

ANÁLISE DE TENDÊNCIA EM ELEVATÓRIAS PELO APRENDIZADO DE MÁQUINAS

Israel Schmittel Spirito¹,
Thamirys Evellyn Rodrigues Neto da Silva¹,
Kevyn Phillipe Gusmão²,
Mirella Gonçalves da Fonseca Miranda da Silva²

¹ Discentes do curso de Engenharia Elétrica do Centro Unversitário Multivix Vitória

² Docentes do Centro Unversitário Multivix Vitória

RESUMO

O gerenciamento, tratamento e controle dos efluentes líquidos tem por finalidade avaliar a origem da sua produção, minimizar os impactos ambientais, realizar a caracterização e supervisão desses resíduos, cumprir com as obrigações legais e avaliar opções para sua administração. O bombeamento é realizado por meio das estações elevatórias de esgoto (EEEs), cujos processos podem ser comprometidos, seja no seu funcionamento ou monitoramento, incorrendo em problemas no desempenho de suas funções. Neste contexto, por meio da análise de tendência em elevatórias pelo aprendizado de máquinas, este estudo tem por objetivo principal criar um modelo preditivo, antecipar se o nível de uma elevatória irá subir ou descer, e identificar potenciais falhas no sistema, permitindo assim uma intervenção precoce e minimizando os riscos. Buscou-se descrever as EEEs, seus componentes, o controle e a operação de funcionamento, os sistemas de controle e os métodos de aprendizado de máquina. A metodologia empregada compreendeu a coleta e tratamento dos dados das elevatórias, seguida da construção e avaliação do modelo de Árvore de decisão. Os resultados de 90% de acurácia, confirmaram a viabilidade de prever com confiabilidade as variações nos níveis das elevatórias e identificar falhas iminentes. A análise valida a robustez do aprendizado de máquinas nesse cenário e indica um caminho inovador para a otimização da gestão desses sistemas.

PALAVRAS-CHAVE

Elevatórias; Aprendizado de Máquinas; Árvore de Decisão; previsão.

ABSTRACT

The management, treatment, and control of liquid effluents aim to evaluate their production origin, minimize environmental impacts, characterize and monitor these wastes, comply with legal obligations, and assess options for their management. Pumping is carried out through sewage pumping stations (SPS), whose processes can be compromised either in their operation or monitoring, leading to performance issues. In this context, through trend analysis in pumping stations using machine learning, this study's main objective is to create a predictive model to anticipate whether a station's level will rise or fall, and to identify potential system failures, enabling early intervention and minimizing risks. The study sought to describe the SPS, their components, control and operation, control systems, and machine learning methods. The methodology employed included data collection and processing from the pumping stations, followed by the construction and evaluation of a Decision Tree model. The results, with 90% accuracy, confirmed the feasibility of reliably predicting variations in station levels and identifying imminent failures. The analysis validates the robustness of machine learning in this scenario and suggests an innovative path for optimizing the management of these systems.

KEYWORDS

Pumping stations; Machine learning; Decision Tree; Prediction.

INTRODUÇÃO

O propósito do gerenciamento e controle dos efluentes líquidos é avaliar a origem da sua produção, diminuir os impactos no ecossistema, realizar a caracterização e supervisão desses resíduos, cumprir com as obrigações legais e avaliar opções para sua administração. A gestão é crucial para a segurança do sistema, assegurando que os efluentes sanitários sejam bombeados adequadamente e tratados de forma eficiente (AQUINO, 2018).

O mecanismo de bombeamento é desempenhado pelas estações elevatórias de esgoto (EEE), também conhecidas como estações de bombeamento, as quais assumem a responsabilidade de impulsionar os esgotos e os resíduos para áreas de elevação superior, viabilizando o transporte eficiente desses detritos dentro da infraestrutura de saneamento básico. Ou seja, impulsionam líquidos ou materiais em um sistema de dutos. Essas instalações realizam uma função vital em sistemas de fornecimento de água, distribuição de água potável, irrigação agrícola, sistemas de drenagem e diversas outras aplicações (TSUTIYA; ALEM SOBRINHO, 2011).

A operação desenvolvida pelas EEEs segue uma sequência de etapas, que incluem a captação dos efluentes em áreas mais baixas, que ficam retidas temporariamente e, posteriormente, são encaminhadas para as regiões mais elevadas com a finalidade de serem devidamente tratadas. Entretanto, as estações de bombeamento podem sofrer falhas nos seus processos, seja no seu funcionamento como defeito nas bombas ou em seu monitoramento como perda de comunicação ou até mesmo configuração incorreta (MACHADO, 2020).

Dentre essas intercorrências, podem estar as falhas mecânicas e elétricas, obstrução nas bombas, danos à tubulação, vazamentos, o que pode acarretar a interrupção do funcionamento. Por conseguinte, esses problemas podem levar a potenciais perigos à segurança, danos aos dispositivos, eficiência operacional comprometida e escassez de informações para fins de análise, com efeitos adversos para o meio ambiente e à saúde pública (MARTINS, 2014).

Na ausência de monitoramento apropriado, quaisquer falhas ou irregularidades nos componentes das EEEs podem passar despercebidas. Isso resulta em um aumento do risco de ocorrência de incidentes. Sendo necessária a constante supervisão do desempenho do sistema. Isso viabiliza a identificação antecipada de possíveis disfunções, irregularidades ou desvios indesejados, viabilizando uma

resposta rápida e eficaz para evitar prejuízos ou interrupções no sistema. Além disso, possibilita a detecção de padrões de desempenho ao longo do tempo, o que auxilia na otimização das operações, na execução de manutenção preventiva e na tomada de decisões informadas por dados concretos (COSTA FILHO, 2014).

