases de E e T são persistidos no banco de dados.

Definição das técnicas e algoritmos: São definidos quais algoritmos serão utilizados e quais técnicas de aprendizagem de máquina, para o reconhecimento de padrões e predição de eventos. Estes algoritmos e técnicas são a base para a implementação de outras técnicas usadas na construção do modelo de predição.

Construção do Modelo de predição: Esta etapa consiste na construção do modelo de predição, com execução das fases de treinamento e teste.

Testes e validação: A fase de testes é relacionada ao modelo de predição. Esta é a fase em que o modelo é submetido a um conjunto de testes para avaliação dos resultados das predições e prognósticos, com o intuito de verificar o desempenho do modelo.

Avaliação dos resultados: Os resultados obtidos na etapa de “Testes e validação” são discutidos, avaliados e tiradas as conclusões em relação à hipótese de pesquisa definida no Capítulo 1.

Por se tratar de uma tarefa majoritariamente de aprendizagem de máquina, na próxima seção será descrita a fase de implementação da Predição de Desligamentos.

4.2 Abordagem para a Predição de Desligamentos

Para treinar o modelo completo, a decisão de utilizar aprendizagem de máquina com abordagem supervisionada, usando reconhecimento de padrões e informações de prognóstico, surgiu devido à necessidade de obter diversas informações extras, a exemplo da chance de um alarme ocorrer ou a sequência em que esse alarme se encontra, a chance de um alarme levar a um

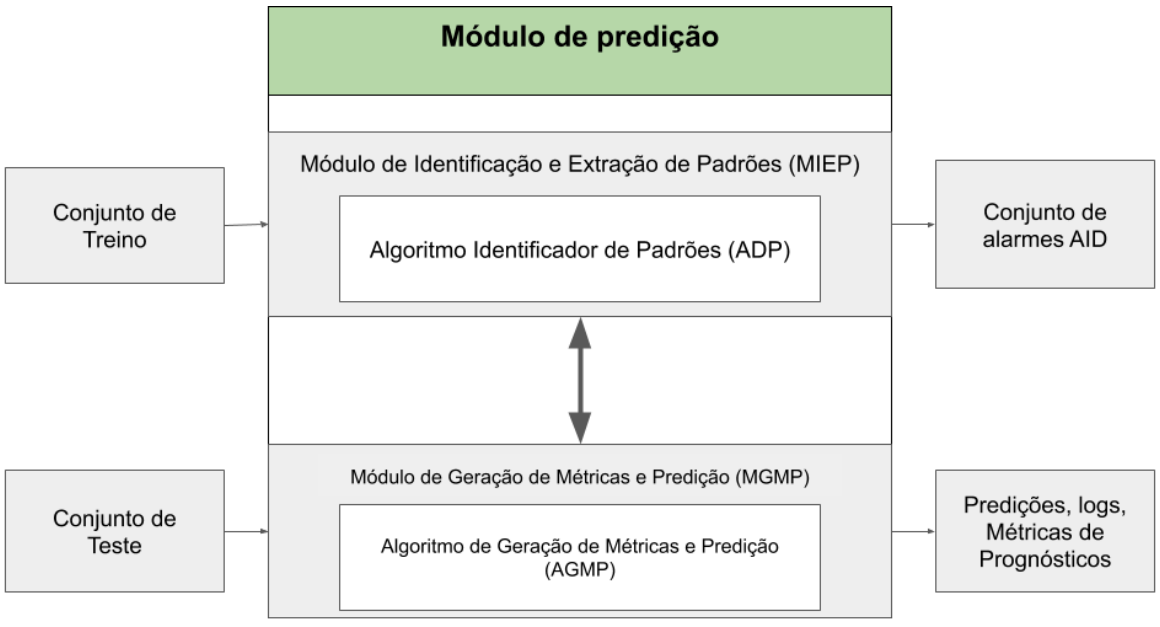
desligamento, o tempo médio da sequência inteira e do último alarme desta sequência levar ao desligamento e por último, a predição de que vai ocorrer o desligamento abrupto.

O objetivo principal com a construção do modelo de predição é que, dado um novo alarme, o modelo é capaz de gerar informações de prognóstico que auxiliem no processo de FDDC.

A construção do modelo é baseada em cinco técnicas (componente 1: Identificação de padrões, construção da Máquina de estados; componente 2: Casamento de padrões, Geração de Métricas e componente 1 e 2: Aprendizagem de máquina).

O modelo é formado pelo Módulo de Identificação de Padrões (MIP) e Módulo de Geração de Métricas e Predição (MGMP). Conforme Figura 12, no MIP é executado o Algoritmo para a Identificação de Padrões (ADP) e no MGMP é executado o Algoritmo de Geração de Métricas e Predição (AGMP).

Figura 12. Modelo de predição.



Fonte: Autoria Própria.

Componente 1: A primeira componente do modelo é o Módulo de Identificação de Padrões composta pelo ADP, que recebe como entrada o conjunto de eventos

tratados e transformados nas fases de ETL, os ordena em sequência pelo instante de tempo e, em seguida, identifica padrões nestas sequências. A saída da Componente 1 será um conjunto de sequências de alarmes consideradas padrões.

Componente 2: A segunda componente é o Módulo de Geração de Métricas e Predição composta pelo Algoritmo Gerador de Métricas e Predição, que recebe uma nova sequência de alarmes e busca identificá-la em alguma outra que está no conjunto gerado pela Componente 1. A saída da componente 2 serão as métricas definidas na Seção 4.3.2.

Saída: A saída do modelo, na fase de treinamento, será o conjunto de padrões identificados pela Componente 1 e, na fase de testes, será a saída da Componente 2.

As componentes que constituirão o modelo de predição são, portanto: o ADP (componente 1), para identificar padrões na sequência de alarmes, através da avaliação dos estados que o alarme assume ao longo do tempo e o AGMP (componente 2), para calcular métricas e gerar uma predição de que haverá ou não desligamento abrupto a partir de uma nova sequência de alarmes emitida pelo sistema de controle.

Por ser um problema de aprendizagem de máquina, duas etapas são necessárias para a construção e avaliação do modelo: treinamento e teste. A primeira é a execução da componente 1, sob um conjunto de treinamento para construção do conjunto de padrões. A segunda, a de testes, baseia-se na execução das componentes 1 e 2 em um conjunto de testes. A componente 1 é executada para identificar os padrões. A saída dessa componente será a entrada da componente 2. Ao executar a fase de testes, o algoritmo busca casar cada padrão de teste com os padrões gerados na fase de treinamento. Sempre que uma nova sequência é identificada nos padrões, o algoritmo encontrou um *episódio*. Após encontrar os episódios, serão geradas as informações de prognóstico, a saber: as probabilidades, o tempo médio até o desligamento e a predição de que haverá desligamento abrupto, todas calculadas a partir dos episódios encontrados.

4.3 Construção do Modelo de Predição

Os algoritmos para a construção e execução do modelo de predição estão apresentados na Figura 13. O algoritmo para extração de períodos de operação (AEP) implementa a demarcação dos períodos de operação da UTE, identificando um par de eventos: o evento “Sync on”, que indica que o motor foi inicializado, e o evento de desligamento, que será ENGSTO, caso tenha ocorrido um desligamento normal e S011SDI, se for abrupto. O segundo Algoritmo é o responsável pela separação dos conjuntos de treino e teste (ASTT). Este algoritmo implementa a recepção do conjunto dos períodos de operação do AEP separando-o em dois: treinamento e teste.

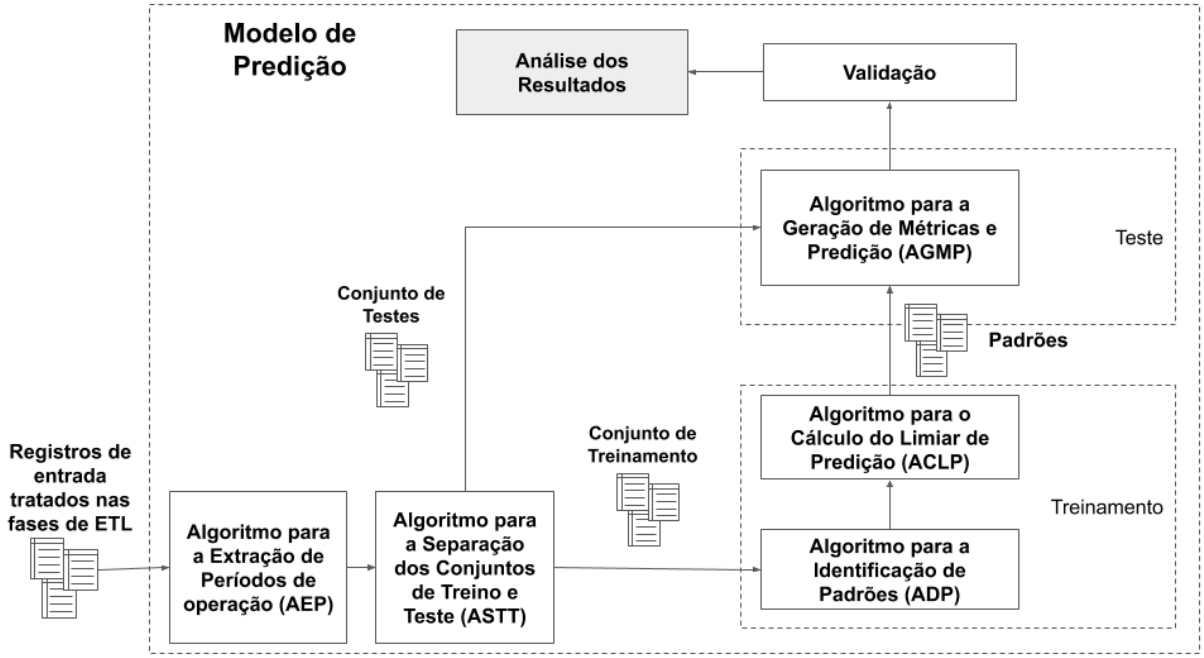
O conjunto de todos os períodos de tempo considerados foi dividido em dez subconjuntos de tamanho, aproximadamente, igual. A técnica utilizada para avaliar o modelo de predição foi a validação cruzada, com método k-fold e k igual a 10. Esta técnica permite que o modelo seja analisado de uma forma mais precisa, pois leva em consideração diferentes perspectivas do conjunto de dados. Por isto, foi escolhida como técnica para avaliação de desempenho do modelo.

O terceiro algoritmo é o de identificação de padrões. Se um alarme aconteceu antes de um desligamento abrupto e permaneceu no estado UNACK_ALM até o momento do desligamento, esse é um forte candidato a estar relacionado ao desligamento. Por isto, um alarme com estas características é considerado um Alarme Indicador de Desligamento (AID). Um *padrão* é uma sequência de alarmes AID ordenados pelo tempo. Na Seção 3.2, estão descritos os detalhes sobre os estados que um alarme pode assumir ao longo do tempo.

A partir dos padrões gerados, é possível encontrar uma métrica importante para o modelo de predição, o limiar para indicar se ocorrerá ou não desligamento. Antes de persistir os dados, é executada uma fase intermediária para o cálculo desta métrica. Este limiar é uma probabilidade (um número real), encontrado a partir da média entre as probabilidades de todos os alarmes levarem ao desligamento. Consiste em um valor limite para o modelo de predição indicar se vai ocorrer ou não um desligamento. Para um valor acima do limiar, o modelo indica desligamento e caso contrário, indica que não haverá desligamento.

Finalizada a fase de treinamento, inicia-se a fase de testes, na qual é executado o AGMP. Este algoritmo necessita de dois conjuntos de dados: o conjunto de testes e o conjunto de padrões gerados pelo AIPD. Para cada período do conjunto de testes, uma sequência de alarmes AID é extraída e, posteriormente, o algoritmo busca em cada padrão a sequência de alarmes AID identificada no período atual. Quando ocorre o casamento com algum padrão, o algoritmo incrementa uma contagem, para ao final estimar as probabilidades de a sequência levar ou não levar a um desligamento abrupto.

Figura 13. Diagrama de execução dos algoritmos para Construção e Execução do Modelo de Predição.



Fonte: Autoria Própria.

Detalhes complementares sobre o funcionamento de cada um dos módulos do modelo de predição, sobre o cálculo das métricas de desempenho e sobre a validação do modelo, estão descritos no Capítulo 5.

4.3.1 Técnicas Aplicadas

O sistema de predição de desligamentos considerado neste estudo considera três técnicas.

Técnica 1: Responsável pela geração de um conjunto de padrões, em que cada padrão é uma sequência de alarmes. Cada alarme desta sequência foi encontrado em um estado considerado de falha (ver Seção 3.2) no instante em que o desligamento ocorreu. Quando isto acontece, caracteriza-se a ocorrência de um AID. O algoritmo que implementa esta técnica foi denominado de *identificador de padrões*, o qual faz parte da componente 1.

Técnica 2: Responsável por, dado uma nova sequência de AID, realizar casamento de padrões para identificar quantas vezes a nova sequência ocorreu dentro do conjunto de padrões gerados pela técnica 1. O algoritmo que implementa esta técnica foi denominado de *gerador de métricas*, o qual faz parte da componente 2.

Técnica 3: Responsável pela predição do desligamento a partir das métricas calculadas pela técnica 2. O algoritmo que implementa esta técnica foi denominado de *preditor*, o qual parte da componente 2.

4.3.2 Máquinas de Estados e Métricas geradas pelo AGMP

A máquina de estados é uma técnica computacional (WAGNER, 2006) capaz de capturar as transições de estado de um sistema, que recebe um sinal como entrada e, a partir de uma função de transição, usando o sinal e o estado atual, a máquina pode mudar ou não para outro estado. Ao ser executada desta forma, a máquina é capaz de capturar mudanças ao longo do tempo.

A máquina implementada considera duas técnicas para reconhecimento de padrões.

Técnica 1: responsável pela geração de uma lista de máquinas de estados, em que cada máquina acompanha um alarme encontrado na sequência principal. O algoritmo que implementa esta técnica foi denominado de *Identificador de Estados*.

Técnica 2: responsável pela análise dos resultados de cada máquina gerada pela técnica 1. O algoritmo que implementa esta técnica foi denominado de *Analisador de Estados*.

O objetivo do acompanhamento e análise das transições é descobrir os AID que consistem em alarmes, em que antes de ocorrer um desligamento, acabaram em um estado diferente de ACK_RTN (ver Seção 3.2), que indicam que as falhas associadas não foram devidamente corrigidas.

A Máquina de Estados é usada para identificar os alarmes em que, dado que houve um desligamento, terminaram no estado ACK_ALM, conforme apresentado na Seção 3.2. Uma máquina é criada para cada tipo de alarme ao longo do tempo, com o objetivo de acompanhar as transições de estado de cada alarme e, ao final do período de operação, identificar quais desses são AID.

O Algoritmo para a Geração de Métricas e Predição (AGMP) é executado para calcular o número de vezes em que a sequência aconteceu no histórico de padrões e também estimar o tempo médio até a ocorrência do desligamento abrupto. Além destas métricas, outras são calculadas. As métricas utilizadas para geração de prognósticos são as seguintes:

- Total de vezes em que a sequência ocorreu;
- Probabilidade de a sequência levar a um desligamento abrupto;
- Probabilidade de a sequência não levar a um desligamento abrupto;
- Probabilidade de o último alarme da sequência levar a um desligamento abrupto;
- Probabilidade de o último alarme da sequência não levar a um desligamento abrupto;
- Tempo médio entre a sequência e o desligamento abrupto;
- Predição do modelo (se vai ocorrer ou não desligamento abrupto); e
- Número do equipamento no qual ocorreu a sequência.

4.4 Discussão

Neste capítulo, foram apresentadas as etapas para execução da pesquisa, desenvolvimento de um método para predição de desligamentos utilizando um

modelo de predição, que busca unir dois componentes que utilizam técnicas de reconhecimento de padrões.

Foram definidas as componentes do modelo de predição, bem como os dados que constituirão suas entradas e técnicas utilizadas para construção de cada componente, bem como os componentes que constituirão o modelo híbrido: o modelo de predição construído a partir de técnicas de reconhecimento de padrões (componente 1, construída para identificar padrões nas sequências de alarmes através da avaliação dos estados que os alarmes assumem ao longo do tempo). Também foi definido que, com os padrões gerados, são calculadas métricas, com o intuito de gerar um prognóstico do sistema e indicar se haverá ou não um desligamento abrupto (componente 2).

Esta combinação busca realizar a predição de forma adequada aos propósitos da pesquisa, para que o prognóstico possa ser realizado em tempo real e de forma acurada. O modelo final apresentado tem uma característica híbrida, por combinar técnicas diferentes para atingir o propósito desta pesquisa. A etapa de treinamento do modelo de predição ocorre quando o algoritmo de identificação de padrões, que utiliza máquinas de estados, é executado.

Novas sequências de alarmes são geradas quando há um novo treinamento, e se há novos alarmes, existirão também novos padrões e, consequentemente, as métricas e as predições serão mais precisas. Isto é o que proporciona a característica de aprendizagem do modelo.

Na fase de testes e validação (conforme Figura 4), é aplicada uma lógica de decisão para gerar um alerta de desligamento. O alerta só será gerado, assim como a predição do modelo de predição, se a probabilidade de ocorrer um desligamento for superior a um limiar. Este limiar é calculado a partir de todos os padrões encontrados com a execução do AIP.

No Capítulo 5, são apresentados e discutidos os resultados obtidos com o uso da abordagem desenvolvida.

Capítulo 5

Metodologia

Neste capítulo, serão apresentadas as técnicas e algoritmos de aprendizagem de máquina, para reconhecimento de padrões, utilizados para construção do modelo de predição.

Para este estudo, foi utilizada uma amostra de dados referente ao período de janeiro de 2017 a novembro de 2018. Os registros de alarmes do sistema de controle são salvos em arquivos com extensão txt. O conjunto de arquivos com extensão .txt coletado no período considerado representa o log de alarmes e constitui o conjunto de dados deste estudo. As fases de ETL são executadas usando este conjunto de dados. Conforme as etapas vão sendo executadas, os conjuntos de entrada de dados de uma etapa representam o conjunto de dados da etapa anterior.

