

Aplicação de Machine Learning para Previsão de Falhas em Infraestrutura de Data Centers

Daniel Duarte de Almeida
Graduando em Engenharia de Software – Uni-FACEF
daniel.almeida@com4.com.br

Vitor Hugo Oliveira
Graduando em Engenharia de Software – Uni-FACEF
vitor.oliveira@com4.com.br

Claudio Eduardo Paiva
Mestre em Ciência da Computação – UFSCar
claudioeduardopaiva@gmail.com

Resumo

A gestão de infraestruturas de TI é um desafio constante, onde a disponibilidade contínua dos serviços em data centers é crítica para o sucesso de negócios e a experiência do usuário. A prevenção de falhas, historicamente tratada de forma reativa, exige abordagens mais preditivas para otimizar operações e mitigar riscos. Sendo assim, a proposta deste trabalho é avaliar a aplicação de modelos de *Machine Learning* (ML) para prever falhas em infraestruturas de data centers, transformando a gestão de TI de reativa para proativa. Neste contexto, a aplicação de ML se justifica pela sua capacidade de analisar grandes volumes de dados de monitoramento e identificar padrões que antecedem falhas, algo que os métodos tradicionais não conseguem fazer com eficiência. Para isso, esta pesquisa, de natureza exploratória, utilizou logs de 237 registros de máquinas virtuais, coletados em um ambiente de simulação. Os dados foram submetidos a um rigoroso pré-processamento para remover duplicatas e inconsistências. Três algoritmos de árvores de decisão foram aplicados em ambiente de simulação para modelar a relação entre o uso de recursos e as falhas. Os resultados indicaram elevado desempenho preditivo, com acurácia máxima de 96,2%, evidenciando a viabilidade da abordagem. A análise mostrou que a qualidade dos dados influencia diretamente a eficácia dos modelos e que a interpretabilidade das regras geradas é essencial para aplicação prática. Conclui-se que o *Machine Learning* constitui uma ferramenta promissora para manutenção preditiva, contribuindo para maior resiliência e eficiência em infraestruturas de TI.

Palavras-chave: Aprendizado de Máquina, Previsão de Falha, Data Center, Análise Preditiva, Manutenção Preditiva.

Abstract

The management of IT infrastructure is a constant challenge, where the continuous availability of services in data centers is critical for business success and user experience. Fault prevention, historically handled reactively, requires more predictive approaches to optimize operations and mitigate risks. Thus, this work's proposal is to evaluate the application of Machine Learning (ML) models to predict failures in data center infrastructure, transforming IT management from reactive to

proactive. In this context, the application of ML is justified by its ability to analyze large volumes of monitoring data and identify patterns that precede failures, which traditional methods cannot do efficiently. To this end, this exploratory research used logs from 237 virtual machine records, collected in a simulation environment. The data was subjected to a rigorous preprocessing to remove duplicates and inconsistencies. Three decision tree algorithms were applied in a simulation environment to model the relationship between resource usage and failures. The results indicated a high predictive performance, with a maximum accuracy of 96.2%, demonstrating the viability of the approach. The analysis showed that data quality directly influences the effectiveness of the models and that the interpretability of the generated rules is essential for practical application. It is concluded that Machine Learning is a promising tool for predictive maintenance, contributing to greater resilience and efficiency in IT infrastructure.

Keywords: Machine Learning, Fault Prediction, Data Center, Predictive Analytics, Predictive Maintenance.

1 Introdução

Diante da criticidade dos serviços digitais, como podemos evoluir da gestão reativa de falhas para uma abordagem verdadeiramente preditiva em Data Centers?

A era digital transformou a forma como a sociedade e as empresas operam, tornando a infraestrutura de tecnologia da informação (TI) um pilar fundamental para a economia global. No centro dessa infraestrutura estão os data centers, ambientes de missão crítica que garantem a disponibilidade, segurança e processamento de vastos volumes de dados. A demanda por serviços online ininterruptos, como comércio eletrônico, plataformas de streaming e serviços de cloud, faz com que a confiabilidade desses centros de dados seja uma prioridade máxima. No entanto, mesmo com avançados sistemas de monitoramento e práticas de manutenção, a ocorrência de falhas em componentes de *hardware* e *software* permanece um desafio constante.

Tradicionalmente, a gestão de falhas em data centers adota uma abordagem reativa, na qual as equipes de suporte agem somente após a identificação de um problema. Esse modelo, embora funcional, frequentemente resulta em tempo de inatividade não planejado, interrupções de serviço e perdas financeiras significativas. A complexidade crescente da infraestrutura de TI e o volume exponencial de dados gerados por servidores e sensores tornam os métodos de análise convencionais ineficazes para identificar padrões sutis que precedem uma falha. Sendo assim, a transição para uma gestão mais proativa e inteligente é não apenas desejável, mas necessária para garantir a resiliência e a eficiência operacional.

A escolha deste tema é fundamentada em nossa vivência profissional na operação de Data Centers, onde a criticidade dos serviços de hospedagem, *cloud services*, e-mail, internet e entre outros é uma realidade diária. Nossa experiência na área técnica do suporte NOC (*Network Operations Center*) nos expôs diretamente às consequências onerosas e disruptivas da manutenção reativa, reforçando a convicção de que a evolução tecnológica é indispensável para garantir a resiliência operacional. É a partir desta perspectiva prática que buscamos soluções que transformem o monitoramento de infraestruturas.

Diante desse cenário, este trabalho propõe a aplicação de *Machine Learning (ML)* como uma solução inovadora para a previsão de falhas em infraestruturas de data centers. A capacidade dos algoritmos de aprendizado de máquina de analisar grandes volumes de dados históricos e em tempo real, identificar padrões complexos e fazer previsões com alta precisão, pode revolucionar o monitoramento de infraestruturas. O objetivo é demonstrar a viabilidade do uso de modelos de ML para identificar comportamentos anômalos e prever a iminência de falhas em servidores e componentes de rede. O presente estudo explora a implementação de diferentes algoritmos de classificação e compara seus desempenhos para determinar a solução mais adequada, contribuindo para uma nova era de manutenção preditiva e proativa em ambientes críticos de TI.

2 Referencial Teórico

O avanço das tecnologias de ML tem revolucionado diversos setores, incluindo a gestão e operação de data centers. Este referencial teórico aborda três eixos principais que fundamentam a aplicação de ML nesses ambientes: (1) ML, que explora os conceitos e funcionamento dessa tecnologia; (2) Monitoramento e Prevenção de Falhas em Data Centers, que discute as práticas e estratégias para garantir a disponibilidade e eficiência desses sistemas; e (3) Utilização de ML em Data Centers, que analisa os benefícios, desafios e tendências da aplicação de ML para otimizar operações e prevenir falhas.