O monitoramento é importante para assegurar a segurança, a eficiência e a confiabilidade dos processos e sistemas, o que, por sua vez, proporciona uma administração mais eficaz e reduz os custos operacionais. Além disso, tem-se a automação, que surgiu como ferramenta para aprimorar a segurança operacional e prevenir futuros contratemplos no sistema, fornecendo dados em tempo real da estação elevatória sanitária buscando a detecção, correção e monitoramento remotos, com ênfase na rapidez e precisão (BRK, 2023). As variáveis de controle em sistemas de bombeamento sanitário são os elementos que podem ser supervisionados e regulados com o objetivo de assegurar o correto funcionamento do sistema de bombeamento (MACHADO, 2020).

Nesse cenário, o foco desta pesquisa foi analisar o sistema de elevação de efluentes sanitários de uma empresa localizada em Vitória. O objetivo geral do estudo foi desenvolver uma proposta de sistema que possa antecipar as falhas das EEEs, com o propósito de ajudar no planejamento, manutenção e operação. Para validar o sistema proposto, buscou-se analisar suas variáveis de funcionamento; avaliar a utilização de linguagem de máquinas que possam prever as falhas; e determinar indicadores de desempenho.

O gerenciamento dos sistemas de monitoramento das EEEs é um aspecto crítico para assegurar a operação segura e eficaz dessas instalações. Neste contexto, essa pesquisa se justifica a partir da relevância do aprimoramento da eficiência das estações de bombeamento, mitigando falhas operacionais, elevando a confiabilidade e otimizando o desempenho do sistema como um todo. Tudo isso tem por meta prevenir problemas e avarias na manutenção, conseqüentemente, minimizando o risco de contaminação.

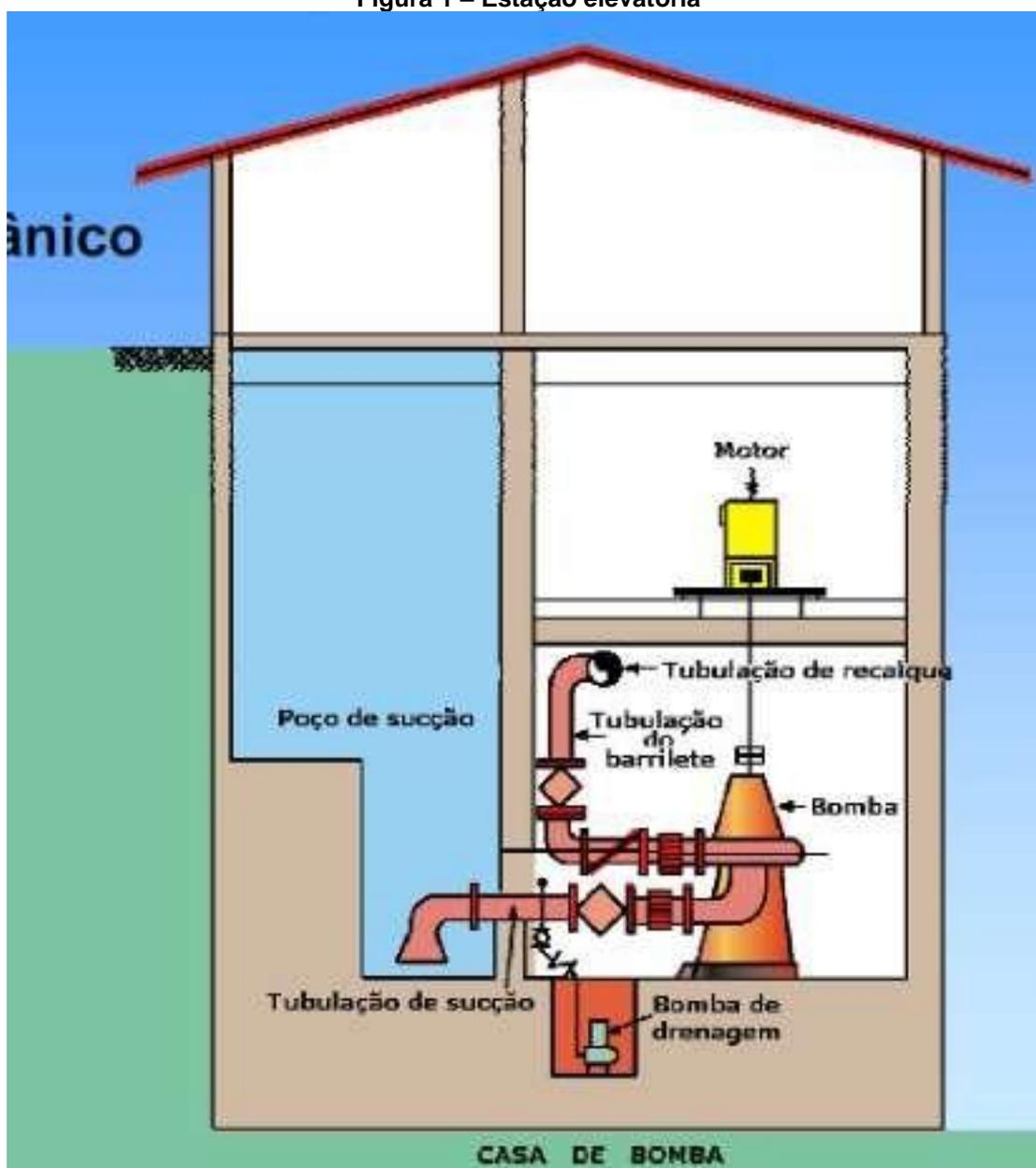
1. REFERENCIAL TEÓRICO

1.1 Estações Elevatórias

Também conhecida como estação de bombeamento, uma estação elevatória é um conjunto de instalações que engloba sistemas hidráulicos, mecânicos e elétricos

com finalidade de transporte e elevação de fluidos. Os principais componentes específicos incluem bomba, motor elétrico, tubulações tanto de sucção quanto de recalque, dispositivos auxiliares e acessórios conforme a imagem abaixo. As estações elevatórias têm aplicações fundamentais que envolvem: a captação de líquidos de fontes inferior ou de superfície com mesmo nível; o aumento da pressão nas redes para fornecer fluidos a pontos situados em níveis superiores ou a distâncias maiores, ou mesmo em ambas as situações simultaneamente; e, por fim, o aumento da vazão nas etapas de coleta e distribuição (SAMAE, s.d.).