5.1 Pré-Processamento, Extração e Transformação dos Dados (ETL)

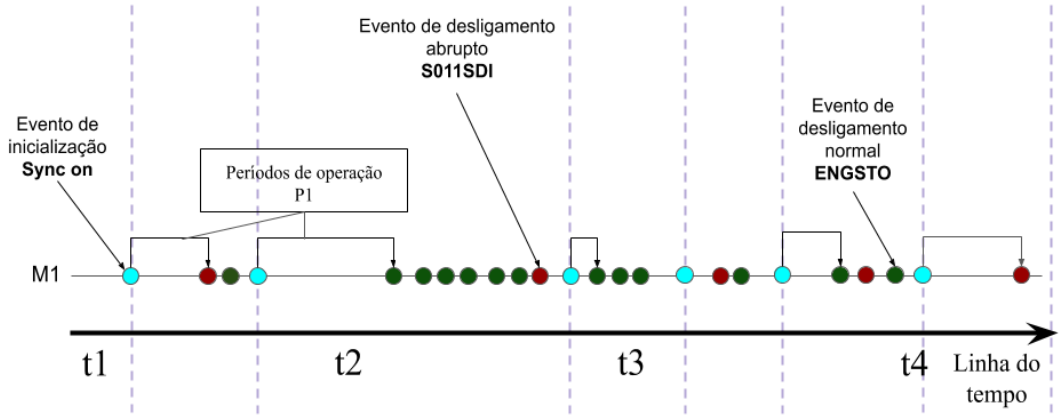
A fase de pré-processamento, é constituída por três das etapas apresentadas na Seção 4.2 (destacadas em amarelo), que são as etapas de ETL. Nestas etapas, os dados do log de alarmes serão lidos, tratados, transformados, novas informações serão extraídas e, por último, os dados obtidos deste processamento serão persistidos na base de dados, nesta sequência.

Os registros que possuem campos com valores inconsistentes são descartados. Além disto, pondo os alarmes em sequência ordenada pelo instante de tempo, conforme mostrado na Seção 3.2, foram identificados os períodos de tempo em que os motores operaram. Na Figura 14, todos os alarmes do motor M1 estão ordenados pelo instante de tempo em que ocorreram. Um filtro é executado para selecionar os eventos de inicialização do motor (círculos de cor azul).

Em seguida, o algoritmo AIP busca, entre um evento de inicialização e outro, o primeiro evento de desligamento, seja abrupto (círculos vermelhos) ou

não (círculos verdes) e, quando encontra, separa o par de eventos de inicialização e desligamento e define o início do período como sendo o instante de tempo do evento de inicialização e o fim como sendo o instante de tempo do desligamento. Se entre um evento de inicialização e outro, o algoritmo encontrar mais de um desligamento, é selecionado o primeiro e ignorado o restante. Se não encontrar nenhum, é descartado o evento atual de inicialização.

Figura 14. Algoritmo de Identificação de Períodos.



Fonte: Autoria Própria.

O tratamento dos dados consiste na verificação nos campos dos registros de alarmes, a exemplo de como testar os campos para verificar se algum dos valores está fora dos padrões especificados no manual do motor da UTE. Caso seja encontrado, será descartado o registro problemático (inválido). Por exemplo, segundo o manual, o formato do instante de tempo em que o alarme ocorreu deve ser o seguinte: MÊS/DIA/ANO HORA:MINUTOS:SEGUNDOS.MILISSEGUNDOS.

Assim, a data 03/22/2017 00:04:56.560 atende ao padrão aceitável, enquanto que 03/22 00:04 não atende. Por não ter como recuperar os valores perdidos e por não apresentar criticidade alta para os resultados do estudo, todos os registros problemáticos foram descartados. Após a transformação, os dados são salvos no banco de dados. No Apêndice A, estão descritos mais detalhes sobre outros campos dos registros de alarmes.

Após o processo de ETL, inicia-se a fase de transformação, que consiste em usar os dados filtrados e pré-processados para extrair novas informações para a construção do modelo de predição. Os campos dos registros brutos vindos do

sistema de controle apresentam informações limitadas, por exemplo, esses não possuem o número do motor no qual o alarme foi emitido, porém, esta informação pode ser extraída do atributo Area. O identificador do alarme (ID) também não é um dos campos do registro de alarmes. Ao invés do ID, o sistema emite a *Tag do Alarme* que possui não só o ID, mas a localização de onde esse foi emitido na planta, como o setor, a área, o conjunto de equipamentos e o equipamento específico. Cada dígito tem um significado. Na Tabela 2 e na Tabela 3, são apresentados os dados originais e os dados com os novos campos extraídos, respectivamente.

Tabela 2. Registros originais de alarmes do sistema de controle.

EventStamp	AlarmTag	Description	Operator	Value	Area
03/22/2018 02:03:54.569	PCA901M001MCE	HFO feed pump 1 control in manual pos.	OP	ON	GENSET_1
03/22/2018 02:03:54.569	PCA901M001MCE	HFO feed pump 1 control in manual pos.	ADM	ON	GENSET_1
03/22/2018 02:03:54.569	PCA901M001MCE	HFO feed pump 1 control in manual pos.	ADM	OFF	GENSET_1

Fonte: Autoria Própria.

Tabela 3. Registros com os novos campos extraídos dos originais.

EventStamp	...	Area	**Date**	**EquipmentNumber**	**Alarm**
03/22/2018 02:03:54.569	...	GENSET_1	**03/22/2018**	**1**	**M001MCE**

Fonte: Autoria Própria.

Além dos campos exibidos na Tabela 3, outros campos são extraídos, como a quantidade de vezes em que o alarme ocorreu e também se, no período de operação em que o alarme foi emitido, o motor foi abruptamente desligado por causa de uma falha ou se foi desligado porque a geração de energia foi concluída com sucesso.

5.2 Treinamento

Nesta seção, serão apresentados os detalhes do treinamento do modelo, todos os algoritmos, técnicas e decisões relacionadas à fase de construção do modelo de predição.

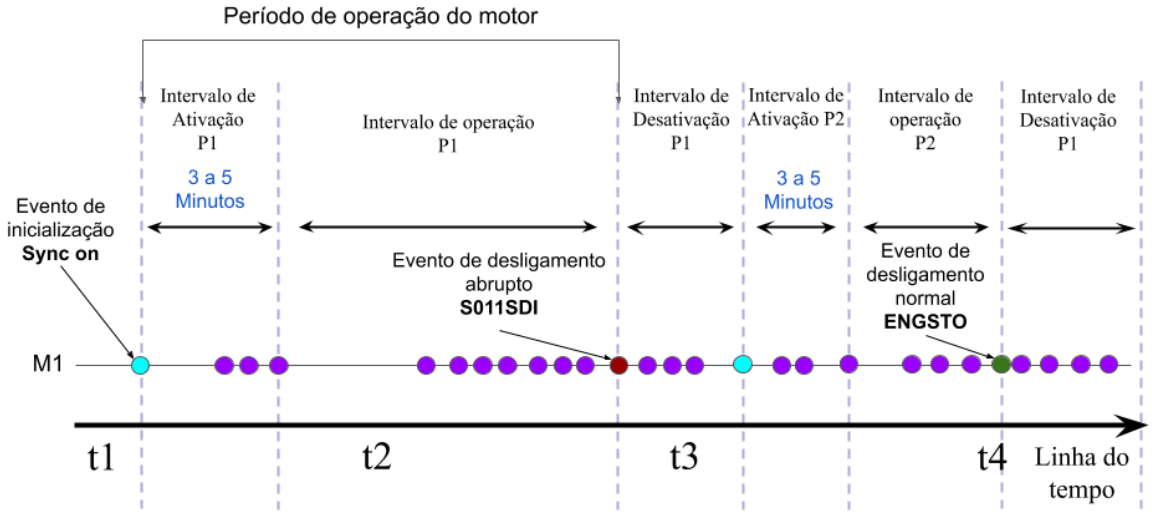
5.2.1 Módulo de Identificação de Padrões

Ao finalizar as etapas de ETL, inicia-se o treinamento do modelo com a execução dos algoritmos do MIP. Os períodos de operação são identificados pelo Algoritmo de Extração de Períodos (AEP) e são a entrada do AIP.

Na Figura 15, há um exemplo hipotético, em que estão destacados os períodos de operação. O trabalho do AEP é identificar os períodos de operação e subdividi-los em intervalos menores que detalhem o processo de partida e desligamento do motor. Quando o motor é ligado, um intervalo de tempo é necessário para a inicialização completa, uma exigência para que o motor atinja os níveis aceitáveis de tensão e potência para operar de forma correta. Este intervalo de partida não pode ser considerado nas análises deste estudo porque irão afetar os resultados de forma incorreta. Por isto, os alarmes gerados durante o período de ativação são sempre descartados.

Quando o operador aciona um motor, um evento com descrição *Sync On* é emitido. Este evento indica que o processo de inicialização começou. Após a emissão deste evento, o motor estará pronto para operar, entre 3 a 5 minutos, conforme definido pelos especialistas da UTE. Assim, definiu-se que o intervalo de operação a ser considerado tem início no instante de tempo correspondente à soma do instante de tempo do evento de inicialização Sync On com 5 minutos. Desta forma, o período de ativação do motor será desconsiderado.

Figura 15. Extração dos alarmes dos períodos de operação.



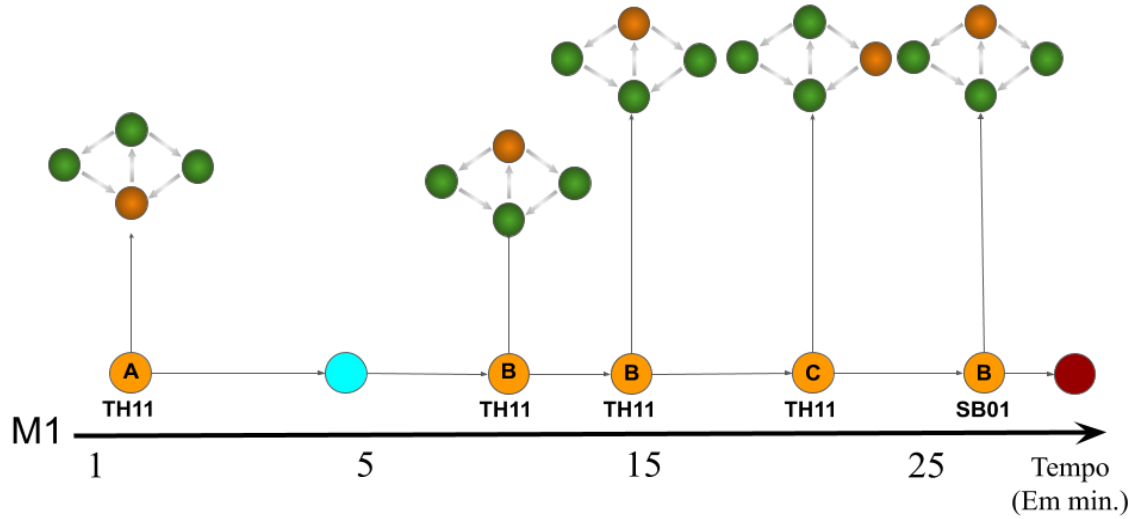
Fonte: Autoria Própria.

Sempre que considerar necessário, a UTE poderá repovoar a base de dados, executando novamente a fase de treinamento. Quando novos alarmes forem emitidos ao longo do tempo, novos padrões poderão ser identificados. Ao treinar novamente o modelo, a base de dados é excluída e os novos padrões poderão ser identificados.

5.2.2 Máquinas de Estados

A Técnica 1, definida na Seção 3.3, é o *Identificador de Padrões*, implementada pelo AEP, responsável pela identificação dos alarmes com potencial de causar desligamento dentro dos períodos de operação. Esta técnica foi executada considerando os estados que os alarmes podem assumir. Na Figura 16, é apresentado um exemplo hipotético de quatro alarmes e suas mudanças de estado ao longo do tempo. Para acompanhar as transições, uma máquina de estados foi criada para cada alarme A, B e C. Sempre que uma cópia do alarme é gerada com estado diferente, a máquina reflete as mudanças realizando transição de estado (ver Seção 3.2), de modo que, ao encerrar a operação, cada máquina de estados estará em um estado. Se alguma máquina terminou no estado UNACK_ALM, então o alarme é considerado AID, conforme explicado na Seção 4.3.

Figura 16. Identificação dos alarmes que possuem relação com a falha usando a máquina de estados.

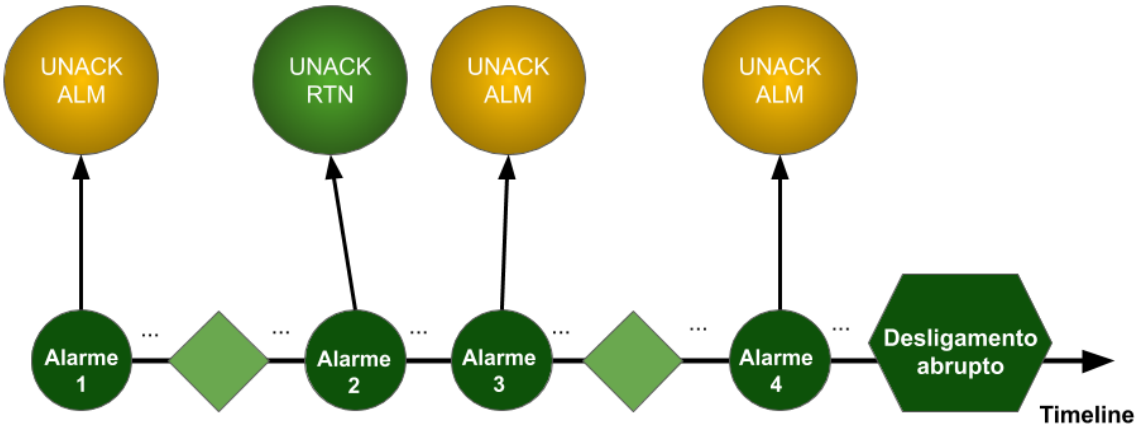


Fonte: Autoria própria.

A Técnica 2 é implementada pelo *Analisador de Estados*, que recebe como entrada as máquinas de estados resultantes do *identificador de padrões* e filtra aquelas em que os alarmes são AID.

Os alarmes AID são ordenados pelo instante de tempo e as sequências destes alarmes constituem os padrões. Na Figura 17, é apresentada uma sequência de alarmes identificada pelo Analisador de Estados. O padrão encontrado será formado pelo alarme 1, 3 e 4 nesta sequência, porque no instante em que o desligamento ocorreu, estes alarmes estavam no estado UNACK_ALM. O alarme 2 não se encontra neste estado, portanto não foi considerado AID e por isso foi ignorado. Após a identificação, todos os alarmes são ordenados para formar uma sequência que constitui um padrão. Todos os registros são persistidos na base de dados.

Figura 17. Detecção de sequências formadas por alarmes indicadores de desligamento.



Fonte: Autoria Própria.

Devido à quantidade de informações que ofuscam os verdadeiros efeitos de um conjunto de alarmes, a central de controle não consegue identificar que determinada sequência levará a um desligamento, o que reafirma a relevância da máquina de estados.

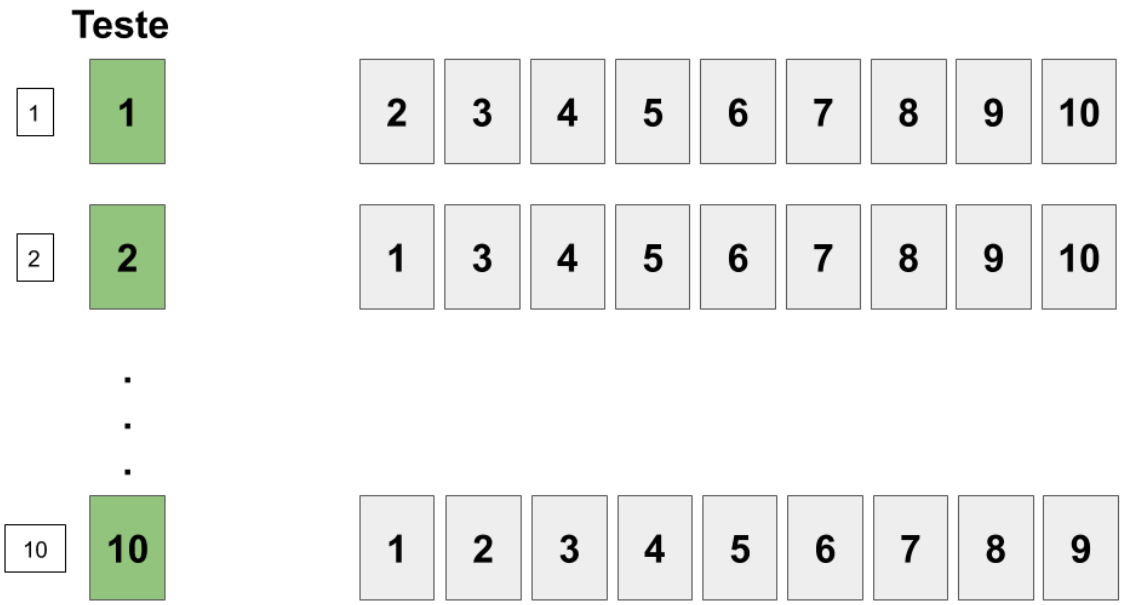
É importante destacar, que a máquina de estados não detecta relacionamentos entre os alarmes, mas gera um histórico de ocorrências que, diante de situações de desligamento abrupto, não foram tratadas.

5.2.3 Validação Cruzada

A etapa de validação cruzada (KOHAVI, 1995) consiste na utilização da técnica de validação cruzada usando o método *k-fold*, com *k* igual a 10, para avaliar o modelo sob diferentes perspectivas do conjunto de dados.

As etapas de treinamento e teste foram repetidas 10 vezes e, em cada uma, um conjunto de treino e teste diferentes é utilizado. Conforme a Figura 18, o conjunto total de períodos de operação foi dividido em dez partes. Para cada etapa da validação cruzada, a décima parte do conjunto foi utilizada para testes e o restante para treinamento. Cada conjunto foi rotulado com um número identificador de 1 a 10.

Figura 18. Definição dos conjuntos de treino e teste para as rodadas da validação cruzada.