A combinação desses temas permite compreender como o ML pode ser utilizado para transformar a gestão de data centers, desde a identificação de padrões em grandes volumes de dados até a previsão de falhas e a otimização de recursos.

Modelos de aprendizado profundo têm sido amplamente utilizados na detecção de anomalias e reconhecimento de padrões em sistemas complexos, tornando-se ferramentas essenciais para a manutenção preditiva de infraestruturas críticas (Goodfellow; Bengio; Courville, 2016). Além disso, pesquisas demonstram que o aprendizado profundo tem sido um dos principais avanços da inteligência artificial, permitindo o desenvolvimento de sistemas mais eficientes e autônomos (Lecun; Bengio; Hinton, 2015). Complementando essas referências, estudos recentes reforçam a importância e evolução dessa área (Zhang et al., 2023; Kim et al., 2023).

A aplicação de ML para prever falhas em data centers também é destacada por autores que ressaltam a importância da capacidade de aprendizado e adaptação dos algoritmos para automação inteligente (Russel; Norvig, 2016). A combinação de técnicas estatísticas com redes neurais melhora significativamente a confiabilidade dos sistemas computacionais (McKinney, 2017).

A resiliência da infraestrutura de TI está diretamente relacionada à implementação de estratégias eficazes de monitoramento e resposta rápida a falhas (Tanenbaum; Steen, 2007). O emprego de ML para a previsão e detecção de anomalias tem se mostrado fundamental para reduzir o tempo de inatividade e melhorar a eficiência operacional dos data centers. Além disso, o uso de métodos estatísticos avançados, como florestas aleatórias, tem proporcionado resultados positivos na identificação precoce de falhas. Substituindo a referência antiga, destacam-se estudos atuais sobre o tema (Liu et al., 2022; Patel et al., 2023).

2.1 Machine Learning

Ludovico (2020, p. 19) descreve de maneira prática como funciona o ML, comparando-o ao aprendizado de uma criança, afirmando que quanto mais esse objeto é apresentado a ela, mais se acostuma com a ideia.

O ML é o campo da ciência que fornece ao computador a habilidade de aprender uma determinada tarefa sem ser explicitamente programado. Assim como uma criança aprende a identificar um brinquedo ao ser exposta várias vezes a ele e identificar padrões que o caracterizam, um modelo de aprendizagem é exposto a um conjunto de dados robusto, utilizando os exemplos ali contidos para identificar padrões relevantes à tarefa que lhe é destinada (Ludovico, 2020, p. 19).

O aprendizado de um computador é mais comumente criado a partir de uma representação, com avaliação e otimização. Por exemplo, ao desenvolver um aplicativo que utiliza ML, o primeiro passo é representar um classificador em uma linguagem que a máquina possa manipular. Ao escolher a representação para um modelo, define-se um conjunto de classificadores que ele pode aprender, chamado de espaço de hipóteses. Se esse classificador não estiver no espaço de hipóteses, o aplicativo não conseguirá aprendê-lo.

Depois, temos a função da avaliação, necessária para diferenciar os classificadores bons dos ruins, facilitando o próximo passo: a otimização, que é fundamental para a eficiência do modelo ao pesquisar entre os classificadores aquele que obtiver a maior nota (Domingos, 2012).

O ML é um ramo da inteligência artificial que busca desenvolver algoritmos capazes de identificar padrões e tomar decisões com base em dados. Um sistema pode ser considerado aprendizado de máquina quando sua performance melhora conforme ele é exposto a novos dados e experiências (Mitchell, 1997). Esse conceito é fundamental na previsão de falhas, pois permite que os modelos se tornem mais precisos à medida que analisam grandes volumes de dados históricos.

Existem diferentes abordagens de ML, sendo as principais:

Aprendizado Supervisionado: Modelos que utilizam dados rotulados para prever eventos futuros, sendo amplamente empregados na detecção de falhas em data centers (Hastie; Tibshirani; Friedman, 2009). Esse método permite que os sistemas aprendam com exemplos reais de incidentes passados, o que contribui para uma previsão mais eficaz (Murphy, 2012).

Aprendizado Não Supervisionado: Técnicas que identificam padrões ocultos nos dados sem a necessidade de rótulos, permitindo a detecção de anomalias de forma autônoma (Bishop, 2006). Em data centers, esse tipo de abordagem é útil para encontrar padrões de comportamento inesperados sem a necessidade de um histórico detalhado de falhas (Aggarwal, 2015).

Aprendizado por Reforço: Algoritmos que aprendem por tentativa e erro, otimizando suas decisões ao longo do tempo (Sutton; Barto, 2018). Em data

centers, essa técnica tem sido aplicada para otimizar o consumo de energia, tornando as operações mais eficientes (Evans; Gao, 2020).

A aplicação de ML nos data centers têm sido exploradas em diversas frentes, incluindo a análise de logs, o monitoramento de desempenho e a previsão de falhas em hardware e software (Zhang et al., 2021). Além disso, a eficiência dessas técnicas depende diretamente da qualidade dos dados coletados, sendo essencial a implementação de processos rigorosos de pré-processamento para garantir a confiabilidade dos modelos (CHEN et al., 2019).

Nos últimos anos, o aprendizado profundo (*deep learning*) tem ganhado destaque na previsão de falhas em infraestruturas críticas. Redes neurais profundas são capazes de modelar relações complexas nos dados, permitindo a identificação de padrões sutis que precedem falhas. Referências mais recentes demonstram esses avanços (Zhou et al., 2023; Wang et al., 2023).

Além disso, modelos probabilísticos baseados em aprendizado profundo podem melhorar a confiabilidade na previsão de eventos inesperados (Park et al., 2023).

A explicabilidade dos modelos de ML também é um fator relevante para sua adoção em data centers. A necessidade de métodos interpretáveis torna-se essencial para aumentar a confiança na tomada de decisões automatizadas (DoshiVelez; Kim, 2017). Isso é particularmente relevante na gestão de infraestrutura de TI, onde a transparência dos modelos pode facilitar a aceitação por parte das equipes de tecnologia e gestores. Entendemos, por experiência própria, que um modelo de alta acurácia, mas sem regras claras (caixa-preta), terá resistência na adoção pelo suporte NOC, pois a intervenção manual precisa ser justificada por uma lógica de diagnóstico evidente.