Figura 1 – Estação elevatória



Fonte: Ministério das Cidades (2008).

1.2 Componentes das Estações Elevatórias

Os componentes de uma estação elevada desempenham funções específicas para garantir o funcionamento adequado do sistema de bombeamento e, assim, são de grande importância (MINISTÉRIO DAS CIDADES, 2008). O Quadro 1, abaixo, apresenta cada componente e sua função.

Quadro 1 – Componentes das estações elevatórias

Elemento		Definição	Função	Referência
Bomba		Equipamentos fluido mecânicos projetados para converter energia mecânica oriundas de uma fonte externa em energia aplicada a um fluido incompressível.	A principal função de uma bomba é transferir ou mover fluidos de um ponto a outro.	Netto; Fernandes (2015)
Motor		O motor elétrico é uma máquina projetada para converter energia elétrica em energia mecânica.	Amplamente utilizados para acionar máquinas, veículos e uma variedade de dispositivos que requerem movimento mecânico.	Petruzella (2013)
Poço		Estrutura de dimensões limitada que recebe as contribuições.	Armazenar o fluido de forma adequada até que ocorra o bombeamento.	SAMAE (s.d.)
Tubulação	Sucção	Refere-se ao trecho finalizado de um aqueduto, composto por tubos individuais ou uma passagem contínua, que pode ser fabricada no local.	Estabelecem a ligação entre o reservatório inferior ou poço de sucção e os conjuntos motor-bomba.	Tsutiya; Alem Sobrinho (2011)
	Recalque		Responsáveis por transportar o fluido pressurizado que foi impulsionado pela bomba para fora do sistema	

1.3 Controle e Operação de Funcionamento de uma Estação Elevatória

Nas instalações das EEEs, as bombas operam tipicamente de forma automatizada. Esse controle automático, em sua essência, depende da detecção da variação de nível do líquido no poço de sucção, o que representa a forma mais comum e simples de estabelecer uma relação entre a vazão de entrada e a vazão de saída do sistema, essa variação é identificada por meio de sensores de nível que são configurados para controlar os pontos de ligar e desligar as bombas (MINISTÉRIO DAS CIDADES, 2008).

O controle automatizado das bombas, em sua maioria, se fundamenta na mudança de nível do líquido presente no poço de sucção. Essa abordagem representa a maneira mais simples e convencional de estabelecer uma conexão entre a quantidade de líquido que entra e a quantidade de líquido que é bombeada de volta.

A detecção da variação de nível do líquido é realizada por meio de sensores de nível, os quais são configurados para controlar os diversos pontos de partida e parada das bombas, ou então para programar sua operação por intermédio de um painel de controle (ITATIAIUÇU, 2020).

Há a necessidade de verificação da ocorrência de vazamentos, bem como da avaliação do desempenho das válvulas e registro. Também, deve-se estar atento aos ruídos incomuns e vibrações que podem envolver motor e bomba. É, ainda, crucial que se realize as inspeções nas conexões elétricas e nas peças, a fim de evitar a ocorrência de oxidação e demais sinais que possam prejudicar o bom funcionamento do sistema (ITATIAIUÇU, 2020). Assim, para garantir que não haja danos, é importante que haja um sistema de controle capaz de assegurar que os resultados sejam mantidos da melhor forma.

1.4 Sistemas de Controle

Os sistemas de controle são concebidos com o propósito de assegurar que o sinal de saída de um sistema ou processo específico, sob uma entrada determinada, permaneça dentro de uma faixa desejada. Suas aplicações são vastas e abrangem setores que vão desde a indústria química até o controle de posicionamento de antenas e satélites, bem como sistemas de navegação, monitoramento e controle de níveis em reservatórios, entre outras aplicações (NISE, 2023). Sua composição é formada por sinal de entrada, sensor, controlador, processo e o sinal de saída

(PENEDO, 2014).

A entrada é a representação do estado desejado do sistema, é o valor que o sistema de controle busca alcançar. Por outro lado, a “saída” é uma variável que estamos controlando. Representa o resultado real do sistema após a intervenção do controlador. Através do sensor a variável é medida em tempo real fornecendo informações sobre o estado atual do sistema (BEUREN, I. M.; SANTOS, V. DOS; THEISS, V, 2018).

Já o controlador é o cérebro do sistema de controle. Ele obtém informações do sensor e, com base nas diferenças entre o estado real e o estado desejado (erro), gera um sinal de controle para influenciar o processo e minimizar o erro (KAVESKI; BEUREN, 2020).

Todas as sequências de mudanças graduais que ocorrem de forma estável, visando atingir um objetivo é chamado de processo, pode ser uma progressão natural ou uma série de ações controladas deliberadamente para alcançar resultados específicos (OGATA, 2010)

Existem dois tipos de sistema de controle, o de malha aberta onde não há realimentação ou feedback da saída para entrada, isso é, a saída é meramente observada a partir de suas amostras instantâneas, sem estabelecer qualquer conexão com o sinal de entrada (DORF; BISHOP, 2018).