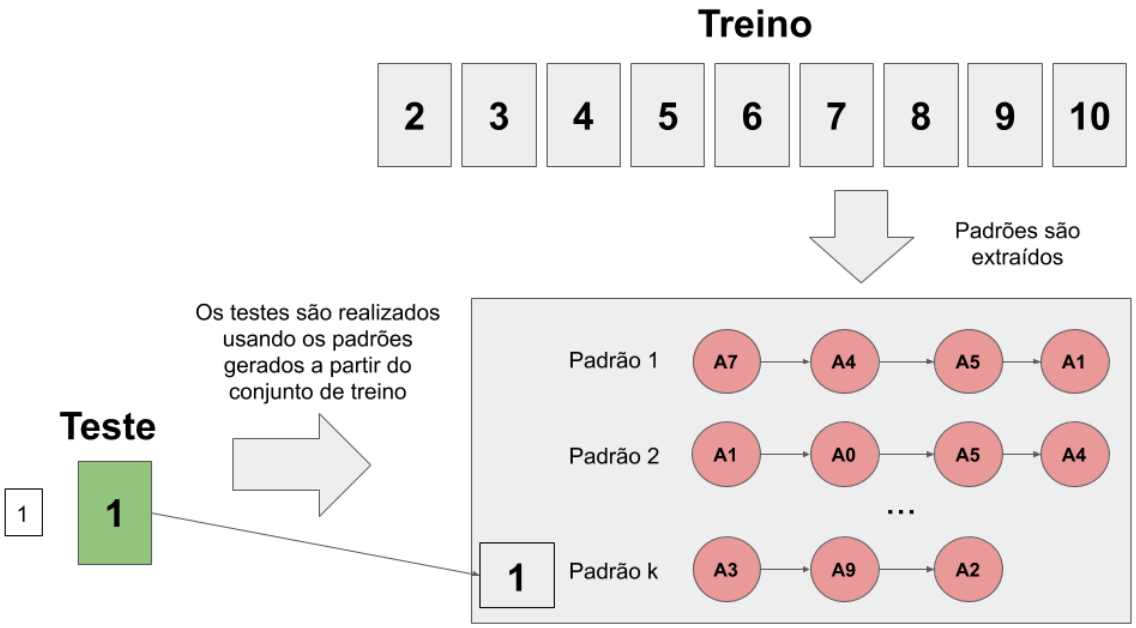


Fonte: Autoria Própria.

Após os conjuntos serem divididos em treino e teste, a etapa de teste, conforme mostrada na Figura 19, consiste no seguinte: dividir o conjunto total de períodos por 10. Um dos dez subconjuntos é separado para teste e cada registro é rotulado com um número que o identifica. Os outros nove subconjuntos são separados para treino. Os padrões são gerados a partir do novo conjunto de treino e cada padrão é rotulado com o mesmo número do conjunto de testes. Esta etapa é repetida até que cada um dos dez subconjuntos tenha sido usado para teste.

Quando o AGMP começa a executar, o primeiro conjunto de testes é selecionado e os padrões rotulados com o seu número são identificados. Então, o AGMP avalia os períodos do conjunto de teste, alarme por alarme e calcula as métricas e a predição a partir do conjunto de padrões correspondente gerado por meio do conjunto de treinamento, rotulados com o mesmo número do conjunto de teste corrente. Desta forma, o modelo é treinado com a décima parte do conjunto e testado com as nove partes restantes. Repetindo este procedimento 10 vezes, o modelo terá sido treinado e testado com todos os subconjuntos e uma avaliação mais adequada dos resultados será realizada.

Figura 19. Etapa de teste do modelo de predição usando o conjunto de testes 1.

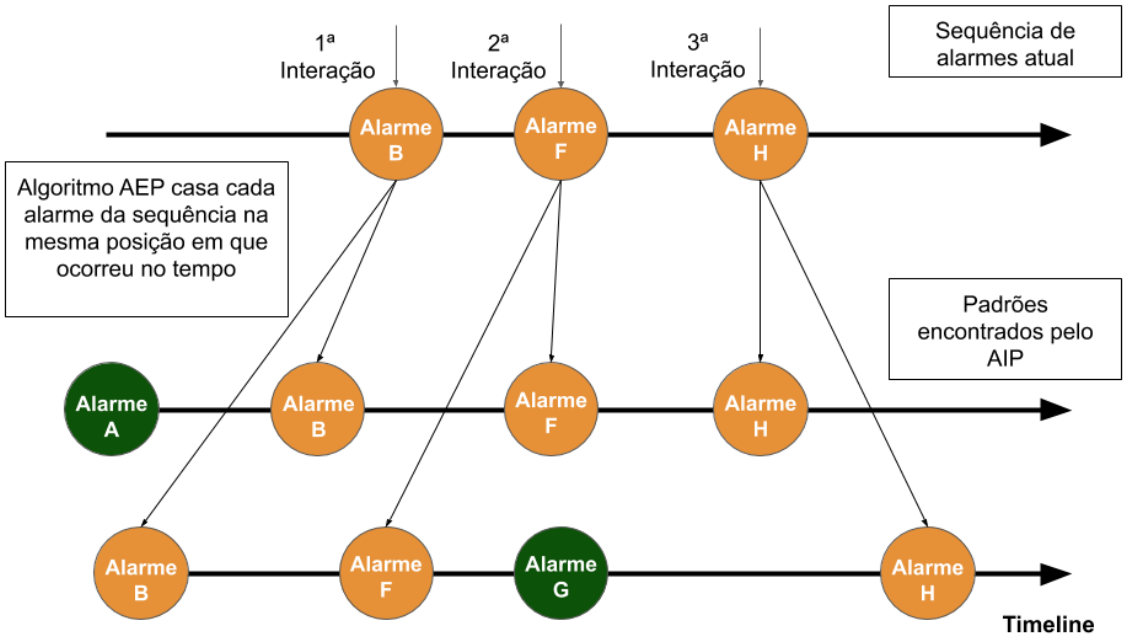


Fonte: Autoria Própria.

5.3 Módulo de Geração de Métricas e Predição

O MGP é responsável por identificar as sequências de alarmes dentro do conjunto de padrões identificados pelo AEP e, em seguida, gerar métricas. Sempre que o MGP identifica uma ocorrência dentro dos padrões, tem-se a ocorrência de um episódio. Conforme mostrado na Figura 20, o AGMP percorre cada um dos padrões dentro do conjunto de padrões e, para cada padrão, é escolhido o primeiro alarme da sequência atual emitida pelo sistema de controle e este alarme é procurado dentro do padrão atual (1ª interação). Quando este alarme é encontrado, avança-se para o próximo alarme da sequência atual, o qual é procurado dentro do mesmo padrão (2ª Interação), porém, considerando apenas os alarmes do padrão que ocorreram após o alarme encontrado na 1ª interação. Ao encontrar o segundo alarme no padrão atual, o AGMP seleciona o terceiro alarme da sequência atual e busca por este dentro do padrão (3ª interação), considerando apenas os alarmes que ocorreram após o alarme encontrado na 2ª interação. Ao encontrar todos os alarmes da sequência, o AEP encontrou um episódio. Após identificar cada episódio possível, são calculadas as métricas detalhadas na Seção 5.3.4. Caso algum dos alarmes não case em algum momento com o alarme do padrão, então não houve um episódio.

Figura 20. Identificação de episódios por parte do AGMP.



Fonte: Autoria Própria.

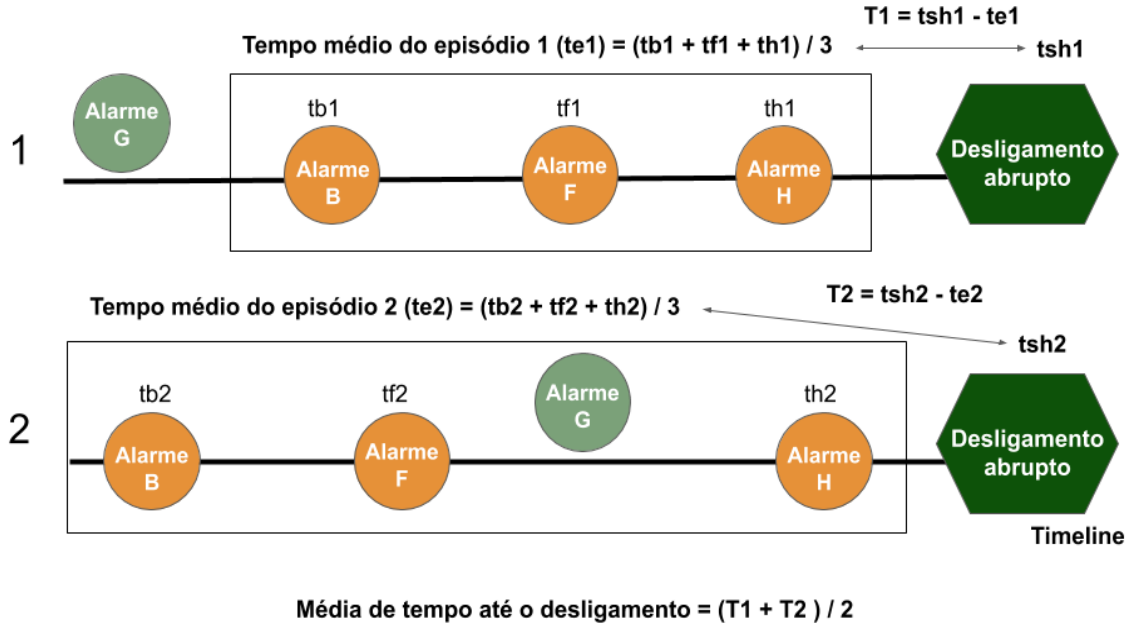
5.3.1 Tempo Médio até o Desligamento Abrupto

Além da contagem dos episódios, o AGMP calcula a média de tempo entre os episódios de uma determinada sequência e o desligamento abrupto. Com este valor, os operadores saberão quanto tempo levará, em média, até que o desligamento aconteça, caso o problema não seja sanado. Para cada episódio, o AGMP calcula o tempo médio do episódio a partir da soma dos tempos de seus alarmes dividida pelo número de alarmes contidos no episódio. A média dos tempos dos alarmes é uma métrica que se mostra adequada para a tomada de decisão, por representar melhor o tempo do episódio inteiro e não só de alguns de seus alarmes.

No exemplo apresentado na Figura 21, o instante de tempo tsh1 é o instante de tempo em que ocorreu o desligamento abrupto 1, tsh2 o desligamento abrupto 2 e assim por diante. Para o cálculo do tempo médio entre a sequência atual e o desligamento abrupto são utilizadas as diferenças entre todos os tempos dos desligamentos abruptos e os tempos médios dos seus episódios. Por exemplo, na Figura 21, a diferença de tempo T1 é calculada a partir da subtração entre o tempo do desligamento abrupto e o tempo médio do episódio te1. O

mesmo raciocínio é válido para T2. Ao final da execução, o algoritmo AGMP terá o tempo médio da sequência atual até o desligamento dado por (T1 + T2) / 2. O AGMP poderá fazer o mesmo para qualquer outra quantidade de episódios.

Figura 21. Cálculo da Média de tempo até o desligamento.



Fonte: Autoria Própria.

5.3.2 Métricas de Prognóstico

A informação predição sobre o desligamento abrupto, unicamente, não é útil para a central de controle. Além de informar se vai ocorrer ou não desligamento abrupto, é útil também que o modelo gere o tempo médio até que o desligamento ocorra e a frequência com que esta sequência ocorreu em períodos de operação, com ou sem desligamento abrupto. Essas informações em conjunto fornecerão ao operador um prognóstico mais completo do que pode acontecer, pois o ajuda a identificar, pelo número de vezes em que a sequência ocorreu, se esta é importante ou não. Caso uma sequência ocorra muitas vezes em períodos de operação normal, é natural afirmar que não ocorrerá nada de grave. Mas, se uma sequência for identificada em períodos de desligamentos, há indícios de que seja a causadora de desligamentos abruptos.

A experiência do operador, unida às informações de prognóstico geradas, proporcionará prognósticos mais fortes à central de controle de que a sequência

identificada levará ou não a um desligamento. Cada uma das métricas calculadas pelo MGMP está apresentada na Tabela 4. A probabilidade de uma sequência levar a um desligamento normal pode ser calculada levando-se em consideração a probabilidade da sequência inteira (pns) ou apenas do último alarme (pna). O mesmo vale para a probabilidade de levar a um desligamento abrupto, pois no cálculo pode ser considerada a probabilidade da sequência completa (pds) ou apenas a probabilidade do último alarme (pda) levar ao desligamento. Isso é importante porque um único alarme pode ser suficiente para levar a um desligamento abrupto.

Outras métricas consideradas são o tempo médio entre a sequência e o desligamento abrupto (tms) e o último alarme da sequência e o desligamento (tma). Por último, o MGMP indica uma predição de que vai ou não ocorrer um desligamento forçado, caso pda ou pds sejam maiores que um determinado limiar. Estas informações poderão auxiliar a central de controle nas tomadas de decisão.

Tabela 4. Exemplos de dados gerados pelo MGMP.

Nº do Motor	Total de Episódios	Probabilidade da sequência alarmes não levar a um desligamento abrupto (pns)	Probabilidade do último alarme da sequência não levar a um desligamento abrupto (pna)	Probabilidade de a sequência de alarmes levar a um desligamento (pda)	Probabilidade do último alarme da sequência levar a um desligamento (pds)	Tempo médio até o desligamento abrupto a partir do tempo médio da sequência (tm)	Tempo médio até o desligamento abrupto a partir do último alarme da sequência (tm)	Ocorrerá Desligamento?
3	5	0,3	0,8	0,4	0,3	10 minutos	5 minutos	Não
7	6	0,9	0,2	0,3	0,2	20 minutos	2 minutos	Não
20	8	0,3	0,1	0,6	0,9	10 minutos	1 minuto	Sim
1	4	0,4	0,4	0,3	0,9	20 minutos	13 minutos	Sim

Fonte: Autoria Própria.

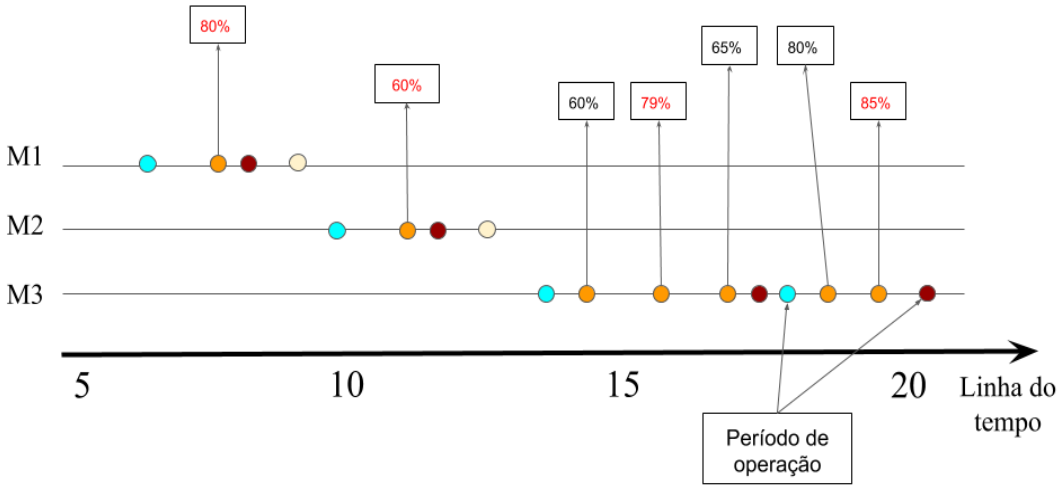
5.4 Algoritmo de Busca do melhor Limiar de Probabilidade (ABLP)

O principal motivo para a inserção de um limiar é limitar a geração de um alerta de desligamento fazendo com que o sistema informe somente quando os dados mostrarem que a sequência de falhas ocorrida possui potencial para gerar um desligamento. A escolha foi feita a partir da média das maiores probabilidades relacionadas à ocorrência do desligamento encontradas no histórico de alarmes.

Após encontrar as sequências formadas por alarmes AID e persisti-las na base de dados, ainda na fase de treinamento, também é executado o Algoritmo de Busca do melhor Limiar de Probabilidade (ABLP). Este limiar será uma probabilidade, usada como valor limite para indicar se haverá ou não desligamento abrupto. Quando uma nova sequência de alarmes ocorrer, o sistema buscará episódios entre os padrões salvos na base de dados, e irá calcular a probabilidade dessa sequência ocorrer. Caso a probabilidade seja maior que o limiar, então o sistema informará que vai ocorrer um desligamento abrupto, caso contrário o sistema informa que não haverá desligamento.

Para cada período de operação separado no conjunto de testes, haverá uma máxima probabilidade de acontecer um desligamento. Para o próximo período, será calculada novamente a probabilidade máxima de haver um desligamento e os valores são guardados e, assim, sucessivamente. A probabilidade máxima é calculada entre a probabilidade de o alarme atual levar ao desligamento e a probabilidade da sequência toda, incluindo o alarme atual, levar a um desligamento abrupto. Ao final, a média de todas as probabilidades máximas é calculada e este valor representa o limiar. Para este estudo, o limiar obtido foi de 50%. Na Figura 22, é ilustrado um exemplo em que as maiores probabilidades dos períodos (valores destacados em vermelho) são somadas para calcular a média. Neste caso, o limiar L é dado por L = (80+60+79+85) / 4 = 76. Assim, para uma sequência, se a probabilidade de ocorrer desligamento for superior a 76%, esta será considerada causadora do desligamento abrupto.

Figura 22. Cálculo do limiar de ativação.



Fonte: Autoria Própria.

5.5 Discussão

Neste capítulo, foi apresentado um método para predição de desligamentos utilizando um modelo de predição, que combina dois componentes que utilizam técnicas de identificação de padrões e máquinas de estados.

Essa combinação busca realizar a predição de forma adequada aos propósitos da pesquisa, para que o prognóstico possa ser realizado em tempo real e de forma acurada. O modelo final tem característica híbrida por combinar técnicas diferentes para atingir o propósito da pesquisa. Nesta abordagem, vale ressaltar, que a etapa de treinamento do modelo de predição ocorre de forma independente da máquina de estados. A máquina de estados quando treinada, gera novas sequências de alarmes baseada nos novos alarmes recebidos.

Na fase de predição/validação é aplicada uma lógica de decisão para gerar um alerta de desligamento. O alerta só será gerado, assim como a predição do modelo de predição, se a probabilidade de ocorrer um desligamento, a partir de uma determinada sequência de alarmes, for superior a um limiar definido a partir de todas as ocorrências encontradas no histórico.

No Capítulo 6, serão apresentados e discutidos os resultados obtidos com o uso da abordagem proposta.

Capítulo 6

Apresentação e Análise dos Resultados

Para facilitar a apresentação e análise dos resultados obtidos, a parte mais significativa desses é apresentada neste capítulo.

6.1 Critérios de Avaliação

A etapa de avaliação consistiu em verificar o desempenho do modelo de predição. Um conjunto de dados foi selecionado de dias de operação na UTE para serem inseridos no sistema e avaliados os resultados segundo os seguintes critérios:

- Se o modelo de predição afirmar que apenas um alarme pode causar um desligamento, este alarme deve, tecnicamente, ser capaz de gerar um desligamento;
- O resultado de cada período de operação na fase de testes é avaliado alarme por alarme; e
- O resultado obtido pelo modelo de predição, para os períodos de operação, leva em consideração o tempo em que o alarme ocorreu. Se este ocorreu e depois foi reconhecido ou voltou ao normal, não será contabilizado.