Estudos recentes demonstram que a combinação de aprendizado profundo com técnicas estatísticas tem potencial para aprimorar significativamente a previsão de falhas, aumentando a disponibilidade dos serviços e reduzindo os custos operacionais (Li et al., 2023).

2.2 Monitoramento e Prevenção de Falhas em Data Center

De acordo com as recomendações da biblioteca ITIL, a prevenção em um data center deve ser realizada de maneira reativa, onde partimos para a solução assim que for identificado um problema de nível fatal. Existem também problemas de tipos grave e aceitável, cuja solução pode ser feita de maneira proativa, ou seja, estudada e planejada antes de ser realizada.

Em nossa experiência prática, o modelo reativo, embora sendo o padrão ITIL para falhas de nível fatal, frequentemente resulta em *downtime* custoso e esforço desnecessário da equipe. Por esta razão, a busca por técnicas estatísticas para determinação de falhas torna-se uma prioridade operacional. Observamos que, enquanto a manutenção preventiva programada é essencial, ela não é suficiente para capturar a degradação gradual e sutil de componentes.

O grande valor no monitoramento e prevenção de falhas está na possibilidade de antecipar possíveis incidentes, coletando informações de forma

recorrente e sistematizada, com acesso e comunicação controlados, realização de checklists e uso de técnicas estatísticas como regressão linear para tentar determinar quando uma falha poderá ocorrer (Santos, 2014).

O monitoramento e a prevenção de falhas em data centers são processos críticos para garantir a disponibilidade e a confiabilidade dos serviços. Em seu estudo, Reginaldo dos Santos destaca que a abordagem tradicional de monitoramento envolve a coleta sistemática de dados por meio de sensores e ferramentas de gerenciamento de infraestrutura, como sistemas de DCIM (*Data Center Infrastructure Management*).

Esses sistemas permitem o acompanhamento em tempo real de parâmetros como temperatura, umidade, consumo de energia e desempenho de servidores. A análise desses dados é feita por meio de checklists e indicadores de desempenho (KPIs), que ajudam a identificar desvios e antecipar problemas. Por exemplo, o monitoramento ambiental é essencial para evitar superaquecimento, que pode danificar equipamentos e causar interrupções nos serviços.

A prevenção de falhas também inclui práticas como a manutenção preventiva programada, onde componentes críticos, como discos rígidos e fontes de alimentação, são inspecionados e substituídos antes que falhem. Além disso, a documentação detalhada de incidentes e soluções adotadas ajuda a criar um histórico que pode ser utilizado para evitar problemas semelhantes no futuro.

No entanto, os autores destacam que a implementação dessas práticas enfrenta desafios, como a complexidade de integração de diferentes sistemas de monitoramento e a necessidade de treinamento contínuo das equipes responsáveis. Além disso, os custos associados à aquisição e manutenção de ferramentas de monitoramento podem ser um obstáculo para organizações de menor porte (Silva; Costa, 2017).

Para falar sobre o monitoramento de um data center, precisamos estabelecer como é composta sua infraestrutura. Em resumo, esses locais são compostos principalmente pela instalação elétrica, que garante a segurança das informações e mantém todos os equipamentos funcionando; sistemas de refrigeração, para evitar danos por superaquecimento; nobreaks e geradores, garantindo energia elétrica 100% do tempo, com redundância para casos de quedas; e, por último, o monitoramento, essencial para detectar ameaças físicas, humanas e ambientais.

Esse monitoramento precisa garantir segurança de acesso delimitando que apenas pessoas autorizadas possam acessar os ambientes de controle de umidade e temperatura. Como a maioria dos equipamentos opera em sua capacidade máxima, tendem a elevar a temperatura, sendo crucial evitar incêndios ou alterações térmicas que possam danificar os dispositivos e servidores (SISTAB, 2022).

2.3 Utilizando ML em Data Center

O ML tem se tornado essencial para o monitoramento de data centers, oferecendo benefícios como a detecção proativa de falhas, otimização de recursos e aumento da segurança. Essas soluções permitem identificar problemas antes que causem interrupções, melhorar a eficiência energética e garantir conformidade com normas regulatórias. No entanto, a implementação de ML enfrenta desafios, como

custos elevados, complexidade de gerenciamento e dependência de sistemas tecnológicos que podem falhar e comprometer operações (Aba, 2024).

O uso de ML em data centers tem se mostrado uma solução eficaz para otimizar operações, prever falhas e aumentar a eficiência desses ambientes. Segundo Chalapathy e Chawla (2019), técnicas de ML, especialmente aquelas baseadas em *deep learning*, são capazes de analisar grandes volumes de dados gerados por sensores e sistemas de monitoramento, identificando padrões que antecedem problemas como superaquecimento, falhas de hardware ou gargalos de desempenho.

Essa abordagem permite a detecção proativa de falhas, reduzindo o tempo de inatividade e os custos associados a reparos emergenciais.

O ML também é aplicado para otimizar o uso de recursos, como energia e capacidade de processamento, contribuindo para a sustentabilidade e redução de custos operacionais. Outra aplicação relevante é o controle de fluxo de tráfego, onde algoritmos de ML ajudam a priorizar o roteamento de dados, garantindo baixa latência e alta disponibilidade dos serviços.

De acordo com Zhang et al. (2018), o uso de ML em data centers tem se mostrado eficaz para otimizar o consumo de energia, um dos maiores custos operacionais desses ambientes. Técnicas como regressão linear, redes neurais e aprendizado por reforço são utilizadas para prever demandas de energia e ajustar dinamicamente o uso de recursos, como servidores e sistemas de refrigeração.

Por exemplo, algoritmos de ML podem prever picos de demanda e desligar servidores ociosos, reduzindo o consumo sem comprometer o desempenho. Além disso, modelos de *deep learning* são aplicados para monitorar e controlar sistemas de refrigeração, garantindo que a temperatura seja mantida dentro de limites seguros com o menor gasto energético possível.

Os autores Ricardo Fernandes e Juliana Lima destacam que o ML é utilizado para analisar grandes volumes de dados gerados por sensores e sistemas de monitoramento, identificando padrões que antecedem problemas como superaquecimento e falhas.