Dorf e Bishop (2018) abordam o sistema de malha fechada, no qual existe realimentação, ou seja, a saída não é apenas monitorada com base em suas amostras instantâneas, mas também influencia a entrada do sistema por meio da criação de um sinal de erro. Para que este sinal seja processador, é frequentemente utilizado o CLP (Controlador Lógico Programável) para desempenhar o papel de um controlador, ajustando constantemente as saídas com base no erro recebido, para manter o sistema no estado desejado.

1.5 CLP

O CLP é um equipamento eletrônico com capacidade executória de instruções. Visa controlar e monitorar outros dispositivos, bem como maquinários industriais e processos em vários níveis de complexidade. É implementado um comando no dispositivo que é capaz de realizar funções lógicas e matemáticas, controle temporal, sequenciamentos (FRANCHI; CAMARGO, 2009). Apesar de existir tamanhos e complexidades diferentes, todos os sistemas de controle podem ser divididos em três

partes com funções bem definidas: atuadores, controladores e os transdutores (sensores).

Um transdutor é um dispositivo que transforma uma característica física do componente sensor, em um sinal elétrico. Esse sinal é então transmitido ao CLP, conectando-se às suas entradas. Já os atuadores têm o papel de converter o sinal elétrico vindo do CLP em uma condição física, geralmente ligando ou desligando algum dispositivo. E os controladores analisam a situação de entrada, conforme esse sinal opera um programa de controle para calcular as condições das suas saídas, esses sinais são convertidos no processo através dos atuadores. Além disso, o operador pode interagir com o controlador, ajustando as configurações de controle.

Alguns drivers também têm a capacidade de exibir informações sobre o estado do processo em um display ou tela. (CAMARGO, 2014).

1.6 Aprendizado de Máquina

O aprendizado de máquina, uma subárea proeminente da inteligência artificial, destaca-se por sua capacidade inerente de processar e analisar vastos volumes de dados complexos, também referidos como "*big data*". Conforme apontado por Mohammed e Al-Zawi (2017), existem algoritmos avançados e técnicas computacionais inovadoras que desempenham papéis cruciais nesta área, a tabela abaixo detalha as etapas envolvidas no processo de aprendizado de máquina, desde a definição do problema até o monitoramento e manutenção do modelo.

Quadro 2 – Etapas do Aprendizado de máquinas

Etapas	Descrição
Definição do Problema	Definir claramente o problema e determinar se o aprendizado de máquina é a solução.
Coleta de Dados	Realizar a coleta dos dados de fontes variadas, tais quais: arquivos, bancos de dados e sensores.
Pré-processamento de Dados	Fazer a limpeza, mutação e divisão dos dados em procedimentos de treinamento, validação e teste.

Escolha do Modelo	Selecione um modelo baseado na natureza do problema (ex.: classificação, regressão).
Treinamento	Desenvolver treinamento com a finalidade de mitigar os erros e ajustar os parâmetros do modelo.
Validação e Ajuste	Avaliar e refinar o desempenho do modelo, possivelmente ajustando hiperparâmetros.
Teste	Avaliar o modelo em um conjunto de teste com dados não vistos anteriormente.
Implementação e Produção	Implementar o modelo otimizado em um ambiente de produção.
Monitoramento e Manutenção	Monitorar o desempenho do modelo em produção e fazer ajustes conforme necessário.

Fonte: Adaptado de Breiman (2001).

1.7 Preparação de Dados

Antes de mergulhar nas técnicas de modelagem, é imprescindível garantir que os dados estejam bem estruturados e de alta qualidade, portanto diversas tratativas são necessárias nesse processo. A abordagem para lidar com essas lacunas varia conforme a natureza do problema e dos dados em questão, estratégias comuns englobam a remoção de registros com dados faltantes ou a imputação de valores usando média, mediana ou, até mesmo, técnicas mais avançadas, como a imputação baseada em modelos (BREIMAN, 2001).

Outro desafio são os Outliers, valores que se desviam significativamente da norma e podem distorcer análises. A identificação e, subsequente, transformação ou remoção desses valores é fundamental para garantir a integridade da análise da mesma forma a normalização e a padronização dos dados, quando se lida com diferentes variáveis em escalas distintas (SEWWANDI; LI; ZHANG, 2024).

Reduzir a dimensionalidade, usando técnicas como a Análise de Componentes Principais (PCA), pode otimizar o desempenho de algoritmos. Além disso, dividir adequadamente os dados em conjuntos de treinamento, teste e validação para uma avaliação justa e eficaz do modelo em questão (ALLEGRETTA, *et al.*, 2020).

1.8 Métodos de Aprendizagem de Máquinas

Dentre os métodos comumente empregados na aprendizagem de máquina, pode-se citar as Redes Neurais Convolucionais (CNNs), notórias por sua eficácia em tarefas de imagem e vídeo, também há as Redes Neurais Recorrentes (RNNs) e suas

variantes, como LSTM (Long Short-Term Memory) e GRU (Gated Recurrent Units), frequentemente empregadas em sequências temporais e linguagem natural.

Mohammed e Al-Zawi (2017) abordam sobre os Transformers, tais como BERT e GPT, representam uma arquitetura em ascensão, especialmente no campo do processamento de linguagem natural. Máquinas de Vetores de Suporte (SVM) são algoritmos de destaque para tarefas de classificação e regressão em espaços de alta dimensão, enquanto as Árvore de decisão (*Random Forests*) baseiam-se na construção de várias árvores de decisão durante o treinamento, resultando em uma classificação ou regressão agregada.