O conjunto de dados possui 1.424.789 registros e o resumo dos dados está descrito na Tabela 5. Dentre estes registros, existem 325 tipos de alarmes, 564 períodos de operação e 952 alarmes de desligamentos abruptos e 1.899 desligamentos normais.

As métricas foram calculadas em função da contagem de episódios. Ao obter estes valores, foram calculadas as probabilidades, a predição e o cálculo dos tempos médios até os desligamentos.

O modelo foi treinado, testado e as métricas de avaliação foram calculadas usando o conjunto de bibliotecas do Scikit-learn (PEDREGOSA, 2011).

Tabela 5. Resumo dos dados.

Desligamentos Abruptos	952
Desligamentos Normais	1.899
Períodos de Operação Válidos	564
Períodos com Desligamento Abrupto	57
Períodos com Desligamento Normal	507
Total de eventos no Log	1.424.789

Fonte: Autoria Própria.

6.2 Avaliação do Modelo de Predição

A avaliação objetivou testar se o modelo de predição é capaz de, dentro de um conjunto de sequências de alarmes AID, prever quando haverá e quando não haverá desligamento abrupto.

Para considerar que uma determinada sequência de alarmes levará a um desligamento, a probabilidade gerada pelo modelo deve ser maior que o limiar escolhido e, além disto, deve ser maior também que a probabilidade de não acontecer desligamento. Desta forma, o modelo garante que o alarme ocorre com mais frequência antes do desligamento abrupto, dando a certeza de que há relação com o desligamento. Na comparação com o limiar, o modelo escolhe a maior das duas probabilidades: a probabilidade da sequência inteira ou somente a probabilidade do último alarme da sequência levar ao desligamento.

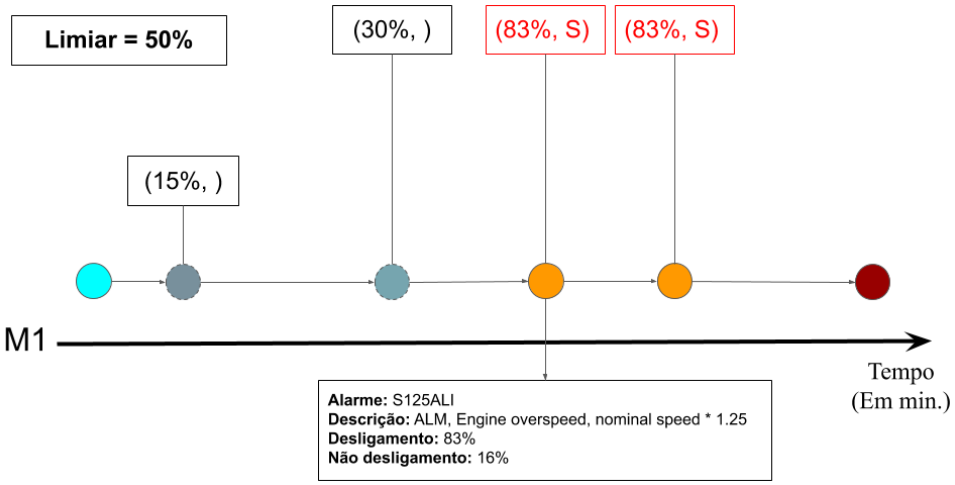
Ao gerar uma indicação de desligamento, o sistema manterá o resultado diante dos próximos alarmes que ocorrerão a não ser que o estado do alarme mude para qualquer um diferente de ACK_ALM. Neste caso, o modelo não o considerará mais como ameaça. Desta forma, os resultados apresentados na Tabela 6 levam em consideração o desempenho do modelo ao longo do tempo.

Na Figura 23, é apresentado um dos resultados obtidos pelo modelo ao realizar uma predição. Neste caso, um alarme (círculos de cor laranja), individualmente, foi indicado como uma ameaça para geração de um

desligamento. Enquanto o alarme não sair do estado ACK_ALM, o sistema continua o indicando como ameaça. Outros alarmes ocorreram (círculos de cor azul escuro), mas nem a sequência na qual estão inseridos e nem individualmente, obtiveram uma probabilidade superior ao limiar de predição para levar ao desligamento abrupto. O alarme S125ALI obteve uma probabilidade de 83% de levar a um desligamento enquanto que o de não levar foi de apenas 16%. A chance de levar ao desligamento também é maior que o limiar que é de 50%. Com isto, o modelo decidiu que o alarme pode levar ao desligamento.

Analisando-se novamente a Figura 23, ao longo do tempo, verifica-se que nenhuma ação foi tomada e o alarme com probabilidade alta (cor laranja) não saiu do estado ACK_ALM. Logo, permaneceu ativa a possibilidade desse alarme levar a um desligamento abrupto (porcentagens destacadas em vermelho). Neste caso, o modelo acertou em ter indicado desligamento, pois este período de operação terminou com um desligamento.

Figura 23. Emissão de alerta de desligamento baseado na probabilidade de a sequência de alarmes levar ao desligamento.



Fonte: Autoria Própria.

Os resultados da fase de testes foram analisados conforme as restrições listadas na Seção 6.1. Os conjuntos de treino e teste para a validação cruzada foram separados pelo ASTT. Dos 593 períodos de operação, em 66 desses ocorreram desligamentos abruptos e nos restantes 523, não houve desligamento. Dos 66 casos com desligamentos abruptos, 9 apresentaram

problemas e foram desconsiderados da fase de treino e teste. Da mesma forma, dos 523 períodos sem desligamentos, 16 apresentaram problemas e também não foram considerados nas fases de treino e teste. As fases de treino e testes foram executadas considerando a técnica de validação cruzada com o objetivo de proporcionar melhor avaliação do desempenho do modelo. Com isto, o modelo foi treinado e testado sob diferentes perspectivas do conjunto de dados. A técnica de validação cruzada consistiu em dividir o conjunto em *n* subconjuntos e treinar o modelo com 90% destes subconjuntos e testá-lo com os 10% restantes. Este procedimento, baseado na seleção de parte dos subconjuntos para treino e os subconjuntos restantes para teste, foi replicado 10 vezes, de forma que cada um dos subconjuntos faça parte do teste.

A validação cruzada foi executada quatro vezes com valores diferentes para o Limiar de Probabilidade. Os valores selecionados foram 0,2, 0,5, 0,7 e 0,9, com o intuito de avaliar o desempenho do modelo em um cenário com muita (0,9) ou pouca (0,2) restrição quanto à chance de os alarmes levarem ao desligamento. Quanto mais próximo de 0,9, menos alertas de desligamento abrupto serão emitidos e quanto mais próximo de 0, mais alertas ocorrerão.

Ao final, foram obtidas as matrizes de confusão construídas a partir dos resultados das fases de testes de cada etapa da validação cruzada. Os números são a soma dos erros e acertos obtidos. O objetivo principal das tabelas é analisar o desempenho do modelo. Os resultados são mostrados na Tabela 6, os quais não se alteram para um limiar abaixo de 0,5. O número de Falsos Positivos é inversamente proporcional ao limiar. Em contrapartida, o número de Falsos Negativos é diretamente proporcional ao limiar.

Tabela 6. Matrizes de confusão geradas a partir do Modelo de Predição para cada um dos limiares escolhidos.

	0,2		**0,5**		**0,7**		**0,9**	
	Sim	**Não**	**Sim**	**Não**	**Sim**	**Não**	**Sim**	**Não**
Sim	390	117	390	117	436	71	449	58
Não	9	48	9	48	34	23	44	13

Fonte: Autoria Própria.

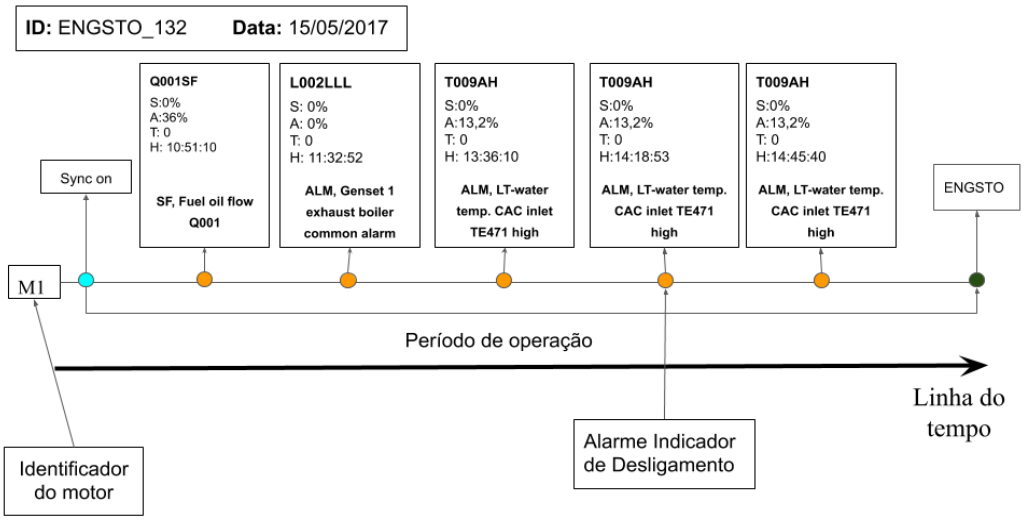
Para este estudo, é mais importante enfatizar o número de Verdadeiros Negativos, pois se o modelo indicar não desligamento e houver um desligamento, tal fato é mais prejudicial para a UTE do que a indicação de que haverá desligamento e na prática, este não ocorrer. Assim, se ocorrer um desligamento, mesmo com o modelo informando que não, os prejuízos para a central de controle serão maiores, pois os operadores não se mobilizarão para amenizar as consequências do desligamento e nem irão se preparar para sanar o problema antes que esse aconteça. Por este motivo, o limiar de 0,5 está mais adequado por apresentar uma taxa de acertos maior para Verdadeiros Negativos.

As sequências apresentadas na Figura 24 e na Figura 25 foram obtidas da validação cruzada, em que o modelo avaliou alarme por alarme e gerou um alerta caso a probabilidade de um desligamento fosse maior que o limiar e se o alarme estivesse no estado ACK_ALM. Se, ao longo do tempo, este alarme sai do estado ACK_ALM e passa para outro, então o alerta gerado pelo modelo é desativado e não é contabilizada a contagem na matriz de confusão. Na Figura 24 é ilustrada a ocorrência de não desligamento e o sistema não informou nenhuma vez que haveria desligamento. Isto foi correto, já que não houve em nenhum momento um alarme com probabilidade alta de levar a um desligamento.

Na Figura 25, é apresentado outro período. Em um determinado momento, ocorreu um alarme que apresentou uma probabilidade considerável de levar a um desligamento abrupto (acima de 50%) e o sistema avaliou como positiva a possibilidade do desligamento ocorrer. Por isto, o modelo alertou o operador sobre a possibilidade de ocorrer um desligamento.

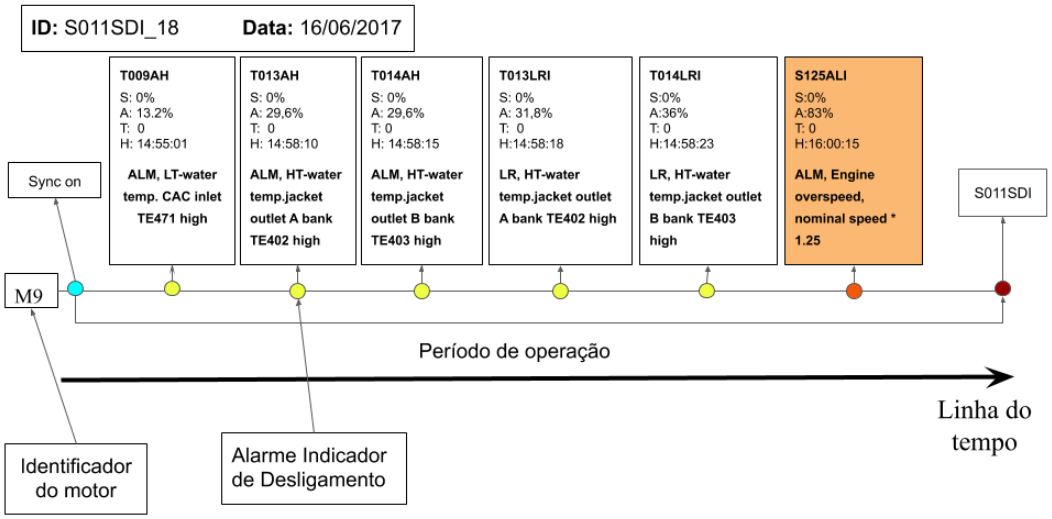
Como são 593 períodos de operação, com muitos alarmes ocorrendo em cada período, torna-se inviável apresentar todos os casos neste documento. Porém, estes são dois exemplos de como o ADP identifica os padrões dentro dos períodos de operação.

Figura 24. Período de não desligamento retirado do conjunto de testes avaliado pelo modelo de predição.



Fonte: Autoria Própria.

Figura 25. Período de desligamento retirado do conjunto de testes avaliado pelo modelo de predição.



Fonte: Autoria Própria.

A partir das matrizes de confusão da Tabela 6, as métricas de desempenho do modelo para cada valor de limiar escolhido, estão apresentadas na

Tabela 7. Todas estas métricas estão definidas no Apêndice B que trata de detalhes sobre o projeto de pesquisa executado neste estudo.

Tabela 7. Métricas de desempenho do modelo de predição.

Métrica de desempenho	0,2	0,5	0,7	0,9
Acurácia Total	78%	78%	81%	81%
Precisão	77%	77%	86%	88%
Recall	98%	98%	93%	91%
F-Measure	86%	86%	89%	89%

Fonte: Autoria Própria.

Conforme

Tabela 7, o desempenho do modelo, refletido nas métricas, aumenta quando o limiar aumenta. Isto porque o modelo passa a acertar mais quando não haverá desligamento abrupto e passa a errar mais quando haverá desligamento. Estes limiares não são apropriados para o modelo, porque o mais importante é que sejam realizadas predições corretas sobre quando haverá desligamento abrupto e não quando não haverá. Um limiar abaixo de 0,5 se mostra mais adequado, devido à importância dos Verdadeiros Negativos para a UTE, já que a central se mobilizará quando o sistema informar que haverá desligamento. Para um limiar de 0,5, o número de falsos negativos é mínimo, o que significa que este valor minimiza o erro de predição para os casos em que haverá desligamento.

6.3 Discussão

A hipótese de pesquisa H0 foi rejeitada, dado que houve indícios de que os alarmes possuem um relacionamento com os desligamentos, conforme ilustrado no exemplo do alarme B7324SD e dos limites de operação de segurança apresentados na Seção 3.2. Este fato técnico indica que há uma relação entre os alarmes e os desligamentos abruptos. Há também a

possibilidade de prever os desligamentos, dado que, para o operador, o modelo irá sugerir a ocorrência ou não do desligamento com antecedência, assim que ocorrer um alarme. O desempenho do modelo é promissor quando se trata de prever os desligamentos abruptos. Além disto, o operador também terá acesso a informações como o tempo médio até o desligamento e a probabilidade deste alarme levar a um desligamento. Estas informações serão dadas com antecedência, proporcionando aos operadores tomar decisões com mais antecedência do que no cenário atual de operação na UTE. Destaca-se, também, que o modelo aprende com novas instâncias, o que proporcionará o aperfeiçoamento do modelo e melhoria dos resultados à medida que a base de dados aumenta com novos registros.

Estes indícios, analisando-se de forma qualitativa, indicam que o modelo é útil no que se propõe a fazer, gerar informações de prognóstico para avaliar, diagnosticar e corrigir falhas. O modelo poderá ser melhorado quando forem adicionados mais dados e seu uso foi positivo em uma UTE, dado seu bom desempenho. Posteriormente, o modelo poderá ser inserido em um sistema WEB para auxiliar a central de controle no processo de FDDC.

No Capítulo 7, serão apresentadas e discutidas as Considerações Finais e Propostas para Pesquisas Futuras.

Capítulo 7

Considerações Finais e Sugestões para Pesquisas Futuras

Neste capítulo, serão apresentadas as considerações finais sobre a pesquisa, as limitações e os pontos principais a serem continuados.

7.1 Considerações Finais

Nesta pesquisa, foi proposta uma abordagem para o prognóstico de situações de falhas (desligamentos) geradas nos motores de uma UTE, baseada no uso de técnicas de aprendizagem de máquina e máquinas de estados, aplicada em uma UTE. A avaliação das situações de desligamento, indicando se esta situação ocorrerá ou não, foi baseada na extração de características correspondentes à frequência dos alarmes em períodos de operação, seguida de sua classificação, e, posteriormente, da identificação e contagem de episódios no histórico de alarmes e do tempo médio até o desligamento. Esta abordagem pode ser considerada, portanto, híbrida, pois une técnicas de aprendizagem de máquina com construção de máquinas de estados.

Neste estudo, foram selecionadas diversas combinações de sequências de alarmes e técnicas de classificação, com o intuito de avaliar o desempenho desta combinação na predição de desligamentos abruptos.

Foram utilizadas abordagens frequentistas ao serem extraídas características dos períodos de operação, geradas a partir do histórico de alarmes (contagem de episódios e tempo médio entre as ocorrências). O método de classificação selecionado foi definido pelo seguinte objetivo: avaliar as sequências de alarmes em busca de padrões que indicam uma ocorrência de desligamento.

O método de classificação proposto na pesquisa é um modelo preditor com uma máquina de estados que agrega a relação entre os alarmes e os desligamentos abruptos.

A partir de um subconjunto da base de dados do sistema de controle, formada por alarmes extraídos de uma UTE, gerados automaticamente por motores e com situações de falhas de operação e de testes dos motores, previamente classificados em duas classes: “Sim” e “Não”, tornou-se possível a análise, em um ambiente real, com um conjunto de dados representativo desse ambiente. Além de falhas de operação, outras geradas a partir de testes realizados nos motores foram usadas, devido à dificuldade para conseguir dados que contenham apenas períodos de operação dos motores, considerando a condição de operação da UTE.

A partir de um conjunto significativo de eventos de alarmes (1.424.789), obtidos do sistema de controle, observou-se a eficácia da abordagem diante das situações avaliadas, pois esses permitiram a avaliação das falhas entre as duas classes avaliadas (“Ocorrerá desligamento” e “Não ocorrerá desligamento”).

As características extraídas dos períodos de operação permitiram caracterizar melhor as sequências de alarmes quanto a sua relação com o desligamento, identificando a frequência com que a sequência ocorreu e o tempo médio até o desligamento.