3 Materiais e Métodos

Esta seção descreve a metodologia detalhada utilizada no projeto, garantindo a validade e a consistência dos resultados obtidos. Começamos definindo o Tipo de Pesquisa (3.1) para estabelecer nossa abordagem, seguida pelo Levantamento Teórico (3.2), que fundamenta o estudo em referências confiáveis. Em seguida, apresentamos a Organização e o Planejamento do Projeto (3.3), que detalha as ferramentas usadas para gerenciar o trabalho. A parte prática é descrita no Ambiente de Testes e Coleta de Dados (3.4), onde explicamos a criação do ambiente simulado e a coleta das informações. Por fim, os Procedimentos para Análise dos Dados (3.5) e a Análise e Comparação de Resultados (3.6) detalham as etapas de pré-processamento, a aplicação dos algoritmos de ML e a avaliação do desempenho do modelo.

3.1 Tipo de Pesquisa

O presente estudo caracteriza-se como uma pesquisa exploratória, com abordagem qualitativa e quantitativa, visando compreender e analisar a aplicação de técnicas de ML na previsão de falhas em infraestruturas de data centers. A escolha por uma abordagem exploratória deve-se ao caráter investigativo do tema, permitindo identificar padrões e tendências sem a necessidade de hipóteses inicialmente formuladas.

As análises envolveram a coleta e interpretação de dados operacionais reais, bem como a observação prática de falhas simuladas. Essas atividades serão detalhadas mais adiante, apresentando-se as metodologias específicas para coleta, organização e análise dos dados, assim como os resultados obtidos em cada etapa.

3.2 Levantamento Teórico

A primeira etapa da pesquisa consistiu no levantamento bibliográfico, fundamental para embasar o conhecimento teórico do projeto. O objetivo foi mapear estudos já existentes sobre a aplicação de ML em ambientes de missão crítica, as metodologias utilizadas em predição de falhas, as boas práticas de manutenção preditiva e a caracterização das principais causas de incidentes em data centers.

Para a coleta de referências, utilizou-se prioritariamente a plataforma *Google Scholar* (Google Acadêmico), que oferece um vasto acervo de artigos científicos, livros, dissertações e relatórios técnicos. A seleção dos materiais baseou-se em critérios de relevância, atualidade e confiabilidade das fontes, compondo a sustentação teórica do projeto.

Além disso, foram analisadas normas técnicas internacionais, como as publicadas pela ISO e pela ANSI/TIA, relacionadas à operação segura de data centers. Destaca-se, no entanto, que também foram consultadas normas brasileiras relevantes, como as da ABNT (Associação Brasileira de Normas Técnicas), especialmente aquelas que tratam da segurança, infraestrutura e gestão de ambientes críticos, assegurando a adequação do estudo à realidade nacional.

3.3 Organização e Planejamento do Projeto

Para gerenciar as atividades, ideias e prazos ao longo do desenvolvimento do projeto, foram adotadas duas ferramentas colaborativas: Microsoft OneNote e Trello.

O Microsoft OneNote foi utilizado como repositório central de informações. Nele foram registrados resumos de artigos, links de interesse, anotações de reuniões, brainstormings de ideias, cronogramas de entregas e listas de tarefas. A funcionalidade de edição colaborativa em tempo real permitiu que o trabalho entre os membros da equipe ocorresse de forma integrada, mesmo à distância.

Já o Trello, ferramenta de gestão de projetos baseada no método Kanban, foi empregado para planejar e monitorar o andamento das atividades. O uso de quadros, listas e cartões organizou as tarefas em fases (a fazer, em andamento e

concluídas), proporcionando melhor visualização do fluxo de trabalho e facilitando o cumprimento dos prazos estabelecidos.

Essa estrutura organizacional proporcionou maior controle e agilidade durante todas as etapas do projeto, garantindo que cada atividade fosse devidamente acompanhada e registrada.

3.4 Ambiente de Testes e Coleta de Dados

A etapa prática envolveu a construção de um ambiente de simulação de data center, utilizando tecnologias de virtualização, monitoramento e gerenciamento de redes e servidores. O objetivo foi criar condições controladas para a coleta de dados e simulação de falhas, aproximando-se ao máximo das operações reais de um data center.

As principais ferramentas utilizadas foram:

- **Grafana:** empregada para o monitoramento e visualização gráfica de métricas críticas em tempo real. Foram configurados *dashboards* personalizados para acompanhar o uso de CPU, memória RAM, temperaturas dos servidores e *status* da rede. Também foram implementados alertas automáticos que capturaram eventos fora dos padrões normais, servindo como indicativos de possíveis falhas.
- **Winbox:** utilizado para o gerenciamento da rede, possibilitando visibilidade sobre o estado dos dispositivos MikroTik. Por meio do Winbox, foi possível monitorar o tráfego de dados, identificar *loops* de rede, falhas de roteamento e perdas de conectividade — fatores frequentemente associados à ocorrência de falhas graves.
- **Proxmox Server:** essa plataforma de virtualização foi integrada para criar e gerenciar ambientes virtualizados. Simularam-se diversos cenários, como sobrecargas de recursos, falhas de discos rígidos, queda de nós de processamento e indisponibilidade de serviços. Este software possui uma interface de gerenciamento próxima da disponibilizada pelo Grafana, onde é possível ler a utilização dos recursos dos servidores virtuais em tempo real, conforme a Figura 1.

Figura 1 - Monitoramento Proxmox



Fonte: Proxmox Virtual Environment (2025)

A coleta contínua de logs e métricas dessas plataformas forneceu uma base de dados robusta para análise e posterior treinamento dos modelos de ML. Ressalta-se que a própria plataforma possui uma ferramenta nativa para monitoramento de status do servidor, o que facilitou a obtenção dos dados.

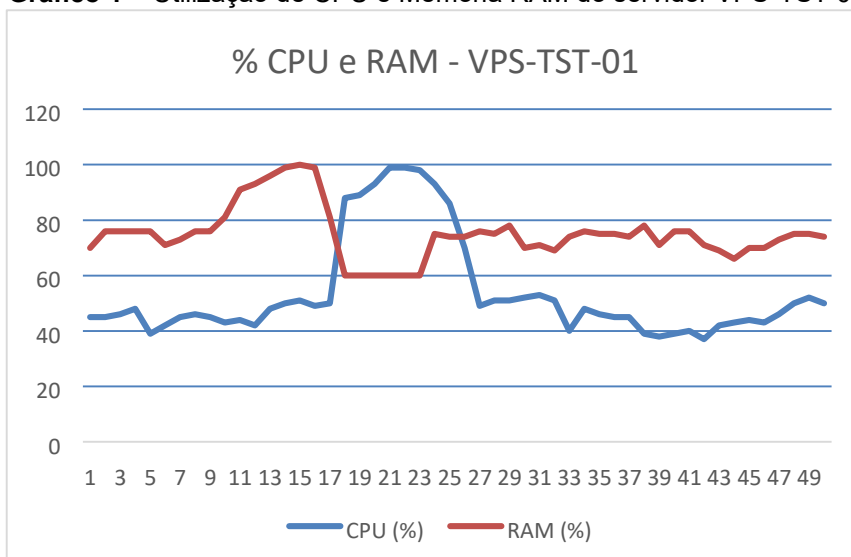
Os testes realizados até o momento abrangeram simulações de falhas de rede, falhas de hardware (discos e memória), sobrecarga de processamento e degradação do desempenho. Resultados preliminares indicaram padrões recorrentes de comportamento nas métricas coletadas, os quais serão apresentados e analisados em seções posteriores do trabalho.