No entanto, o aprendizado de máquina vai além de apenas reconhecer padrões, ele contém o poder de extrair *insights* e informações profundas que, muitas vezes, não são identificáveis por meio de testes estatísticos convencionais, análises de correlação, diagramas e tabelas de contingência (LI, W.; ARUNDEL, S. T., 2022). É válido ressaltar que existem métodos tradicionais de análise, como o modelo *Box-Jenkins* ou ARIMA, que se baseia na ideia de que os valores observados em determinados períodos são funções lineares de valores anteriores, contudo, tais modelos requerem uma compreensão profunda dos dados e muitas vezes necessitam de suposições específicas. Isso pode torná-los desafiadores em determinados cenários e limitados em aspectos como a necessidade de coletar diversos parâmetros e realizar calibrações prévias (SANAL, M. G. et al., 2019).

2. METODOLOGIA

O estudo foi desenvolvido com base em dados de um sistema de estação elevatória de esgoto de uma empresa multinacional do ramo de mineração localizada no município de Vitória, há mais de 80 anos a mineradora chegou no estado, a princípio o seu sistema de esgoto era realizado por fossa, e em 2022 foi implantado o sistema de esgoto sanitário através de elevatórias. Essas elevatórias são divididas em quatro regiões, totalizando cinquenta e um elevatórias.

A divisão das elevatórias ocorre da seguinte maneira: há 13 elevatórias no setor administrativo, 11 elevatórias nas usinas de 1 a 4, 12 elevatórias nas usinas de 5 a 7, conforme imagem abaixo e 15 elevatórias no porto. Dado que se trata de um projeto recente dentro da empresa, a atenção do estudo se concentra na região das usinas de 5 a 7, isso ocorre porque esse projeto é pioneiro e oferece a maior probabilidade de obter os dados necessários para a pesquisa.

Figura 2 – Layout das elevatórias



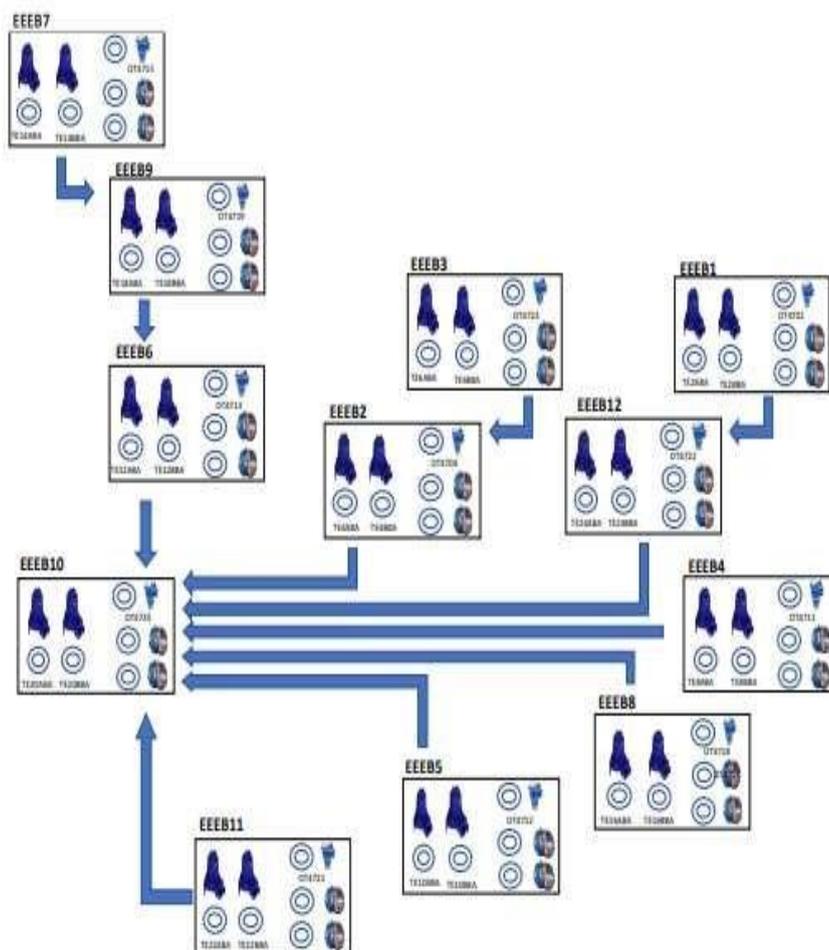
Fonte: Google Earth, adaptado pelo autor

O sistema projetado para esta área tem a função de realizar a coleta e o direcionamento de resíduos advindos de banheiros e vestiários para que seja feito o seu devido tratamento. Este sistema possui doze elevatórias sanitárias, com a capacidade de atender às necessidades de 2000 funcionários, aproximadamente. A organização na qual este sistema de elevatórias funciona, há, ainda, outras diferentes elevatórias que se conectam em série, conforme sua localização e sua capacidade de realizar a coleta.

A seguir apresenta-se o fluxograma (figura 3) para melhor compreensão, que basicamente se comporta da seguinte maneira: as elevatórias 7, 9 e 6 estão dispostas em série, com a elevatória 7 ocorrendo como montante e direcionando o rejeito para a elevatória 9, que por sua vez o encaminha para a elevatória 6 e, finalmente, para a elevatória 10, que se encontra a jusante.

Da mesma forma, as elevatórias 2 e 3 operam em série, com a 2 na posição montante, e as 1 e 12 encaminham seus resíduos para a elevatória 10, que está a jusante. Por fim, os elevadores 4, 5 e 8 têm a função de coletar e enviar seus resíduos diretamente para a elevatória 10.

Figura 3 – Fluxograma das elevatórias



Fonte: Acervo do Autor (2023)

2.1 Obtenção de Dados

A equipe de Gerência de Serviços Ambientais é responsável integralmente pela manutenção e monitoramento das elevatórias sanitárias da mineradora. O projeto em questão foi concluído no primeiro semestre de 2022, e passou por revisões na lógica de funcionamento para garantir um desempenho adequado. Devido a essas modificações, os dados disponibilizados abrangem o período de agosto de 2022 a agosto de 2023.