Ao final da pesquisa, considerando-se os estudos realizados e os resultados obtidos, conclui-se que o prognóstico do desligamento dos motores pode ser realizado, de forma eficaz, considerando o histórico de alarmes. O prognóstico pode se mostrar, portanto, como um auxílio à tomada de decisão sobre o estado de funcionamento dos motores da UTE e sobre a prevenção de desligamentos abruptos. Além disto, com os resultados obtidos, há indícios para rejeitar a hipótese H0, logo é possível prever e tratar os desligamentos com antecedência.

A principal contribuição da modelagem proposta visa a colaborar com a diminuição dos desligamentos abruptos em uma unidade geradora de energia, favorecendo ações preventivas minimizando, assim, os custos provenientes da retirada dos motores de operação e impactando positivamente no processo de geração de energia em usinas termoelétricas.

7.2 Contribuições da Pesquisa

Ao final da pesquisa, podem ser elencadas as contribuições destacadas a seguir.

- O desligamento pode ser identificado com antecedência, o que permite ao operador do sistema agir antecipadamente para resolver problemas causados pela falha.

- O uso da Técnica 1, para extrair padrões, da Técnica 2, para analisar padrões e da Técnica 3, para gerar métricas de prognóstico e predizer se haverá o desligamento abrupto, se mostraram eficazes para gerar predições.

- A utilização de um sistema de predição mostrou ser eficaz e útil no processo de FDDC da UTE, com desempenho significativo.

- O uso do histórico de alarmes demonstrou ser eficaz no processo de predição de desligamento.

O modelo de predição se mostrou, portanto, eficaz para a predição de desligamentos apresentando resultados de desempenho significativos.

7.3 Sugestões para Pesquisas Futuras

Ao final da pesquisa apresentada, várias questões podem ser identificadas para pesquisas futuras, com destaque para as seguintes.

- O modelo de predição não consegue prever bem para instâncias de alarmes que nunca aconteceram e isto pode causar a geração de

informações erradas. Este problema pode ser estudado em pesquisas futuras, com a adição, por exemplo, de informações heurísticas sobre os relacionamentos das falhas, para amenizar seus impactos.

- O modelo de predição pode ser melhorado, se ao invés de usar somente os alarmes, considerar também desvios nos valores dos gráficos de outras variáveis de processo. Modificações bruscas no comportamento das variáveis podem indicar problemas no motor envolvendo as grandezas físicas monitoradas pela variável.

- Os períodos de operação podem ser definidos de forma mais precisa, pois o sistema de controle é capaz de identificar os horários em que os motores começaram a operar e, além disto, emite um alarme naquele horário, com o valor produzido correspondente. Ao realizar um cruzamento entre os dados dos horários de produção e o conjunto de alarmes, é possível descobrir de forma mais precisa o início da operação, excluindo-se o período de inicialização calculado.

Referências Bibliográficas

ALAEDDINI, A; DOGAN, I. **Using Bayesian networks for root cause analysis in statistical process control**. Expert Systems with Applications, 38(9), pp. 11230–11243, 2011.

BOGHEY, R.; SINGH, S. **Sequential Pattern Mining: A Survey on Approaches**. In International Conference on Communication Systems and Network Technologies (CSNT), pp. 670–674, 2013.

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. **Classification and Regression Trees,** no, Ed. Belmont, CA: Wadsworth International Group. Machine Learning, 1984.

BREIMAN, L. Random forest. **Machine Learning**, v. 45, n. 5, pp. 1–35, 1999.

BÜHLMANN, P.; YU, B. **Boosting**. Wiley Interdisciplinary Reviews: Computational Statistics. 2010.

CHIANG, L.; RUSSEL, E.; BRAATZ, R. **Fault detection and diagnosis in industrial systems**. Springer-Verlag, London. 2001.

COOK, J. A.; RANSTAM, J. **Overfitting**. British Journal of Surgery, vol. 103, n. 13, pp. 1814–1814, 2016.

DAHLSTRAND, F. **Consequence analysis theory for alarm analysis**. Knowledge-Based Syst., vol. 15, pp. 27-36, 2002.

DAVIS, J; GOADRICH, M. **The relationship between Precision-*Recall* and ROC curves**. 2006.

DIMITRIADIS, S. I; LIPARAS, D. **How random is the random forest? Random forest algorithm on the service of structural imaging biomarkers for Alzheimer's disease: From Alzheimer's disease neuroimaging initiative (ADNI) database**. Neural Regeneration Research, 13(6), 2018.

DORGO, G.; ABONYI, J. **Sequence mining based alarm suppression**, IEEE Access, vol. 6, pp. 15365–15379, 2018. Disponível: https://ieeexplore.ieee.org/document/8268070/

ESLING, P.; AGON, C. **Time-series data mining**, ACM Comput. Surveys, ed. 1, vol. 45, 2012.

FLACH, P.; KULL, M. **Precision-*Recall*-Gain Curves: PR Analysis Done Right**. Proc. Int. Conf. Advances in Neural Information Processing, pp. 838-846, 2015.

FOLMER, J.; SCHURICHT, F.; VOGEL-HEUSER, B. **Detection of Temporal Dependencies in Alarm Time Series of Industrial Plants**. IFAC Proceedings Volumes, vol. 47, n. 3, pp. 1802-1807, 2014. Disponível: http://linkinghub.elsevier.com/retrieve/pii/S1474667016418744. Último acesso em: 03 de fevereiro de 2020.

FOLMER, J.; VOGEL-HEUSER, B. **Computing dependent industrial alarms for alarm flood reduction**. 9th IEEE int multi-conference syst. Signals Devices, pp. 1–6, 2012.

FOURNIER-VIGER, P. et al. A survey of sequential pattern mining. **Data Science and Pattern Recognition**, vol. 1, n. 1, pp. 54–77, 2017.

FRANK, P. M. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy—A survey and some new results, **Automatica**, vol. 26, n. 3, pp. 459 - 474, 1990.

FREUND, Y.; SCHAPIRE, R. (1996). Proceedings of the Thirteenth International, **Experiments with a new boosting algorithm,** Conference on Machine Learning, pp. 148 -156. Morgan Kaufmann, San Francisco.

HALL, M. et al., **The WEKA data mining software: An update**, SIGKDD Explorations, vol. 11, n. 1, pp. 10-18, 2009.

HU, W.; CHEN, T.; SHAH, S. L. **Detection of frequent alarm patterns in industrial alarm floods using itemset mining methods**, IEEE Trans. Ind. Electron., vol. 65, n. 9, pp. 7290-7300, 2018.

HU, W.; Wang, J.; Chen, T. A local alignment approach to similarity analysis of industrial alarm flood sequences. **Control Engineering Practice**, vol. 55, pp. 13–25, 2016.

ISERMANN, R. **Supervision, Fault-Detection and Fault-Diagnosis Methods - An Introduction**. Control Engineering Practice, vol. 5, n. 5, pp. 639-652, 1997.

IZADI, I. et al. **A framework for optimal design of alarm systems**, Proceedings of 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, pp. 651-656, 2009.

IZADI, I. et al. **An introduction to alarm analysis and design**, Proceedings of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, Barcelona, Espanha, pp. 645–650, 2009.

IZADI, I. et al. **Optimal alarm design**. Proceedings of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, pp. 651–656, 2009.

JULISCH, K. **Clustering intrusion detection alarms to support root cause analysis**. ACM Transactions on Information and System Security(TISSEC), 6(4), 443–471, 2003.

KOHAVI, R. **A study of cross-validation and bootstrap for accuracy estimation and model selection**. Proceedings of the 14th international joint conference on Artificial intelligence, vol. 2, 1995.

KONDAVEETI, S. R. **Advanced Analysis and Redesign of Industrial Alarm Systems**. PhD tese, University of Alberta, 2013.

LANGONE R. et al. **Alarm prediction in industrial machines using autoregressive LS-SVM models**, Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining (CIDM), IEEE, pp. 359–364, 2014.

LEEMANS, M.; VAN DER AALST, W.M.P. **Discovery of frequent episodes in event logs**. Data-Driven Process Discovery and Analysis, Springer, pp. 1-31, 2014.

LEVITT, J. **Complete Guide to Predictive Maintenace**, ISBN-10: 0831134410, ed. 2, 2011.

LI, Y. et al. **Discovering calendar-based temporal association rules**. Proceedings of the 8th International Symposium on Temporal Representation and Reasoning, pp. 111–118, 2001.

MAKI, Y.; LOPARO, K. A. **A neural-network approach to fault detection and diagnosis in industrial processes**. IEEE Transactions on Control Systems Technology, vol. 5, pp. 529–541, 1997.

MANNILA, H.; TOIVONEN, H.; VERKAMO, A. I. **Discovery of frequent episodes in event sequences**. Data Mining and Knowledge Discovery, vol. 1, pp. 259–289, 1997.

NOYES, J. **Alarm systems: a guide to design, management and procurement**. Eng. Manag. J, vol. 191, ed. 3, pp. 226–226, 2002.

PARIYANI, A. et al. **Dynamic Risk Analysis Using Alarm Databases to Improve Process Safety and Product Quality: Part I – Data Compaction**. American Institute of Chemical Engineers, AIChE Journal, 58(3), pp. 812–825, 2011.

PARIYANI, A. et al. **Dynamic risk analysis using alarm databases to improve process safety and product quality: Part II - Bayesian analysis**. AIChE J, 58 (3). Disponível em: http://dx.doi.org/10.1002/ aic. 2012.

PEDREGOSA, F. et al. **Scikit-learn: Machine learning in Python**. Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.

POWERS, D. M. W. e AILAB. **EVALUATION: FROM PRECISION, *RECALL* AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS & CORRELATION**. Journal of Machine Learning Technologies ISSN, vol. 2, pp. 37- 63, 2011.

ROTHENBERG, D. **Alarm Management for Process Control**, A Best-Practice Guide for Design, Implementation, and Use of Industrial Alarm Systems, Momentum Press, 2009.

SARTORI, I. et al. **Detecção, Diagnóstico E Correção De Falhas: Uma Proposição Consistente De Definições E Terminologias**. Science & Engineering Journal, vol. 21, n. 2, pp. 1983–4071, 2012.

SASAKI, Y. et al. The truth of the F-measure. **Teach Tutor mater**, vol. 1, n. 5, pp. 1-5, 2007.

SCHLUTER, T.; CONRAD, S. **About the analysis of time series with temporal association rule mining**. IEEE Symposium on Computational Intelligence and Data Mining, pp. 325–332, 2011.

SOARES, V. B. **Detecção e diagnóstico de falhas em plantas industriais com base em padrões de alarmes.** 184f. Tese (Doutorado em Engenharia Química) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2016.

THUAN, N. D.; TOAN, N. G.; TUAN, N. L. V. **An Approach Mining Cyclic Association Rules in E-Commerce**. 15th International Conference on Network-Based Information Systems, New York, vol. 12, pp. 408–411, 2012.

SAMMUT, C.; WEBB, G.I. Confusion Matrix. **Encyclopedia of Machine Learning and Data Mining**. 2. ed. New York: Springer US, 2017, p. 209.

VENKATASUBRAMANIAN, V; RENGASWAMY, R.; KAVURI, S. N. **A review of process fault detection and diagnosis Part II: Qualitative model and search strategies**. Computers and Chemical Engineering, vol. 27, pp. 313-326, 2003b.

VENKATASUBRAMANIAN, V. et al. **A review of process fault detection and diagnosis Part I: Quantitative model-based methods**. Computers and Chemical Engineering, vol. 27, pp. 293-311, 2003a.

VENKATASUBRAMANIAN, V. et al. **A review of process fault detection and diagnosis Part III: Process history based methods**. Computers and Chemical Engineering, vol. 27, pp. 327-346, 2003c.

VOGEL-HEUSER, B; SCHULTZ, D; FOLMER, J. Criteria-based alarm flood pattern recognition using historical data from automated production systems (aPS). **Mechatronics**, vol. 31, pp. 89-100, 2015. Disponível em: http://dx.doi.org/10.1016/j.mechatronics.2015.02.004. Último acesso em: 15 de agosto de 2019.

WAGNER, F. et al. **Modeling Software with Finite State Machines: A Practical Approach**. Number 0-8493-8086-3. Taylor & Francis Group, LLC, ed. 1, 2006.

WANG, J. et al. **A data similarity based analysis to consequential alarms of industrial processes**. Journal of Loss Prevention in the Process Industries, vol. 35, pp. 29-34, 2015.

WILLSKY, A. S. **Survey of design methods for failure detection in dynamic systems**. Automatic, vol. 12, pp. 601-611, 1976.

YANG, F.; SHAH, S. L; D. Xiao, **Correlation analysis of alarm data and alarm limit design for industrial processes**, Proc. Amer. Control Conf., Baltimore, MD, USA, 2010, pp. 5850–5855.

ZHAO, Q.; BHOWMICK, S.S. **Sequential pattern mining: A survey**. ITechnical Report, Nayang Techno- logical University Singapore, pp. 1-26, 2003.

ZHU, J. et al. Multi-class Adaboost. **Statistics and interface**, vol. 2, pp. 349-360, 2009.

ZHU, J. et al. **Dynamic alarm prediction for critical alarms using a probabilistic model**. Chinese Journal of Chemical Engineering, 24(7), pp. 881-885, 2016. Disponível em: http://www.sciencedirect.com/science/article/pii/S1004954116303044. Último acesso em: 15 de agosto de 2019.

Apêndice A

Análise de Dados e Construção dos Modelos de Aprendizagem Preliminares

Neste capítulo, é apresentada a análise estatística realizada nos dados de alarmes e também é apresentada uma abordagem inicial para o modelo de aprendizagem, usada antes da abordagem híbrida final.

Na análise estatística considerou-se apenas o histórico de alarmes, o que é suficiente para avaliar, inicialmente, a consistência e formato dos dados. Os outros eventos não foram descartados, mas apenas separados para fins específicos, como delimitar os períodos de operação. O estudo foi realizado conforme os passos a seguir.

1. Análise descritiva dos dados
    a. Contagem da quantidade de alarmes no histórico, dos tipos diferentes de alarmes, desligamentos, linhas repetidas, dentre outros.
    b. Cálculo das métricas, como a média, mediana, desvio padrão e IQR.
        i. Considerando *outliers*; e
        ii. Desconsiderando *outlier*s.
    c. Obtenção da frequência com que cada tipo de alarme ocorre desconsiderando o equipamento no qual o alarme ocorreu;
        i. Repetidas vezes;
        ii. Antes dos outros alarmes.
2. Extração, transformação e carga das características a partir dos dados brutos de registros de alarmes.
    a. Criação de novas características;
    b. Subdivisão das sequências de alarmes por equipamento, em blocos de tempo (turnos), segundo o funcionamento da usina.

3. Análise dos dados

a. Comparação entre métricas obtidas das distribuições e gráficos comparativos.

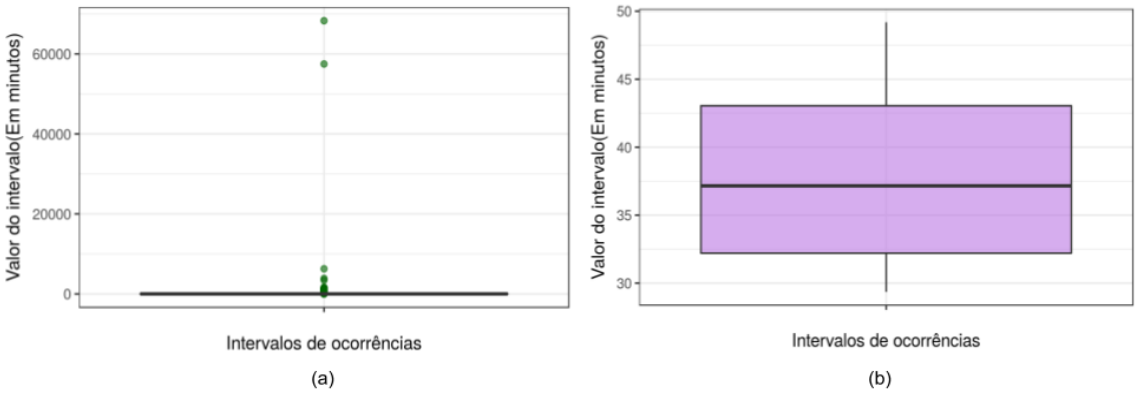
4. Obtenção e análise dos Resultados
5. Apresentação das Conclusões

1. Análise Descritiva dos Dados

Para esta etapa, por questões de tempo de processamento de dados e quantidade de recursos computacionais disponíveis, foi utilizada apenas uma amostra do conjunto total coletado, referente ao período de janeiro a março de 2017. Na época em que esta etapa foi executada, havia mais de 500 mil registros no subconjunto de alarmes, 325 tipos de alarmes diferentes, 31 ocorrências de desligamentos nos 129 turnos considerados em que a usina operou.

Como a dimensão tempo é importante, foi necessário considerar os intervalos entre cada ocorrência. Na Figura 26, são apresentados os gráficos de *box plot* com as distribuições, em minutos, dos intervalos de tempo entre alarmes adjacentes desconsiderando o equipamento no qual ocorreram. O gráfico (a) foi construído considerando os dias inoperantes em que a produção de energia na UTE foi próxima de zero e, por isso, um número considerável de pontos fora da curva (*outliers*) alterou a “forma” da distribuição. Por se tratar de intervalos de tempo referentes a dias inoperantes, em que os motores não estão ligados, porém outros equipamentos estão funcionando, então esses pontos não são representativos e foram removidos. O gráfico (b) exibido na Figura 26 considera a retirada dos *outliers* para melhor visualização dos valores dos intervalos de tempo.

Figura 26. Distribuição dos intervalos de ocorrência: Considerando todos os dias (a) e considerando apenas os dias em que a usina esteve em operação (b).



Fonte: Printscreen retirado da ferramenta R.

A função filtro utilizada para retirar os pontos fora da curva é descrita da seguinte forma: Seja Q1 e Q3, intervalos entre dois alarmes, em que Q1 é o primeiro e Q3 o terceiro quartil da distribuição. Então, a remoção dos intervalos de tempo considerados *outliers* foi feita conforme as Equações 1 e 2. O conjunto resultante da filtragem dos valores, é dado pela Equação 2.