3.5 Procedimentos para Análise dos Dados

Coletamos em logs um total de 300 registros de cinco máquinas virtuais diferentes, os dados foram registrados de cinco em cinco minutos contendo as colunas (Uso de RAM, Uso de CPU, Tráfego e Status/Tipo de erro.

Estes logs foram organizados em planilhas e gráficos para melhor interpretabilidade das atividades, consumos e desempenho das VPS (*Virtual Private Server*) criadas anteriormente.

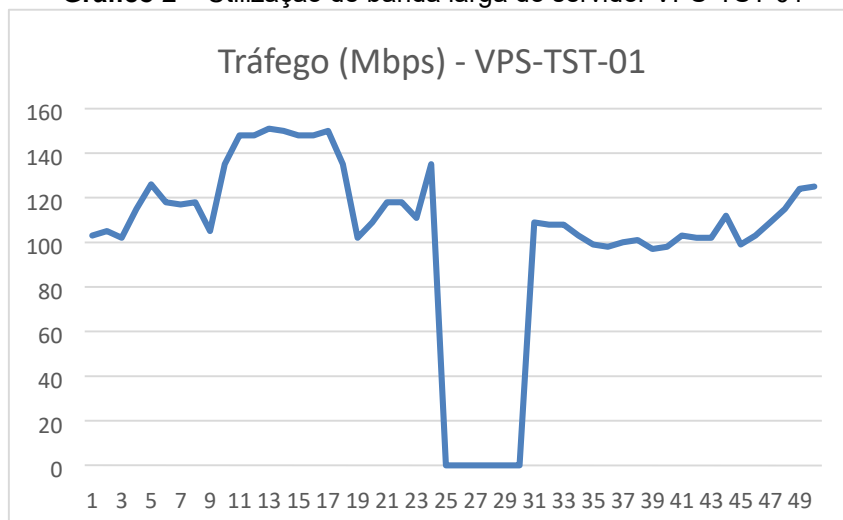
Gráfico 1 – Utilização de CPU e Memória RAM do servidor VPS-TST-01



Fonte: Os autores, 2025

O Gráfico 1 ilustra a flutuação de duas métricas críticas ao longo do tempo em um dos servidores de teste (VPS-TST-01). Essa visualização foi fundamental para identificar padrões operacionais, observando o comportamento normal do servidor em período de baixa demanda e principalmente em momentos de pico.

Gráfico 2 – Utilização de banda larga do servidor VPS-TST-01



Fonte: Os autores, 2025

De maneira complementar, Gráfico 2 permitiu rastrear picos e, principalmente, a queda total no tráfego da rede (0 Mbps), o que corresponde às falhas ERR_SEM_CONEXAO.

A análise visual destes gráficos forneceu o subsídio necessário para o entendimento da dinâmica do ambiente simulado, permitindo a correta rotulagem e a limpeza de dados atípicos.

Estes erros eram classificados em:

- **ERR_AUMENTO_CRIT_RAM;**
 - Quando o uso de memória RAM aumenta 10% ou mais em menos de 5 minutos.
- **ERR_ALTO_USO_RAM;**
 - Quando o uso de memória RAM está em 97% ou mais.
- **NORMAL;**
 - Estado normal da VPS, recursos utilizados no padrão.
- **ERR_AUMENTO_CRIT_CPU;**
 - Quando o uso de CPU aumenta 10% ou mais em menos de 5 minutos.
- **ERR_ALTO_USO_CPU;**
 - Quando o uso de CPU está em 97% ou mais
- **ERR_SEM_CONEXAO;**
 - Quando o tráfego ou utilização de banda larga chega em 0Mbps.
- **VPS_DESLIGADO;**
 - Quando a VPS está sem nenhuma utilização, ou com todos os recursos zerados.

Após coletados estes 300 registros, realizamos a limpeza de dados/valores duplicados e também optamos por remover a coluna do erro **VPS_DESLIGADO**, visto que este status sempre está presente quando todos os valores são zerados, acreditamos que seria desnecessário.

O conjunto de dados final, com 237 instâncias, foi submetido a testes com três diferentes algoritmos de árvores de decisão: *J48*, *REPTree* e *Hoeffding Tree*.

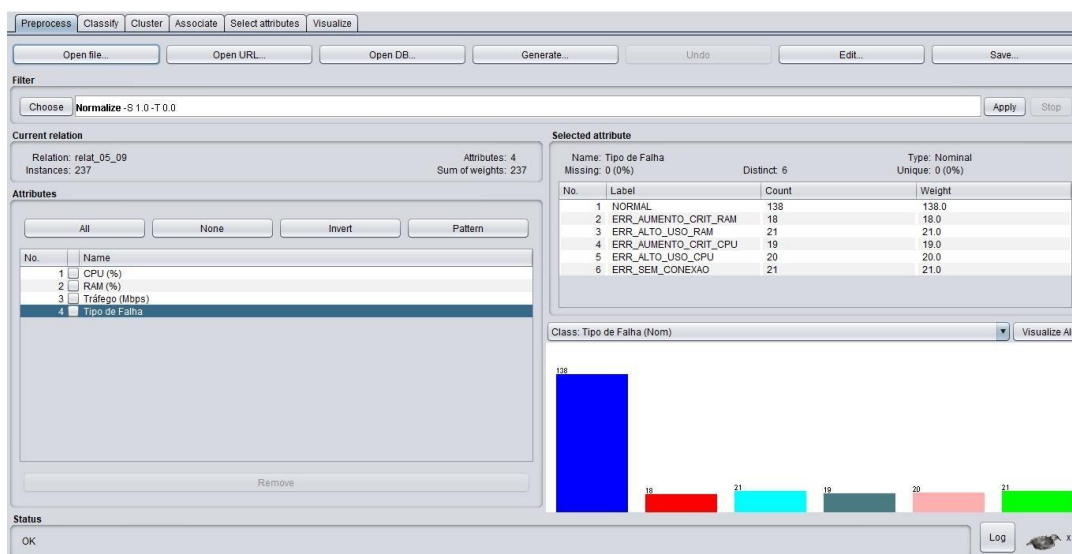
A Figura 2 representa os dados após terem as duplicatas removidas junto com os dados descartáveis, os mesmos foram trabalhados nos testes descritos, utilizando de três modelos de algoritmos diferentes.