Esses dados foram adquiridos por meio do *software PI System*, que permite a coleta e análise detalhada dos dados, através do *software* pode-se obter os dados a cada segundo, porém foi liberado pelo gestor da área apenas os dados de 10 em 10 minutos. Para facilitar a obtenção de dados, o *software* conta com a extensão *PI DataLink* que é um suplemento do Microsoft Excel que concede a recuperação de dados do seu *PI Vision* para uma planilha.

É relevante destacar que os dados foram fornecidos pela equipe de engenharia e que todas as elevatórias seguem uma padronização, empregando bombas do modelo SULZER EJ40BX e sensores de nível VEGAPULS C 21, não contemplando monitoramento de pressão, perda de carga e vazão, sendo assim, não é possível realizar o monitoramento dessas variáveis.

2.2 Tratamento de Dados

Os conjuntos de dados fornecidos foram em formato CSV (*Comma-Separated Values*), que é um formato de arquivo usado para armazenar dados tabulares. Era esperado que os valores viessem ordenados por data e valores de 0 a 100 que indica o nível do poço da elevatória, assim foi conduzido uma limpeza dos dados utilizando os métodos de tratamento de valores faltantes, normalização ou padronização, detecção e tratamento de *outliers*.

Após o tratamento dos dados, identificou-se que as elevatórias 6, 7 e 9 não apenas apresentavam o maior volume de registros, mas também se destacavam em relação à qualidade dessas informações, pois havia mudanças frequentes de níveis, isto é, a confiabilidade e a precisão dos dados coletados dessas elevatórias foram superiores. Além disso, os dados coletados consistência com o restante do conjunto, diminuindo a possibilidade de discrepâncias e erros. Assim, justificando a escolha das elevatórias 6, 7 e 9 como o foco central deste projeto.

2.3 Aprendizagem de Máquina

O experimento utilizou técnicas robustas, tais como redes neurais e Árvore de Decisão. Estas técnicas têm a vantagem de não necessitar de múltiplas variáveis para funcionar de maneira eficaz, adequando-se perfeitamente ao cenário do projeto. Especificamente, a Árvore de Decisão é estruturada de forma hierárquica, assemelhando-se a uma árvore, cada nó corresponde a um teste ou questão sobre um atributo determinado, cada ramificação indica uma decisão tomada e cada folha simboliza um resultado ou conclusão.

Esta técnica é suportada por bibliotecas em *Python*, uma linguagem amplamente utilizada no aprendizado de máquina. Assim, com os recursos de *Python*, desenvolveu-se um *script* para implementar a Árvore de Decisão nos dados obtidos. Para os primeiros passos na codificação do modelo, optou-se pelo *Jupyter*, uma IDE (*Integrated Development Environment*) reconhecida por sua versatilidade. O

Jupyter facilita a importação de bibliotecas e proporciona um ambiente interativo para testes e avaliação da aplicabilidade da Árvore de Decisão ao conjunto de dados. Como visto previamente, identificou-se a influência direta das elevatórias 7 e 9 sobre a elevatória 6. Então pode-se ter como alvo a previsão do nível da elevatória 6 pois com registros históricos indicavam maior sensibilidade, tornando-a um ponto focal para previsões.

Para realizar a divisão dos dados em conjuntos de treino e teste, uma prática comum em aprendizado de máquina é destinar 70% para treino e 30% para teste. Durante esse processo, foi introduzida uma variável *booleana* denominada "*Target*", projetada para prever se os níveis subsequentes seriam superiores (1) ou inferiores (0) ao anterior. No processo de alimentação da Árvore de Decisão, os dados foram introduzidos sequencialmente ao algoritmo. Após treinamento com 70% dos dados, o modelo foi testado com os 30% restantes, visando verificar sua performance ao prever a variável *Target*.

2.4 Avaliação da Confiabilidade dos Resultados

A performance do modelo foi avaliada com base na sua acurácia. A acurácia é uma métrica utilizada em classificação e é usualmente utilizada para quantificar o desempenho do modelo. Ela representa a proporção de previsões corretas feitas pelo modelo em relação ao total, fornecendo uma ideia clara de quão frequentemente o modelo acerta em suas previsões, permitindo entender rapidamente a eficácia global do modelo em suas previsões, e matematicamente, a acurácia (Equação 1) é definida como:

Equação 1 – Formula de acurácia.

$$\text{Acurácia} = \frac{\text{Números de previsões corretas}}{\text{Números total de previsões}}$$

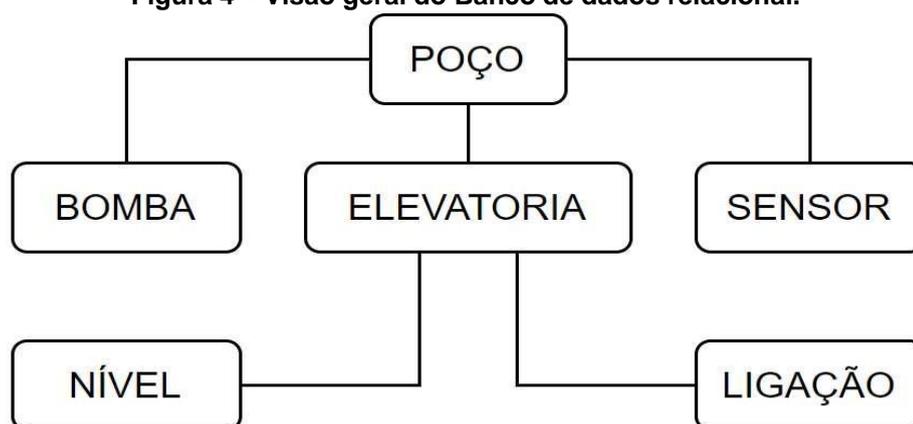
Fonte: Galera Monico, João Francisco et al., 2009.