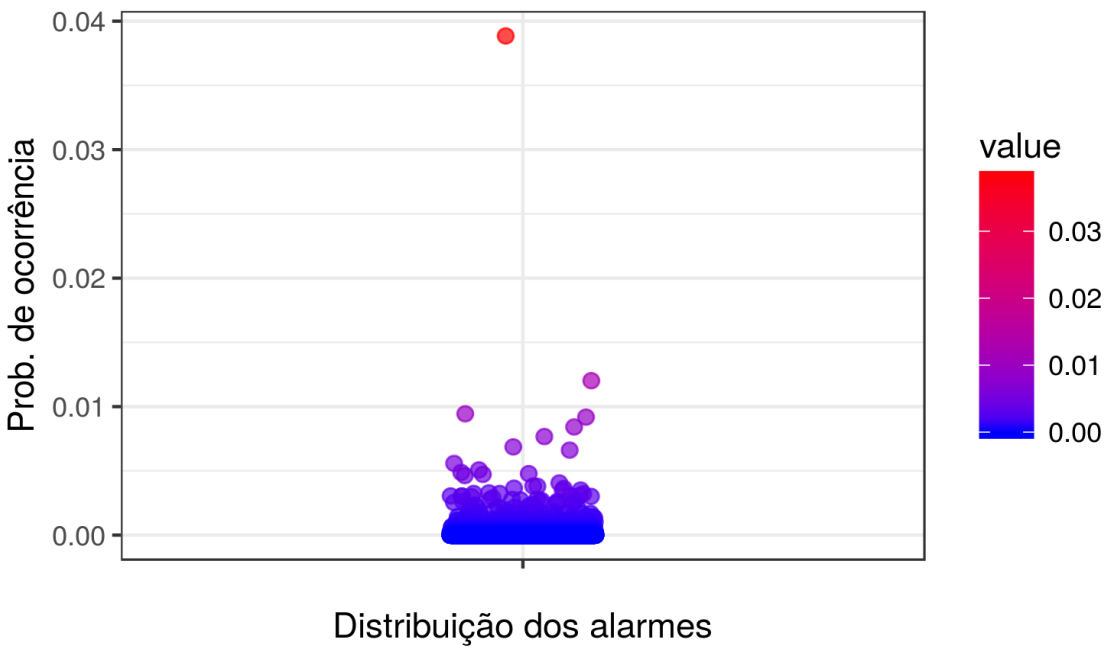
$$amplitude = (Q3 - Q1) \text{ x } 3. \quad (1)$$

$$X = (Q1 - amplitude) < valor < (Q3 + amplitude). \quad (2)$$

A mediana da distribuição corresponde a, aproximadamente, "37,15" minutos e o desvio padrão a, aproximadamente, “5,99”.

É importante observar a distribuição das frequências de cada alarme para observar quais são os alarmes que mais ocorrem na planta. Na Figura 27, são apresentadas as frequências com que cada alarme ocorreu. Pelo gráfico, é possível perceber que um alarme aconteceu mais que os outros e, embora não seja uma diferença significativa na probabilidade de ocorrência, este alarme foi emitido 1.444 vezes em apenas dois meses de operação. A partir de uma busca simples, foi possível descobrir que o alarme que mais ocorreu foi o alarme *T009AH*.

Figura 27. Distribuição das probabilidades de ocorrências de cada alarme.



Fonte: Printscreen retirado da ferramenta R.

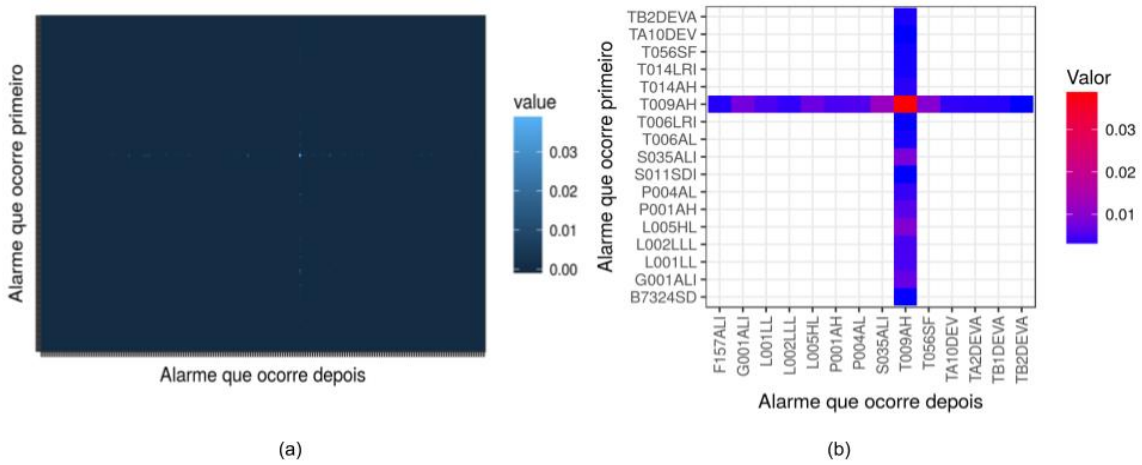
Quando uma falha ocorre, outra pode acontecer depois, dentro de um determinado intervalo de tempo. A falha que “ocorre antes” pode ser a causa da que “ocorre depois”.

Para identificar causalidade entre duas falhas, é necessário levar em consideração o instante de tempo em que ocorreram e o intervalo de tempo entre elas.

No gráfico (a) da Figura 28, é possível observar os resultados da contagem das ocorrências de cada alarme com todos os outros, selecionados dois a dois, inclusive com o próprio alarme. No eixo X, estão os alarmes que “ocorreram depois” e no eixo Y todos os alarmes que ocorreram antes, no período de tempo que vai do início do turno até o instante de tempo do alarme que “ocorreu depois”. No gráfico (a) da Figura 28, observa-se também que o caso que mais aconteceu foi o alarme *T009AH* (ponto vermelho no topo da distribuição). Esta descoberta sugere que a falha correspondente a este alarme é a que mais ocorre na UTE. Estes alarmes podem ser falsos positivos causados por defeitos ou descalibração dos sensores.

Os 17 casos com maior probabilidade de ocorrência apresentados na Figura 28 (a), são visualizados de forma destacada no gráfico (b) da Figura 28. Aparentemente, a quantidade de vezes em que o alarme *T009AH* ocorreu, foi alta, se comparado com os outros casos, porém analisando individualmente, uma probabilidade de aproximadamente 0,03 pode não ser um valor significativo que represente algum problema para a usina.

Figura 28. Contagem das ocorrências entre alarmes, selecionados dois a dois (a). Pares de ocorrências que mais aconteceram (b).



Fonte: Printscreen retirado da ferramenta R.

Com a análise descritiva, foi possível observar o formato da distribuição de intervalos de tempo entre os alarmes assim como descobrir quais alarmes mais ocorreram.

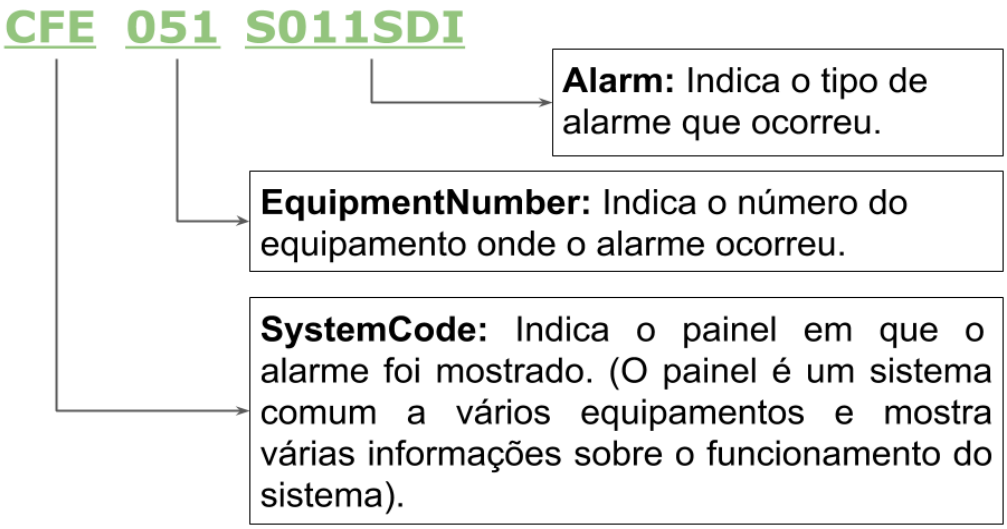
2. Extração, Transformação e Carregamento dos Dados

É importante destacar que, inicialmente, a abordagem para identificar os períodos para análise das relações entre os alarmes e os desligamentos foi diferente da escolhida para a abordagem final. Inicialmente, foi utilizada uma abordagem com turnos de operação e, no que diz respeito à correlação entre os alarmes não houve alteração e, portanto, os resultados aqui apresentados são válidos.

Para as etapas subsequentes, tornou-se necessário extrair novas informações dos dados de entrada, pois as informações geradas foram

insuficientes. A partir da base original, foi possível gerar novos atributos, como por exemplo, o número do equipamento. Na Figura 29, é apresentado o atributo *TagName,* que pode ser subdividido em várias partes, cada uma com um significado.

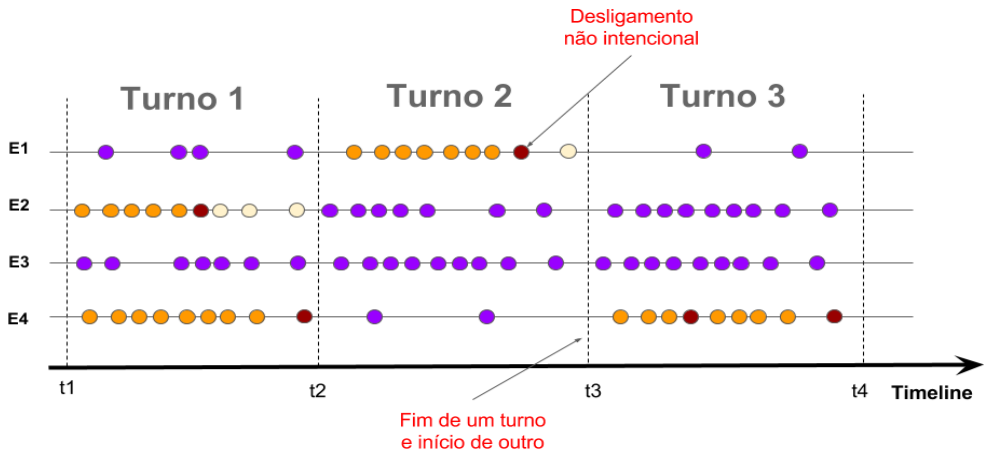
Figura 29. Identificação dos dígitos do campo TagName.



Fonte: Autoria própria.

O manual da planta contém informações mais detalhadas sobre cada dígito do campo *TagName*, porém, estas são as consideradas relevantes para a pesquisa. Além dessas características, foram extraídas outras duas, o número do turno correspondente e a data (sem hora) em que o alarme ocorreu. Com os novos atributos definidos, foi possível ordenar os registros pelo tempo, dividi-los por turno e agrupá-los por equipamento. Cada ponto apresentado na Figura 30 é um registro contido em uma parte do histórico de alarmes, correspondente a um período de três turnos.

Figura 30. Divisões da sequência de eventos.



Fonte: Autoria própria.

As informações sobre os momentos em que a planta estava em operação foram fornecidos pelos administradores da UTE. No processo de extração, os dados foram filtrados e transformados para as etapas subsequentes. Os registros incorretos e linhas duplicadas foram removidos, pois não apresentam nenhuma informação útil. Além disto, também foram considerados apenas os alarmes com estado UNACK_ALM, conforme mostrado na seção 3.2 do Capítulo 3, pois no momento, a ideia consiste em avaliar apenas os alarmes ativos do histórico. Os outros alarmes foram separados para serem utilizados em outras etapas, conforme evolução do estudo.

Um período de funcionamento é demarcado pelo instante de tempo de início do turno e pelo instante de desligamento do motor, que pode ser intencional ou não, conforme está destacado no gráfico da Figura 30. Avaliar os alarmes apenas em períodos de funcionamento restringe o escopo de análise e melhora os resultados obtidos, pois não é possível encontrar relação de causalidade fora do período de operação.

A separação em turnos permitiu a contagem da quantidade de ativações de cada tipo de alarme e se essa quantidade precedeu ou não um desligamento. Como exemplo, observando novamente a Figura 30, um período que pode ser considerado na contagem é o intervalo entre o início do turno 2 e o desligamento considerando o equipamento 1 (E1). No caso do equipamento 2, o período de análise consiste no início e fim da operação. No primeiro caso, relacionado ao equipamento E1, o valor da variável resposta será 1, pois houve

um desligamento, enquanto que no segundo caso, a variável resposta será 0, pois o equipamento foi desligado normalmente. O melhor caso seria realizar a análise exatamente no intervalo de tempo em que o motor foi iniciado e finalizado. Este intervalo corresponde ao período em que realmente a falha pode acontecer.

Para o caso do turno 3, em que ocorreram dois desligamentos, dois períodos são considerados do início do turno até o primeiro desligamento e do início do turno até o segundo desligamento. Então, a quantidade de ocorrências de cada tipo de alarme nestes dois períodos, considerando também o equipamento no qual ocorreram, foi contabilizada.

O último conjunto de dados extraído, necessário às próximas fases, foi a tabela de turnos ou períodos de operação. As células desta matriz são preenchidas por *flag*s que indicam a ocorrência do alarme no turno. Os dados foram separados em dois conjuntos: “turnos com desligamento” e “turnos sem desligamentos”. O algoritmo calculou, a partir de cada turno, para cada conjunto, os valores definidos na seção de métricas. As distribuições dos valores de cada métrica construídas na etapa de *Extração, transformação e processamento*, a partir de cada conjunto (“Turnos com desligamento” e “Turnos sem desligamento”), foram usadas para avaliar se o formato das sequências de alarme sofre modificações quando há desligamento.

A métrica do Total de Pontos foi usada para comparar a quantidade de alarmes que ocorre em cada situação, o IQR para medir o quão “espalhados” são os pontos na distribuição, o desvio padrão para comparar o quanto cada conjunto possui os valores afastados da média e a mediana para verificar o quanto os valores da distribuição estão próximos de algum valor representativo. Para comparar as distribuições de cada métrica, foram considerados os intervalos de confiança das medianas. A média não foi usada devido às alterações sofridas nesta métrica para valores muito altos devido ao período de tempo entre um alarme e outro.

3. Análise dos Dados

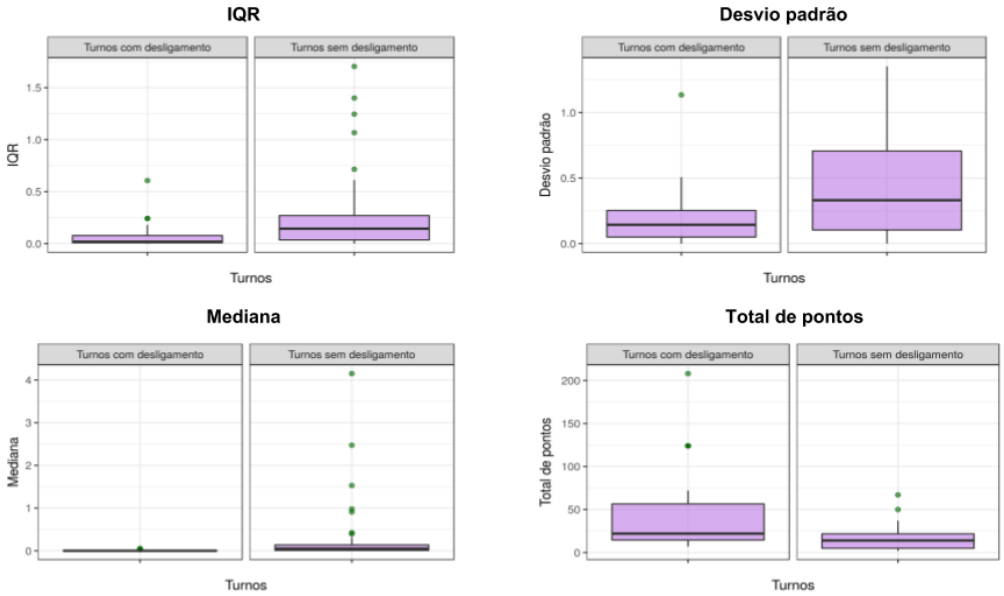
Após as transformações realizadas na base de dados, a fase de análise consistiu principalmente em obter uma análise descritiva dos dados e resultados que apresentassem indícios correlação entre os alarmes. Essa parte consistiu basicamente em comparar as métricas calculadas no conjunto “turnos com desligamento” com o conjunto “turnos sem desligamento” e, ao final, afirmar com significância de 95%, se há diferença ou não entre os conjuntos e, consequentemente, se os desligamentos alteram ou não a forma da distribuição de alarmes.

O teste estatístico usado foi o *bootstrap*, que é uma técnica baseada nas replicações aleatórias da amostra para obter uma nova distribuição e verificar se os valores das métricas não são aleatórios. O principal objetivo para usar o *bootstra*p é calcular os intervalos de confiança da mediana das distribuições geradas para cada métrica. O número de replicações selecionado para o *bootstrap* foi de 20.000, ou seja, ao fim da execução do *bootstrap*, para cada métrica foi gerada uma distribuição com vinte mil pontos.

Pode-se considerar que o *bootstrap* é executado com um nível de significância $\alpha = 0{,}05$, pois o intervalo de confiança encontrado é delimitado pelos valores que ocorrem em 95% dos casos. Os gráficos da Figura 31 mostram os resultados do *bootstrap* para cada métrica de cada um dos conjuntos: “Turnos com desligamento” e “Turnos sem desligamento”. Aparentemente existe uma diferença entre estes dois conjuntos considerados, pois os turnos em que ocorreu um desligamento apresentam um Total de pontos mais alto.

A partir do gráfico, observa-se que a mediana para os turnos com desligamento geralmente é menor, o que indica que os intervalos de ocorrências entre os alarmes são menores, o desvio padrão e IQR também são menores, o que indica uma baixa dispersão dos dados, ou seja, o tempo entre dois alarmes adjacentes é menor em turnos que ocorrem desligamentos. Isso caracteriza uma maior frequência de ativação de alarmes.

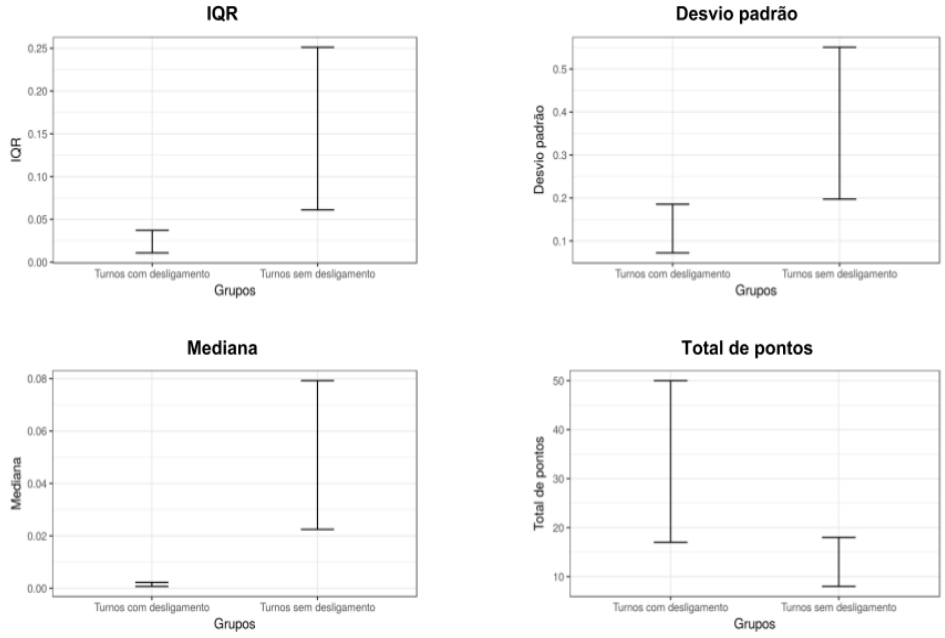
Figura 31. Comparação entre as distribuições das métricas geradas pelo algoritmo de Bootstrap.