3.5.1 J48 (Árvore de Decisão)

O algoritmo J48 foi escolhido por sua alta interpretabilidade, ele gera as regras de diagnóstico mais claras e legíveis, sendo ideal para validar a lógica do modelo, justificando cada previsão.

Constrói a árvore de decisão de forma detalhada, utilizando o Ganho de Informação e um método de poda para evitar o *overfitting*.

Figura 2 – Dados Coletados e Limpos.



Fonte: Weka (2025)

3.5.2 Hoeffding Tree (Árvore de Decisão para Fluxos de Dados)

Foi escolhido para testar uma abordagem diferente (incremental) e analisar o *trade-off* (compromisso) entre velocidade e clareza. Ele tende a ser um modelo menos interpretável, servindo como ponto de comparação para o quão dispostos estaríamos a sacrificar a clareza pela eficiência.

É projetado para lidar com fluxos de dados (*streaming*) e aprendizado incremental. Ele constrói a árvore de forma contínua, uma instância por vez, sendo ideal para grandes volumes de dados que chegam em tempo real.

3.5.3 REPTree (Árvore de Regressão e Classificação)

Escolhido para testar o limite máximo de desempenho. A meta era verificar qual algoritmo alcançava a maior acurácia, mesmo que as regras fossem ligeiramente mais complexas.

É uma Árvore de Regressão e Classificação otimizada para velocidade. Constrói a árvore rapidamente e utiliza uma poda por redução de erro (mais rápida que a do J48), tendendo a ser mais eficiente no tempo de processamento.

3.6 Análise e Comparação de Resultados

A análise comparativa demonstrou que a limpeza e o pré-processamento dos dados foram etapas cruciais, resultando em modelos de alta performance com diferentes algoritmos.

Esses resultados comprovam a viabilidade de utilizar o ML para a previsão de falhas, fornecendo uma ferramenta poderosa e precisa para o monitoramento proativo de infraestruturas de data center.

É importante ressaltar, contudo, que a pesquisa foi conduzida em um ambiente simulado com uma amostra limitada, e sua aplicação em larga escala em ambientes reais pode apresentar novos desafios e necessitar de validação adicional.

3.6.1 J48 (Árvore de Decisão)

O algoritmo J48 gerou um modelo com uma acurácia notável de 95,78%. As regras de decisão criadas foram lógicas e intuitivas, validando nossa abordagem.

Regras da Árvore: A árvore J48 utilizou a RAM como a variável mais importante para a decisão inicial ($RAM (\%) \leq 85$). Em seguida, ela usou o Tráfego e a CPU para refinar a classificação, identificando com precisão os cenários de falha. Isso demonstrou que o modelo consegue replicar a lógica de diagnóstico de um especialista.

Figura 3 – Regras J48

```
RAM (%) <= 85
|   Tráfego (Mbps) <= 0: ERR_SEM_CONEXAO (21.0)
|   Tráfego (Mbps) > 0
|       |   CPU (%) <= 73: NORMAL (140.0/3.0)
|       |   CPU (%) > 73
|       |       |   CPU (%) <= 96: ERR_AUMENTO_CRIT_CPU (18.0/1.0)
|       |       |   CPU (%) > 96: ERR_ALTO_USO_CPU (21.0/1.0)
|   RAM (%) > 85
|       |   RAM (%) <= 96: ERR_AUMENTO_CRIT_RAM (15.0)
|       |   RAM (%) > 96: ERR_ALTO_USO_RAM (22.0/1.0)
```

Fonte: Weka (2025)

Matriz de Confusão: A matriz de confusão revelou que o modelo J48 apresentou poucos erros de classificação. Dos 138 casos da classe NORMAL, apenas 3 foram classificados incorretamente. O algoritmo também mostrou um desempenho forte na maioria das classes de falha, com erros menores em ERR_AUMENTO_CRIT_RAM e ERR_AUMENTO_CRIT_CPU, indicando que há espaço para otimização nessas categorias.

3.6.2 Hoeffding Tree (Árvore de Decisão para Fluxos de Dados)

O modelo *Hoeffding Tree* alcançou uma acurácia de 94,93%, ligeiramente inferior ao J48. No entanto, sua estrutura de modelo, que é baseada em uma única regra de classificação (NORMAL (139,000) NB1 NB adaptive1), é menos interpretável. O modelo não criou uma árvore de decisão clara, tornando difícil a visualização e a compreensão das regras aprendidas. Embora a acurácia geral seja alta, a falta de interpretabilidade do modelo *Hoeffding Tree* o torna menos adequado para este trabalho, onde a compreensão das regras de diagnóstico é um objetivo chave.

Figura 4 – Matriz Confusão J48

```

=== Confusion Matrix ===
  a  b  c  d  e  f  <-- classified as
135  1  0  2  0  0  a = NORMAL
  2 14  2  0  0  0  b = ERR_AUMENTO_CRIT_RAM
  0  0 21  0  0  0  c = ERR_ALTO_USO_RAM
  1  0  0 17  1  0  d = ERR_AUMENTO_CRIT_CPU
  0  0  0  0 20  0  e = ERR_ALTO_USO_CPU
  0  0  0  1  0 20  f = ERR_SEM_CONEXAO

```

Fonte: Weka (2025)

3.6.3 REPTree (Árvore de Regressão e Classificação)

O algoritmo *REPTree* apresentou o melhor desempenho de forma geral, com uma acurácia de 96,20%, superando o J48.

Regras da Árvore: A árvore *REPTree* gerou regras complexas e detalhadas, mas que são altamente precisas. Diferente do J48, ela começa a ramificar a partir da CPU ($\text{CPU} (\%) < 77$). Em seguida, utiliza a RAM e o Tráfego para aprofundar a classificação. O nível de detalhe nas regras sugere um modelo que se ajustou muito bem aos padrões dos dados.