Dentro dos padrões comuns de problemas de classificação, uma acurácia entre 80% e 90% é vista como um indicativo de um modelo de alta qualidade. Isso significa que, em 90% das vezes, o modelo indicaria corretamente se o nível da elevatória 6 estaria para aumentar ou diminuir na próxima leitura. Esta alta confiabilidade é crucial para fundamentar decisões futuras.

Com o modelo devidamente validado, avançamos para a apresentação dos dados por meio de uma interface *web*. Embora *Python* ofereça capacidades para desenvolvimento *web*, foi optado por utilizar PHP (*Hypertext Preprocessor*), uma linguagem de *script server-side* amplamente utilizada para desenvolvimento *web* e sendo mais amigável para iniciantes. Para o armazenamento e gestão de dados, *SQL Server*, um sistema de gerenciamento de banco de dados relacional, fornecendo uma estrutura robusta e otimizada para consulta e manipulação de dados.

Para a aplicação efetiva do modelo, é vital integrar o *Python*, PHP e *SQL Server*, portanto, a criação de um banco de dados relacional no *SQL Server* como mostra a figura 4, que serve como elo entre o PHP e o *Python*. Em PHP, é possível inserir novos dados manualmente ou via API (*Application Programming Interface*), que são armazenados no *SQL Server*. Já a execução do modelo em *Python* é agendada, podendo ser disparada manualmente ou através de gatilhos de execução predefinidos.

Figura 4 – Visão geral do Banco de dados relacional.



Fonte: Autor, 2023.

Uma vez processados os dados pelo modelo, o PHP assume o controle, possibilitando ações como, enviar notificações via e-mail/SMS sobre potenciais falhas, instruir o sistema de controle para acionar bombas, realizar testes de acionamento (conhecidos como testes de degrau), verificar a integridade dos sensores de nível,

entre outras funcionalidades.

3. RESULTADOS E DISCUSSÕES

A relevância deste estudo reside primordialmente na otimização do gerenciamento das estações elevatórias de saneamento. A antecipação de falhas, possibilitada pelo uso de técnicas avançadas como o aprendizado de máquina, não só garante a eficiência operacional, evitando paradas inesperadas, como também protege recursos hídricos e a saúde pública. Cada falha não detectada em uma estação pode resultar em desastres ambientais e em potenciais riscos à comunidade.

Além disso, o estudo traz inovações ao campo do saneamento aplicando a Árvore de Decisão, evidenciada como técnica moderna, podendo ser adaptada a setores tradicionais, potencialmente abrindo portas para mais integrações tecnológicas no futuro. Do ponto de vista da gestão, a capacidade de prever falhas transforma a maneira como os recursos são alocados e como as decisões são tomadas, levando a um sistema de saneamento mais robusto, confiável e sustentável, além de fortalecer a confiança do público e das partes interessadas no sistema de saneamento.

O cenário ideal para o monitoramento é a obtenção de todos os dados no intervalo de 0 a 100 e do tipo real. Entretanto, o pré-processamento revelou que este não era o caso pois havia inconsistências nos dados coletados o que mostrou uma heterogênea dos dados, como dados do tipo *string*, valores fora do range, indicando que o monitoramento estava comprometido. Essas inconsistências não só afetam a análise e a tomada de decisões, mas também sugerem potenciais problemas no sistema de coleta ou nos próprios equipamentos de monitoramento como mostra o Quadro 3.

Quadro 3 – Valores disponibilizados

Erro Detectado	Manifestação	Possíveis Causas
Fora do intervalo	Valores como “99999”	Falha do sensor ou erro no sistema de coleta.
Sequências anormais	Múltiplos valores de “100”	Falha do sensor, extravasamento excessivo ou erro no sistema.
Erro de conexão	“Scan Timeout”	Problemas de rede ou falhas de <i>hardware</i> .
Erro Genérico	“Bad”	Falha genérica no sistema ou no sensor.

Fonte: Produzido pelo autor (2023)

Dos 630.744 registros disponíveis, apenas 425.011 registros, equivale a cerca de 67,29% do conjunto de dados, são úteis para análise. No entanto, é fundamental observar que esses dados devem pertencer ao mesmo dia e horário para garantir a confiabilidade de relação, enfatizando a importância não apenas da quantidade, mas também da qualidade.