Fonte: Printscreen retirado da ferramenta R.

Na Figura 32, são apresentados os intervalos de confiança para a média de cada uma das distribuições da Figura 31, para um nível de significância de 95%. A comparação entre os intervalos na Figura 32 sugere que os grupos são diferentes.

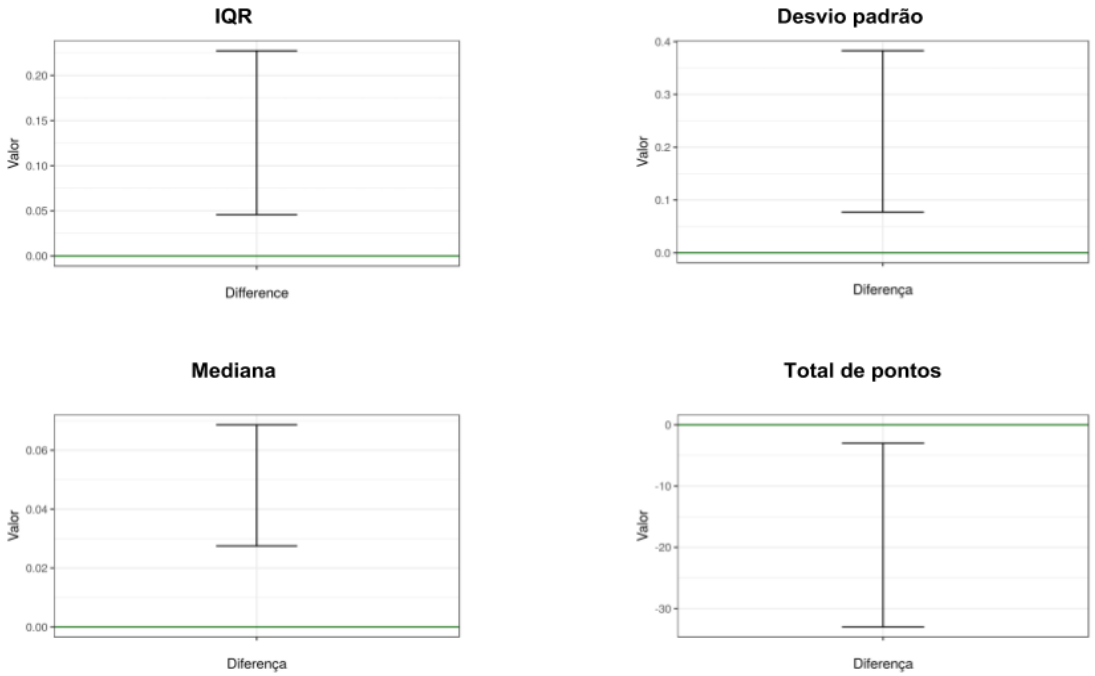
Figura 32. Comparação entre os intervalos de confiança de cada métrica.



Fonte: Printscreen retirado da ferramenta R.

Por último, foram comparados os intervalos de confiança das diferenças entre as métricas dos conjuntos, conforme apresentado na Figura 33. Nenhum dos intervalos contém o zero, o que indica que as métricas não são aproximadamente iguais. Os intervalos de confiança dão indícios de que os conjuntos são diferentes, ou seja, em turnos com desligamentos o formato da distribuição, a quantidade de alarmes emitidos e o tempo entre essas apresentam diferenças significativas, o que pode indicar que existem alarmes emitidos em um determinado intervalo de tempo antecedentemente, associados às falhas que desencadearam no desligamento.

Figura 33. Intervalos de confiança das métricas.



Fonte: Printscreen retirado da ferramenta R.

Os resultados dão indícios de que a Hipótese Nula de QP2 deve ser rejeitada. Logo, existe correlação entre os alarmes da base histórica de registros.

4. Modelo de Aprendizagem

Todos os resultados obtidos, as técnicas e os métodos utilizados nessa seção são referentes a abordagem usando árvores de decisão. O uso desta abordagem teve o mesmo intuito da apresentada nos capítulos de 1 à 7 deste

estudo: prever desligamentos abruptos. A utilização de árvores de decisão com a máquina de estados é promissora, e com potencial para funcionar bem no quesito predições de desligamentos abruptos, mas o desbalanceamento entre as classes das *features* de entrada do modelo, tornou a árvore de decisão imprecisa. Há 325 casos de tipos diferentes de alarmes encontrados no histórico e os valores das instâncias de treinamento e teste foram geradas muito desproporcionais, de modo que havia muito mais alarmes desativados do que ativados, o que causou um mal desempenho nas árvores de decisão. Diante disso, esta abordagem foi trocada pela abordagem híbrida usando o modelo de predição e a máquina de estados, também promissora.

Deste ponto em diante, será apresentada a abordagem inicial que foi descartada, usando árvores de decisão e não tendo qualquer relação com a abordagem principal usada na pesquisa.

Uma parte da pesquisa foi abandonada, devido aos resultados apresentados. Porém, nesta seção estão descritos os passos e os motivos pelos quais uma abordagem híbrida, usando um modelo de aprendizagem de máquina supervisionado, foi excluída do fluxo principal, mas sendo importante e relevante para os resultados gerais desta pesquisa. A obtenção do modelo resultou da comparação de cinco modelos de aprendizagem usando técnicas de treinamento diferentes que serão apresentadas nas próximas seções.

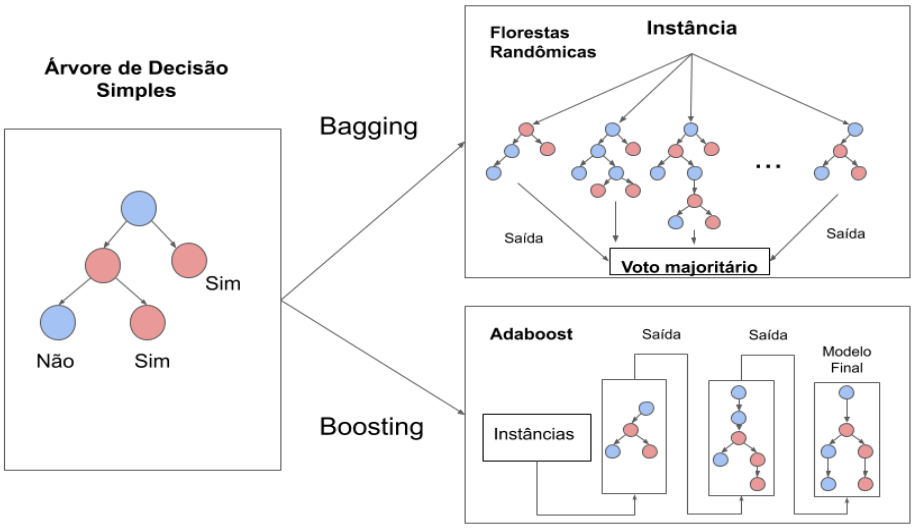
Para selecionar o modelo de predição, tornou-se necessário escolher dentre vários modelos candidatos, aquele que, comparado aos outros, segundo métricas de desempenho, apresentou maiores taxas de acertos. Como se trata de “apresentar conclusões” a partir dos “sintomas do sistema”, o problema possui características de um problema de decisão. As Árvores de Decisão (BREIMAN et al., 1984) se encaixam bem quando a ideia é classificar ou prever baseado em um conjunto de características observadas. Quando o problema exige a tomada de decisão, diante de um conjunto de informações, a árvore é um modelo adequado.

Além da Árvore de Decisão, técnicas de aprendizagem de máquina mais sofisticadas foram utilizadas com o objetivo de incrementar o desempenho das árvores simples, como as Florestas Aleatórias (BREIMAN, 1999) e o AdaBoost

(ZHU et al., 2009). Na Figura 34, é exibido como a árvore de decisão é utilizada com estas técnicas. As Florestas Randômicas geram as árvores em paralelo, e obtém como resposta final, a predição mais comum entre elas (voto majoritário). Enquanto que o AdaBoost gera um modelo final a partir de uma sequência, em que cada modelo subsequente é treinado conforme os resultados da modelagem anterior. Desta forma, a cada etapa, o modelo deve se ajustando a tendência dos dados.

Ambas as técnicas consistem em replicar várias instâncias de Árvores de Decisão simples e as combina para gerar resultados mais precisos. Estas abordagens apresentam melhores resultados do que apenas um modelo simples (BREIMAN et al., 1999; ZHU et al., 2009). Estes algoritmos constroem os modelos a partir da geração de subamostras de uma amostra original.

Figura 34. Árvore de decisão e técnicas de aperfeiçoamento de Bagging e de Boosting.

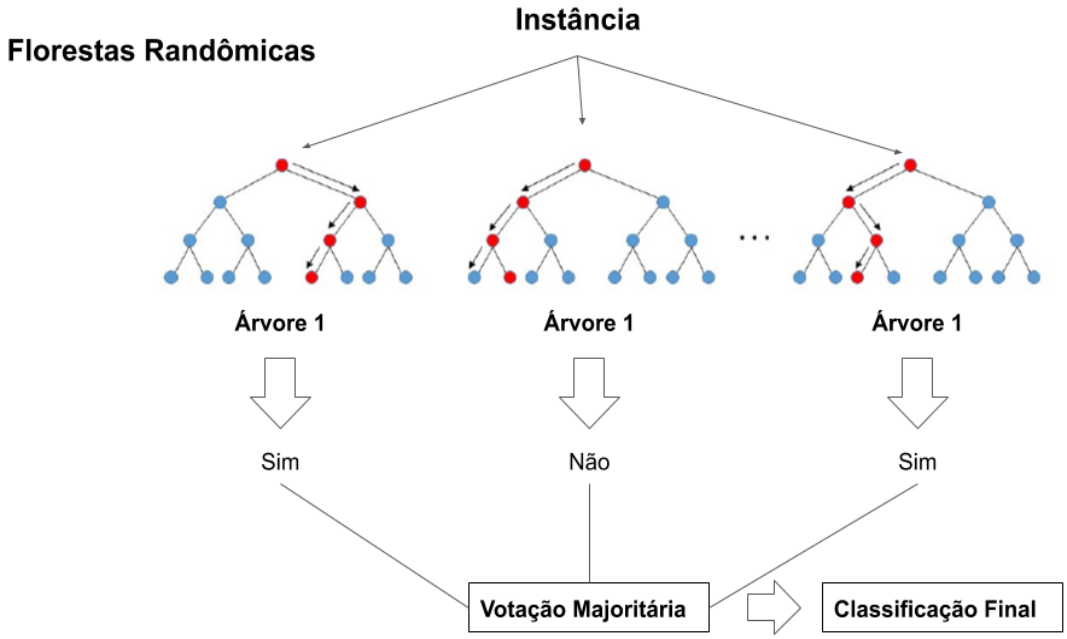


Fonte: Autoria Própria.

Cada uma das instâncias geradas pelos algoritmos de Florestas Randômicas e Adaboost é treinada a partir de subamostras da amostra original, geradas através da seleção dos itens com reposição usando uma técnica conhecida como bootstrap (KOHAVI, 1995). Ao final da execução, uma combinação de árvores de decisão será escolhida e a resposta do modelo será a predição mais comum gerada pelas árvores (votação majoritária) conforme apresentado na Figura 35.

O segundo modelo que considera uma melhoria na árvore de decisão, é baseado na técnica de Boosting (BÜHLMANN et al. 2010; FREUND et al. 1996) que consiste em uma estratégia de construir modelos base como as árvores de decisão em série para que as classificações subsequentes sejam ajustadas a favor das instâncias classificadas negativamente por classificações anteriores. Assim, as instâncias erradas pelo modelo anterior terão maior peso para o modelo subsequente. A implementação principal da técnica de Boosting é o AdaBoost.

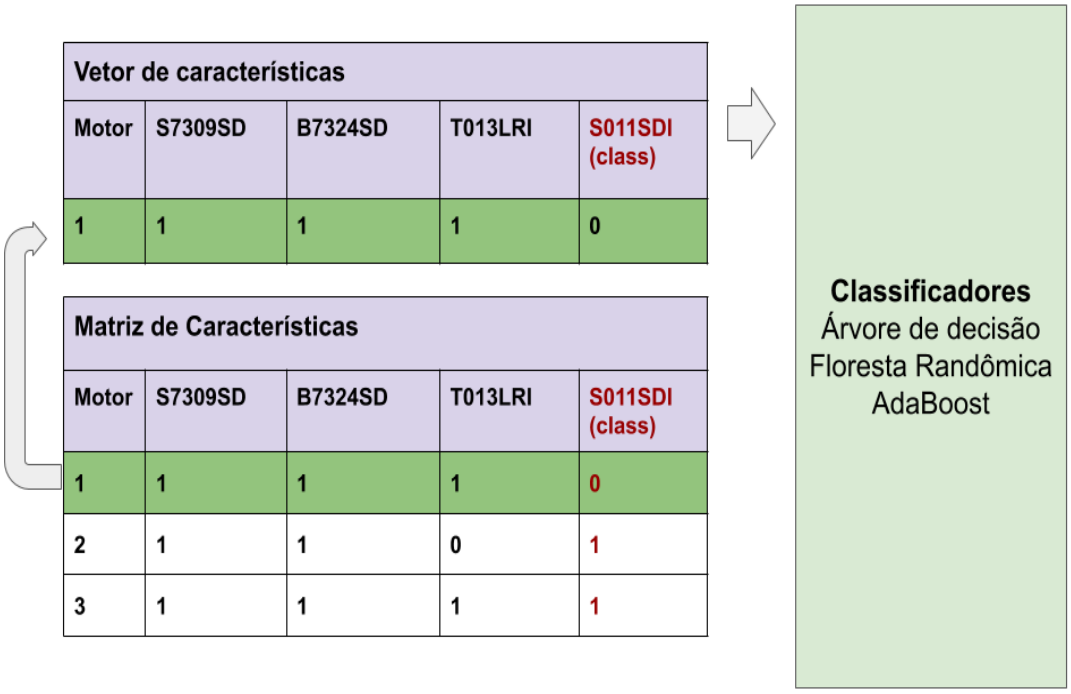
Figura 35. Florestas Randômicas.



Fonte: Adaptada de DIMITRIADIS et al. 2018.

O modelo de predição final tem como objetivo, gerar predições sobre os desligamentos. A entrada do modelo é um vetor de características (VCR), chamado assim porque cada posição do vetor indica uma característica presente no período de operação, que neste caso é identificada pelos alarmes. Cada posição do vetor contém uma *flag*, que indica a ocorrência ou não do alarme correspondente àquela posição. A saída também é uma *flag*, que indica se vai ou não ocorrer um desligamento. O conjunto de todos os VCR possíveis, extraídos do histórico de alarmes forma a Matriz de características (MCR) que será usada para treinar e testar os modelos de predição conforme apresentado na Figura 36.

Figura 36. Vetor e Matriz de Características como entradas dos modelos de predição.



Fonte: Autoria Própria.

5. Métricas de Avaliação

As métricas foram escolhidas conforme a natureza do problema. O *Recall* (FLACH et al., 2015) é uma métrica útil para avaliar a taxa de Verdadeiros Positivos e Falsos Negativos. No contexto da pesquisa, é mais importante um baixo número de Falsos Negativos do que de Verdadeiros Positivos, então o *Recall* foi a métrica adequada. As outras métricas selecionadas foram AUROC (DAVIS et al., 2006), Precisão (POWERS et al., 2011) e F1 score (SASAKI et al., 2007), com esta ordem de prioridade e calculadas a partir da matriz de confusão (SAMMUT et al., 2017).

As métricas foram geradas a cada ensaio da validação cruzada, para cada modelo. Na validação cruzada adotada, o conjunto de dados foi dividido em dez partes e vários ensaios foram realizados. A décima parte foi usada para teste e o restante para treinamento. Os ensaios da validação cruzada acabam quando todos os conjuntos gerados forem usados como teste. A validação cruzada foi definida no capítulo

6. Avaliação do Modelo de Aprendizagem de Máquina

Na primeira abordagem utilizada da pesquisa, a Árvore de Decisão foi o modelo selecionado, juntamente com duas técnicas para melhoria de desempenho. Os motivos para esta escolha serão apresentados nesta seção.

Em seguida, foi avaliada a Floresta Aleatória, apresentou o melhor desempenho, e por isto foi escolhida para compor o componente de predição do modelo híbrido preliminar. Os detalhes sobre a construção da Árvore de Decisão, das Florestas Randômicas e a comparação entre estas estão descritos a seguir.

O conjunto de dados tem 325 tipos diferentes de alarmes (características) e um desses, o alarme de desligamento abrupto, é a classe alvo. No final do experimento, um modelo de cada tipo foi selecionado e o desempenho desses foi comparado. Os três modelos escolhidos foram comparados novamente usando as métricas *Recall*, FScore, AUROC e Precisão, com esta ordem de prioridade e calculadas a partir da matriz de confusão, gerada a cada ensaio da validação cruzada para cada modelo.

Os algoritmos e modelos escolhidos como proposta inicial de solução para o problema desta pesquisa eram estado da arte, no que se refere à classificação/predição. Para selecionar o melhor, tornou-se necessário comparar o desempenho entre esses com o objetivo de escolher o que apresenta melhor taxa de acertos, conforme as métricas de desempenho.

Cada modelo foi treinado e testado com o mesmo conjunto de dados, usando a técnica de validação cruzada. Os modelos avaliados foram uma Árvore de Decisão unitária e combinada usando as técnicas Adaboost e Florestas Randômicas.

O modelo escolhido foi o construído a partir das Florestas Randômicas por ter apresentado as melhores métricas na fase de avaliação. A matriz de características gerada na etapa de ETL, é o conjunto de entrada do modelo. A implementação se divide em duas fases principais - treinamento e predição.