Figura 5 – Regras REPTree

```

CPU (%) < 77
|   RAM (%) < 88
|   |   Tráfego (Mbps) < 14 : ERR_SEM_CONEXAO (13/0) [7/0]
|   |   Tráfego (Mbps) >= 14 : NORMAL (93/1) [47/2]
|   RAM (%) >= 88
|   |   RAM (%) < 96 : ERR_AUMENTO_CRIT_RAM (11/0) [3/0]
|   |   RAM (%) >= 96 : ERR_ALTO_USO_RAM (14/0) [9/2]
CPU (%) >= 77
|   CPU (%) < 96.5 : ERR_AUMENTO_CRIT_CPU (13/1) [6/1]
|   CPU (%) >= 96.5 : ERR_ALTO_USO_CPU (14/0) [7/1]

```

Fonte: Weka (2025)

Matriz de Confusão: A matriz de confusão do *REPTree* mostrou um desempenho sólido em todas as classes. Os erros de classificação foram mínimos, com a maioria dos acertos concentrados na diagonal principal. A precisão na identificação de falhas como *ERR_AUMENTO_CRIT_RAM* e *ERR_ALTO_USO_RAM* foi particularmente impressionante, com um bom balanço entre acertos e erros.

Figura 6 – Matriz Confusão REPTree

```

=== Confusion Matrix ===

  a   b   c   d   e   f   <-- classified as
137   0   0   1   0   0   a = NORMAL
  2  14   2   0   0   0   b = ERR_AUMENTO_CRIT_RAM
  0   0  21   0   0   0   c = ERR_ALTO_USO_RAM
  2   0   0  16   1   0   d = ERR_AUMENTO_CRIT_CPU
  0   0   0   0  20   0   e = ERR_ALTO_USO_CPU
  0   0   0   1   0  20   f = ERR_SEM_CONEXAO

```

Fonte: Weka (2025)

4 Resultados de Análise e Projeto

Esta seção apresenta os resultados da pesquisa, detalhando a análise do conjunto de dados e o desempenho dos modelos de ML. Os resultados validam a viabilidade do uso de algoritmos de aprendizado de máquina na predição de falhas em data centers, oferecendo uma abordagem proativa e precisa.

4.1. Análise Exploratória e Pré-processamento dos Dados

O conjunto de dados inicial, com 300 registros de cinco máquinas virtuais, foi submetido a uma análise exploratória para identificar a distribuição dos eventos. A coleta de dados incluiu métricas como uso de RAM, CPU e Tráfego, além de um status do evento, que foram categorizados em: ERR_AUMENTO_CRIT_RAM, ERR_ALTO_USO_RAM, NORMAL, ERR_AUMENTO_CRIT_CPU, ERR_ALTO_USO_CPU e ERR_SEM_CONEXAO.

Durante a etapa de pré-processamento, foram removidos registros duplicados e a classe VPS_DESLIGADO. Essa decisão foi tomada devido à natureza da classe, que apresentava um padrão de dados com todos os valores zerados, e à sua sub-representação nos dados, que poderia comprometer o desempenho dos modelos. O conjunto de dados final, após a limpeza, consistiu em 237 instâncias, conforme ilustrado na Figura 2.

4.2. Desempenho dos Modelos de Classificação

Para avaliar a capacidade preditiva do ML, três algoritmos de árvores de decisão foram treinados com o conjunto de dados limpo, e seus desempenhos comparados.

4.2.1. J48 (Árvore de Decisão)

O modelo J48 demonstrou um alto nível de eficácia, alcançando uma acurácia de 95,78%. Conforme a Figura 3, as regras de decisão geradas são lógicas e intuitivas, validando o modelo. A regra principal utilizou a RAM como o atributo de maior importância, seguida pelo Tráfego e CPU, o que confirma a correlação entre o

aumento do uso de recursos e a ocorrência de falhas. A Matriz de Confusão (Figura 4) revela a solidez do modelo, com poucos erros de classificação e um desempenho forte nas classes de falha mais críticas.

4.2.2. *Hoeffding Tree* (Árvore de Decisão para Fluxos de Dados)

O *Hoeffding Tree* apresentou uma acurácia de 94,93%, um resultado expressivo. No entanto, sua estrutura de modelo é baseada em uma única regra de classificação, o que compromete a interpretabilidade das suas decisões. Por esse motivo, este algoritmo se mostrou menos adequado para o objetivo do projeto, que busca não apenas prever, mas também fornecer uma lógica clara para a tomada de decisões de diagnóstico.

4.2.3. *REPTree* (Árvore de Regressão e Classificação)

O algoritmo *REPTree* obteve a maior acurácia, com 96,20%, estabelecendo-se como o modelo de melhor desempenho geral. Conforme a Figura 5, sua árvore de decisão é mais complexa, ramificando a partir do uso de CPU. A Matriz de Confusão (Figura 6) confirma que o *REPTree* apresentou os menores índices de erros de classificação, demonstrando sua capacidade superior para se ajustar aos padrões dos dados.

4.3. Conclusão da Análise de Resultados

A análise comparativa comprova a viabilidade de aplicar o ML para a previsão de falhas em data centers. A fase de pré-processamento, embora tenha reduzido o volume de dados, foi fundamental para garantir a qualidade do treinamento, resultando em modelos robustos e precisos.

Para facilitar a comparação direta entre os algoritmos testados, os resultados e características de cada modelo são resumidos no Quadro 1.

Quadro 1 – Comparativo de Desempenho e Características dos Algoritmos

Algoritmo	Acurácia	Interpretabilidade	Pontos Fortes e Fracos
J48	95,78%	Alta	Ponto Forte: Regras lógicas e de fácil entendimento, ideal para validação por especialistas. Ponto Fraco: Desempenho ligeiramente inferior ao REPTree.
Hoeffding Tree	94,93%	Baixa	Ponto Forte: Processamento otimizado para grandes volumes de dados (fluxos). Ponto Fraco: Modelo "caixa-preta", com baixa interpretabilidade das regras.
REPTree	96,20%	Média	Ponto Forte: Maior acurácia e excelente desempenho geral na classificação. Ponto Fraco: Árvore de regras mais complexa, exigindo mais esforço para interpretação.

Fonte: Os autores (2025)

Entre os algoritmos testados, o *REPTree* se destacou como a melhor opção, alcançando a maior acurácia. Por sua vez, o J48 também se mostrou extremamente eficaz, oferecendo a vantagem de ter regras mais simples e interpretáveis, o que facilita sua adoção em ambientes de produção. O estudo demonstra que o ML é uma ferramenta poderosa para transformar a gestão de infraestruturas de TI, permitindo uma abordagem mais proativa e eficiente.

5. Considerações Finais

Este trabalho teve como objetivo aplicar e avaliar técnicas de ML para previsão de falhas em infraestruturas de data centers, buscando transformar a gestão de TI de um modelo reativo para uma abordagem proativa. A partir da coleta de dados em ambiente simulado, com 300 registros iniciais e 237 instâncias após o pré-processamento, foi possível treinar e comparar três algoritmos de árvores de decisão: *J48*, *REPTree* e *Hoeffding Tree*.

Os resultados obtidos confirmaram a viabilidade da proposta. O algoritmo *REPTree* apresentou a maior acurácia (96,20%), destacando-se como o modelo mais eficaz para identificar padrões associados a falhas. O J48, com acurácia próxima (95,78%), demonstrou robustez e, principalmente, interpretabilidade, característica essencial para a adoção em ambientes reais. Já o *Hoeffding Tree*, embora tenha alcançado 94,93% de acerto, mostrou-se menos adequado pela dificuldade de interpretação de suas regras.

A análise comparativa evidenciou que a qualidade dos dados e o pré-processamento são fatores críticos para o desempenho dos modelos. Além disso, verificou-se que diferentes algoritmos podem apresentar níveis distintos de precisão e interpretabilidade, sendo importante equilibrar esses aspectos conforme a aplicação prática.

Essa abordagem exploratória demonstra um grande potencial para o setor, fornecendo aos *data centers* uma ferramenta estratégica para a tomada de decisões proativas, que pode aprimorar a resiliência da infraestrutura e a eficiência na alocação de recursos.

Dessa forma, conclui-se que o uso de ML representa uma solução promissora para a manutenção preditiva em data centers, possibilitando maior resiliência, eficiência operacional e redução de custos relacionados a falhas inesperadas. Embora os testes tenham sido realizados em ambiente simulado, os resultados fornecem uma base sólida para futuras implementações em cenários reais, com potencial de integração a sistemas de monitoramento corporativos.

Como limitações, destaca-se o tamanho reduzido da base de dados e a ausência de eventos coletados em operações reais de larga escala. Para trabalhos futuros, recomenda-se expandir o volume de dados, explorar outras técnicas de ML como *Random Forests* e redes neurais profundas e avaliar a aplicabilidade em ambientes produtivos. Essas etapas poderão consolidar ainda mais o papel da inteligência artificial como ferramenta estratégica para a gestão moderna de infraestruturas críticas de TI.

Referências

Chalapathy, Raghavendra; Chawla, Sanjay. *Deep Learning for Anomaly Detection: A Survey*. Preprint, 2019.

Utilizado para fundamentar o conceito de detecção de anomalias com *deep learning*, que serviu de base para discutir técnicas modernas de previsão de falhas em data centers.

Chen, J. et al. *Machine Learning for Predictive Maintenance in Data Centers*. *Journal of Artificial Intelligence Research*, v. 55, p. 1–30, 2019.

Referência central para o embasamento teórico sobre manutenção preditiva com ML, reforçando a importância do pré-processamento e da qualidade dos dados.

Evans, Richard; Gao, Jeff. *DeepMind AI Reduces Google Data Centre Cooling Bill by 40%*. DeepMind Blog, 2020.

Citado como exemplo prático do uso de ML na otimização de energia em data centers reais, mostrando o impacto da inteligência artificial em operações corporativas.

Fernandes, Ricardo; Lima, Juliana. *Aplicações de Machine Learning em Data Centers*. Revista Brasileira de Computação Aplicada, 2018.

Forneceu o contexto nacional da aplicação de ML em ambientes de data center, conectando a pesquisa a iniciativas brasileiras na área.

Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron. *Deep Learning*. Cambridge: MIT Press, 2016.

Base teórica para compreender o funcionamento de redes neurais profundas, utilizada na explicação dos fundamentos do aprendizado profundo.

Homem, William Ludovico. *Apostila de Machine Learning*. PET Engenharia Mecânica, UFES, 2020.

Utilizado para ilustrar, de forma didática, o processo de aprendizado de máquina, com a analogia do aprendizado infantil citada no referencial teórico.

Lecun, Yann; Bengio, Yoshua; Hinton, Geoffrey. *Deep Learning*. *Nature*, v. 521, p. 436–444, 2015.

Referência que consolidou o embasamento sobre avanços em aprendizado profundo, reforçando sua relevância em aplicações modernas de IA.

McKinney, Wes. *Python for Data Analysis*. 2. ed. Sebastopol: O'Reilly Media, 2017.

Serviu de suporte técnico para o processamento e manipulação de dados, alinhando as boas práticas de análise utilizadas durante os testes no Weka.

Mitchell, Tom M. *Machine Learning*. New York: McGraw-Hill, 1997.

Fundamentou o conceito de aprendizado supervisionado, essencial para a explicação dos algoritmos de classificação empregados na pesquisa.

Murphy, Kevin P. *Machine Learning: A Probabilistic Perspective*. Cambridge: MIT Press, 2012.

Complementou a base teórica de aprendizado supervisionado, apoiando a justificativa do uso de modelos probabilísticos de decisão.

Russel, Stuart; Norvig, Peter. *Artificial Intelligence: A Modern Approach*. 3. ed. Upper Saddle River: Pearson, 2016.

Utilizado para sustentar a discussão sobre automação inteligente e o papel da IA na tomada de decisão em ambientes críticos.

Santos, Reginaldo Hugo Szezupior dos. Modelo de gestão de predição de falhas no gerenciamento da infraestrutura de datacenter. 2014.

Referência prática brasileira usada para contextualizar modelos de predição de falhas em infraestruturas de data center.

Silva, João; Costa, Maria. Monitoramento e Prevenção de Falhas em Data Centers: Abordagens Tradicionais. *Revista Brasileira de Tecnologia da Informação*, 2017.

Base para o comparativo entre modelos tradicionais e preditivos de monitoramento, explorado na seção 2.2 do trabalho.

Tanenbaum, Andrew S.; Steen, Maarten van. *Distributed Systems: Principles and Paradigms*. 2. ed. Upper Saddle River: Pearson, 2007.

Fundamentou a parte teórica sobre resiliência e disponibilidade de sistemas distribuídos, conectando-se à motivação do estudo.

Zhang, L. et al. *Predictive Maintenance for Data Centers: A Machine Learning Approach*. *ACM Computing Surveys*, 2018.

Utilizado para validar a eficácia de algoritmos de ML na previsão de falhas, inspirando a metodologia aplicada com os modelos J48, REPTree e Hoeffding Tree.