Após a etapa de avaliação do melhor modelo, para treinar o modelo final que será executado em produção, foram usadas todas as instâncias de períodos de operação do conjunto de dados e não mais uma parte desses. Como o modelo obtido pelas Florestas Randômicas apresentou os melhores resultados, somente este será tratado no restante desta seção.

A Floresta Randômica exige que alguns parâmetros sejam escolhidos, de forma a calibrar o modelo e obter resultados mais precisos. O primeiro parâmetro calibrado foi o número de árvores de decisão simples que deveriam ser treinadas. O total de 1000 Árvores foi o valor que saturou o erro de predição. Mais do que isso, o erro ficou estático. O segundo parâmetro importante foi uma flag que define se as árvores de decisão devem possuir profundidade limitada. Setar essa flag para falso, permite que o algoritmo escolha a configuração da árvore com uma maior liberdade, o que pode afetar os resultados do modelo.

A fase de treino é custosa e demanda horas para ser concluída. Por isto, deve ser programada para executar em um período que não afete o desempenho de produção da UTE.

Para a fase de avaliação, usando validação cruzada, a amostra foi perfeitamente balanceada com o número igual de instâncias para cada classe. Os resultados da comparação entre os modelos estão listados nas matrizes de confusão da Tabela 8, Tabela 9 e Tabela 10. Os valores consistem na média para cada métrica, obtida na validação cruzada.

Tabela 8. Matriz de Confusão: Árvore de decisão.

	Sim	**Não**
Sim	388	42
Não	16	414

Fonte: Autoria própria.

Tabela 9. Matriz de Confusão: AdaBoost.

	Sim	**Não**

Sim	404	26
Não	46	384

Fonte: Autoria Própria.

Tabela 10. Matriz de Confusão: Florestas Randômicas.

	Sim	**Não**
Sim	401	29
Não	10	420

Fonte: Autoria Própria

A partir das métricas das Tabelas 8, 9 e 10, foram calculadas métricas mais robustas para avaliação de características específicas de cada modelo. Os resultados estão listados na Tabela 11. O desempenho do modelo Florestas Randômicas foi superior a todos os outros. Embora os valores sejam muito próximos, é necessário escolher apenas o que proporcionou os valores mais elevados. Como o Random Forest utiliza uma abordagem que seleciona aleatoriamente o conjunto de características candidatas para compor o próximo nó ao invés de simplesmente selecionar o melhor entre todos os disponíveis, este modelo gera aleatoriedade e diminui a possibilidade de *sobreajuste* (COOK et al., 2016). Isto garante que os resultados não estão associados a uma adequação ao conjunto de treinamento.

Tabela 11. Métricas de desempenho dos modelos de predição usando a técnica de validação cruzada.

Modelo	**Precisão**	***Recall***	**F1 score**	**AUROC**
Árvore de Decisão simples	95,4%	95%	95%	97,6%
AdaBoost	92,6%	92,4%	92,4%	97,8%
Random Forest	96,2%	96%	96%	99,2%

Fonte: Autoria própria.

7. Análise dos Fatores Relacionados

Para esta análise foi usada a ferramenta WEKA (HALL et al., 2009) e o conjunto de dados pré-processados apresentado na seção Extração, Transformação e Carregamento dos Dados. Em cada um dos intervalos de tempo definidos na etapa de extração, todas as ocorrências de alarmes foram contabilizadas, o número de ocorrências foi computado conforme apresentado na Tabela 1 para cada um dos 325 alarmes presentes na amostra fornecida pela usina em cada um dos 129 períodos considerados. Cada linha da tabela corresponde a um período. Se um alarme A1 ocorreu 12 vezes no período 1, então o valor da célula da tabela correspondente é 1.

Para todos os alarmes que não ocorreram, os valores serão 0 nas células correspondentes. Para o alarme de desligamento, que é considerado a variável alvo, se o alarme ocorreu uma ou mais vezes, então a célula correspondente terá valor “Sim”, pois a variável é categórica e não numérica. Se ele não ocorreu nenhuma vez, o valor será “Não”. No Quadro 3, por exemplo, o alarme de desligamento ocorreu nos períodos 3 e 4. Cada alarme do histórico corresponde a um fator e os níveis de fatores são os valores contabilizados para cada tipo de alarme em cada período. Por exemplo, conforme o Quadro 3, os níveis de fatores apresentados para o fator A1 foram 1, 1, 0 e 1.

Quadro 3. Exemplo da computação de cada período.

		A1	A2	A3	Desligamento
Período	1	1	0	1	Não
	2	1	0	1	Não
	3	0	1	0	Sim
	4	1	1	1	Sim

Fonte: Autoria Própria.

O algoritmo usado para avaliar a relação entre os alarmes foi o de avaliação de atributos correlacionados, usando o método de ranking como

método de pesquisa, que busca quais fatores possuem maior índice de correlação com a classe. Para o caso da pesquisa, os fatores são os alarmes e a classe seria a ocorrência ou não do desligamento. A correlação de cada alarme e o desligamento é calculada a partir do coeficiente de correlação de Pearson, conforme definido na seção Métricas no Apêndice B.

Todos os alarmes encontrados na análise de fatores relacionados foram submetidos a análise dos especialistas da usina para confirmação da validade dos resultados. Alguns já foram confirmados como sendo influentes como é o caso do alarme B7324SD cujas ocorrências antecederam muitos dos desligamentos registrados. Isso ocorreu por causa da natureza de algumas falhas registradas no histórico.

A relação forte deste alarme com o desligamento sugere que há relação entre os desligamentos e outras falhas, consequentemente os alarmes correspondentes também possuem relação. Se os outros alarmes também forem confirmados como relacionados ao desligamento, será rejeitada a hipótese P1-H0 e se confirmará a hipótese de que os desligamentos possuem relação com outros alarmes.

8. Etapas para a Execução do Modelo de Aprendizagem

Os resultados apresentados nesta seção sobre modelos de aprendizagem se referem aos resultados iniciais obtidos a partir de avaliações iniciais, utilizando uma abordagem com turnos.

É importante destacar, para esta fase, o conceito de causalidade. A causalidade não se trata de algo verdadeiro, absoluto que ocorre sempre da mesma maneira, repetindo as mesmas condições, mas de causas possíveis. Por exemplo, se um poste cai na rua, os motivos podem ser os mais diversos. Pode ter sido um acidente com carro, uma árvore que caiu, um desmoronamento, o próprio poste que quebrou e etc. Um levantamento de todos os incidentes, a partir da base de dados da prefeitura, mostraria as causas que fizeram os postes caírem nos últimos anos. Os motivos podem variar ou terem sido o mesmo em todos os casos. Algumas causas podem ser

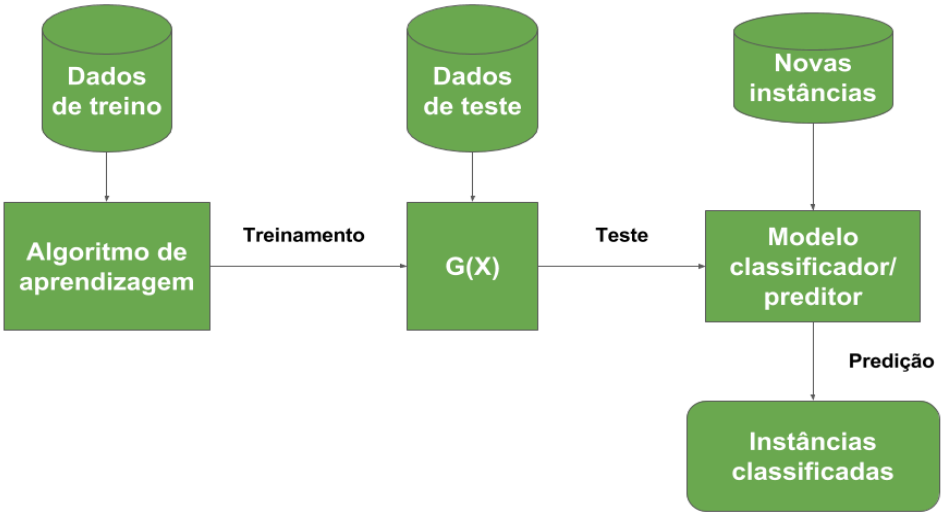
mais frequentes que outras ou alguns casos, mesmo sendo possíveis, nunca ocorreram.

Usando a mesma analogia no contexto da pesquisa, isso significa que quando um determinado alarme ocorre, as causas que desencadearam a sua ativação podem ser das mais diversas. Um desligamento pode ocorrer por centenas ou milhares de motivos, as causas não são sempre as mesmas. Por isso, ao realizar a predição, o modelo também deverá informar a probabilidade do próximo alarme ocorrer. A probabilidade equivale a quantidade de vezes, dentro do histórico, que a sequência se repetiu.

A saída do modelo será a previsão de que irá ou não ocorrer o próximo alarme, um desligamento, por exemplo, com uma probabilidade associada a essa previsão, considerando as condições atuais do sistema, refletida nas sequências de eventos emitidas.

O modelo de predição desenvolvido foi baseado na aprendizagem de máquina e no conjunto de dados, que compõem os parâmetros das fases de treinamento e teste do modelo. O processo de construção de um modelo de aprendizagem é mostrado na Figura 37. A relação entre as características extraídas dos dados é dada por uma função f desconhecida. O objetivo do modelo de aprendizagem é encontrar uma função g que seja igual ou próxima de f, capturando a relação entre as variáveis envolvidas.

Figura 37. Etapas para construção do modelo de aprendizagem de máquina.



Fonte: Autoria própria.

9. Considerações Finais

A partir dos resultados obtidos, pode-se concluir que:

- Os alarmes encontrados, a partir da Análise de Fatores Relacionados, usando-se a métrica coeficiente de correlação de Pearson e análise de turnos, apresentaram indícios de que existe relação entre as ocorrências.

- A precisão dos modelos construídos apresentou indícios de que é possível construir um classificador capaz de prever os desligamentos, porém, ainda não se pode afirmar se isto pode ser feito com antecedência necessária para a tomada de ações.

- A abordagem com árvores de decisão foi abandonada porque a abordagem utilizando o modelo de predição foi suficiente. Com o modelo de aprendizagem usando florestas randômicas, o resultado estava entrando em contradição com o MGMP. Isto ocorreu porque as árvores não são boas classificadoras para *features* com poucas informações. No caso da pesquisa, 324 *features* que são os valores correspondentes à ocorrência de cada alarme na matriz de características possuem muitos valores zerados.

Apêndice B

Análise Preliminar dos Experimentos

Neste apêndice, serão descritos os dados usados para entrada do modelo de predição, as métricas de avaliação, informações sobre a coleta de dados, experimentos realizados e outras questões relacionadas à pesquisa.

1. Fatores

Foram utilizados os dados brutos coletados dos sensores dos equipamentos, dos quais foram extraídas as informações do tipo booleana, para treinamento

do modelo híbrido supervisionado de predição. A quantidade de fatores é o número de tipos de alarmes encontrados na base histórica.

2. Métricas

Para avaliação da correlação entre os alarmes, foram utilizadas as métricas a seguir.

Coeficiente de correlação de Pearson: É uma métrica usada para calcular o quanto duas variáveis são proporcionais geometricamente. O coeficiente de correlação de Pearson é dado pela Equação 3.

$$\rho = \frac{\sum_{i=1}^{n}(xi - \bar{x})(yi - \bar{y})}{\sqrt{\sum_{i=1}^{n}(xi - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(yi - \bar{y})^2}} \ . \quad (3)$$

As variáveis xi e yi são a quantidade de vezes em que os alarmes X e Y ocorreram em um determinado período de tempo, em que i é a i-ésima ocorrência, $\bar{x}$ é a média da variável X e $\bar{y}$ é a média da variável Y. O valor de ρ varia de -1 a 1 em que -1 indica uma forte correlação inversamente proporcional e 1 forte correlação diretamente proporcional entre X e Y.

Total de pontos: É a quantidade de intervalos de tempo entre dois alarmes dentro de um turno.

IQR: É a diferença entre o 3º e o 1º quartil da distribuição de intervalos entre os alarmes.

Desvio padrão: É o nível de dispersão da distribuição de intervalos.

Mediana: É o valor que divide ao meio a distribuição dos intervalos de tempo ordenados de forma crescente.

P2: F1 score (S) e RMSE.

F1 score: É a média harmônica entre a Precisão e o *Recall*. O valor 1 indica que ambos, Precisão e *Recall* são “perfeitos”. É considerada uma boa métrica para conjuntos desbalanceados.

A Precisão de um modelo indica o número real de instâncias positivas classificadas corretamente dentre todas as que foram classificadas como positivas pelo classificador, e o *Recall* indica o quanto o modelo classifica corretamente a quantidade de instâncias pertencentes a uma determinada classe. O *F1 score* e o *RMSE* são calculados conforme apresentado nas Equações 4 e 7, respectivamente.

$$F1\ score = \frac{2 \text{ x } Precisão \text{ x } Recall}{Precisão + Recall} \ . \tag{4}$$

A Precisão e o *Recall* são obtidos respectivamente, conforme as Equações 5 e 6.

$$Precisão = \frac{VP}{VP + FP} \ . \tag{5}$$

$$Recall = \frac{VP}{VP + FN} \ . \tag{6}$$

O valor “Sim” na variável alvo indica que ocorreu um desligamento e “Não” indica que não ocorreu. O valor “Sim” representa o valor positivo enquanto que “Não” é o valor negativo para predição. Então, VP são os verdadeiros positivos, aquelas instâncias que o modelo acertou como sendo “Sim”, VN são os verdadeiros negativos, aquelas instâncias em que o modelo acertou como “Não”, FP são os falsos positivos, aquelas instâncias que o modelo previu como “Sim” (positivas), mas na verdade eram “Não” (negativas) e FN são os falsos negativos, aquelas instâncias em que o modelo previu como “Não”, mas na verdade eram “Sim”.

Os FN apresentam uma maior importância na construção do modelo

pois, na prática indicam aquelas instâncias que foram classificadas como “Não” mas na verdade são “Sim” são aquelas situações em que o modelo mostrou que não iria acontecer um desligamento mas na verdade aconteceu. Se na prática o modelo não acertar bem as situações em que ocorrerá desligamento então não amenizará os prejuízos causados pelos desligamentos e o sistema de predição será inútil.

RMSE: medida que indica o erro do modelo usando o conjunto de teste, calculado conforme a Equação 7.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (yi - \hat{y}i)^2} \quad . \tag{7}$$

Na equação, *n* é o total de instâncias do conjunto de teste, em que yi é o valor real e $\hat{y}i$ é o valor predito pelo modelo. Na prática, o algoritmo usará 1 para “Sim” e 0 para “Não”.

3. Forma de Coleta de Dados para construção do modelo híbrido

Os dados que compõem a base para esta parte da pesquisa constituem uma amostra coletada de janeiro a agosto de 2017, de forma automática, pelo sistema da usina. Os alarmes serão ordenados pelo instante de tempo em que foram emitidos e, posteriormente, a sequência será subdividida em blocos delimitados por um instante inicial (momento em que o motor foi iniciado) e um instante de tempo final (momento em que o motor foi desligado de forma abrupta ou não). Cada bloco foi rotulado segundo a ocorrência ou não de um desligamento durante o período correspondente. Após isso, foram extraídas as contagens das ocorrências de cada tipo de alarme, que, juntamente com os rótulos, compuseram a entrada do modelo de predição. O Apêndice B, na seção *Extração, Transformação e Carregamento dos Dados***,** são apresentados mais detalhes sobre o processo de extração de informações.

4. Design do Experimento

A pesquisa é específica, pois a realização é de interesse da UTE e *offline*, pois a base de dados não é alimentada em tempo real. O experimento está sendo realizado com o objetivo de escolher o modelo de predição que apresentar as melhores métricas, listadas na seção Métricas para a Questão de Pesquisa P2.

5. Significância

Todos os testes estatísticos possuem 95% de significância, ou seja, α = 0,05.

6. Unidades experimentais

As unidades experimentais do estudo de caso são os subconjuntos de treinamento e de teste gerados a cada ensaio na divisão da base de dados.

7. Projeto

O projeto foi referente à análise descritiva dos dados, para identificação das distribuições das frequências de alarmes e correlação entre os alarmes, com apenas um fator de n níveis (o valor n será igual a quantidade de modelos escolhidos na Atividade 8). Os conjuntos de dados utilizados para treinamento e teste foram gerados de forma aleatória. O experimento teve 10 ensaios, uma quantidade suficiente para comparar os modelos de predição.

8. Execução do experimento

Cada um dos 10 ensaios do experimento, usando a técnica de validação cruzada, foi dividido em três etapas:

1. Treinamento do modelo preditor;
2. Teste do modelo preditor.

Os conjuntos de treinamento e teste na validação cruzada foram divididos, conforme o Quadro 4.

Quadro 4. Divisão do conjunto de dados para a etapa de treinamento e validação dos modelos de predição.

Modelo preditivo	Porcentagem de dados de treino	Porcentagem de dados de teste
<Tipo de modelo de predição>	70%	30%

Fonte: Autoria Própria.

O tipo de design selecionado para o experimento é o Design Randomizado Pareado, pois há somente um fator (Modelo de Predição), com *n* níveis (todos os modelos que serão avaliados) e existe a influência de *extraneous variables* como o desbalanceamento dos conjuntos de treinamento e teste. Isto altera os resultados, pois a precisão dos modelos pode ser afetada pela falta de exemplos.

Após a execução de todos os ensaios, o modelo que apresentou maior F1 score e menor RMSE, calculados no conjunto de teste, foi a Árvore de Decisão construída com o algoritmo J48. O valor do F1 score alto e RMSE baixo calculados no conjunto de testes, caracteriza um modelo que prever bem os desligamentos em situações desconhecidas.

9. Ameaças à Validade

Ameaça de conclusão

a. Os resultados podem apresentar erros do tipo I ou II devido à quantidade insuficiente de amostras de períodos de operação contendo exemplos de desligamento normal e abrupto.
b. O total de tipos diferentes de alarmes encontrados na base de dados pode tornar o modelo de predição impreciso para ocorrências devido à ocorrência de alarmes que nunca foram emitidos;
c. A forma como as informações serão extraídas pode influenciar nos resultados se houver dados inconsistentes.

d. Conclusões incompletas podem ser tomadas se os períodos de operação não forem definidos de forma correta, devido as dificuldades envolvendo a identificação exata dos períodos de operação.

Ameaça de Design (constructo): Os fatores (tipos de alarmes) podem não ser suficientes para refletir toda informação idealizada pelo experimento.

Ameaça Externa: Há possibilidade de que os resultados não sejam correspondentes ao funcionamento real dos equipamentos.

Ameaça Interna: Não foram encontradas ameaças internas.