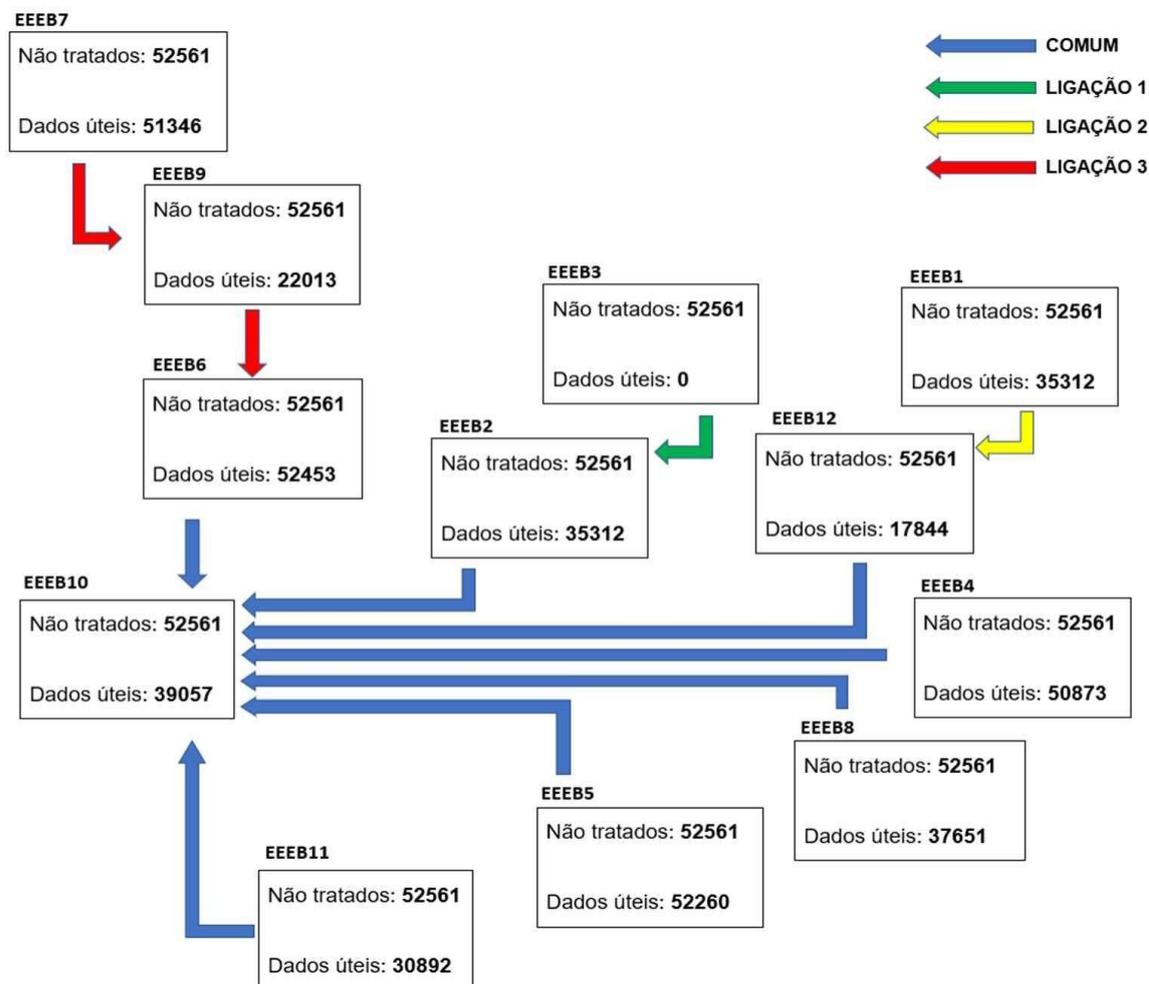
Quando se detecta de forma precoce os dados inconsistentes, é possível evitar a tomada de decisões erradas, assegurando que o sistema de monitoramento trabalhe de maneira eficaz, o que promove a confiabilidade. Dessa maneira, observa-se que é necessário que a informação seja enviada aos envolvidos a fim de que a correção dos possíveis erros seja rapidamente implementada.

Adicionalmente, a presença destas inconsistências por longos períodos, pode sinalizar a ausência de um plano de manutenção ou de um Plano de Ação Emergencial (PAE). Essas falhas não apenas comprometem a integridade do sistema, mas também diminuem significativamente a confiabilidade dele, no contexto do saneamento, a perda de funcionalidade de uma estação elevatória pode resultar em extravasamentos de esgoto, levando à contaminação do solo.

Ao direcionar o foco de estudo para elevatórias específicas, foi levado em consideração não apenas a qualidade dos dados disponíveis, mas também a interconexão e a comunicação entre as elevatórias.

Esta interconexão influencia diretamente o fluxo e a operação de cada elevatória. Por exemplo, a elevatória 10, que recebe fluxos de todas as outras elevatórias, torna-se inviável para previsão devido à sua natureza centralizada e ao alto grau de dependência com as outras, como evidencia a figura 5.

Figura 5 – Dados disponibilizados
DADOS DISPONIBILIZADOS



Fonte: Autor, 2023.

Ao avaliar os conjuntos de dados de cada ligação, observou-se que LIGAÇÃO 1 não continha conjunto de dados para serem utilizados como teste, pois todos os dados da elevatória 3 possuíam erros, assim, a LIGAÇÃO 3 possuía o maior volume de dados, oferecendo uma base mais robusta para análise e previsão, portanto, mais viável e promissora para o estudo.

Reconhece-se a relevância de distribuir de forma equilibrada os dados para treinamento e teste, assim, foi alocado 70% do conjunto total para treinamento e os 30% restantes para teste, utilizando a biblioteca *Sklearn.model_selection* como mostra a figura 6. Posteriormente, a avaliação da performance do modelo foi feita usando a função *score* da mesma biblioteca, permite calcular a acurácia.

Figura 6 – Código em Python

```
In [27]: # Aplicação do modelo
y = treino['target']
x = treino.drop('target',axis = 1)

from sklearn.model_selection import train_test_split
x_treino, x_teste, y_treino,y_teste = train_test_split(x,y,test_size =0.3)

from sklearn.ensemble import ExtraTreesClassifier

modelo = ExtraTreesClassifier()
modelo.fit(x_treino,y_treino)

resultado = modelo.score(x_teste,y_teste)
print ("Acurácia:", resultado) # Resultado da acurácia

Acurácia: 0.9001684274996172

In [11]: prevBASE # Dados para previsão

Out[11]:
      EEEB9  EEEB7  EEEB6
21770  35.06   23.6   59.8

In [29]: modelo.predict(prevBASE) # Entrada para previsão
print ("Resultado da previsão:", modelo.predict(prevBASE)) # Resultado previsão

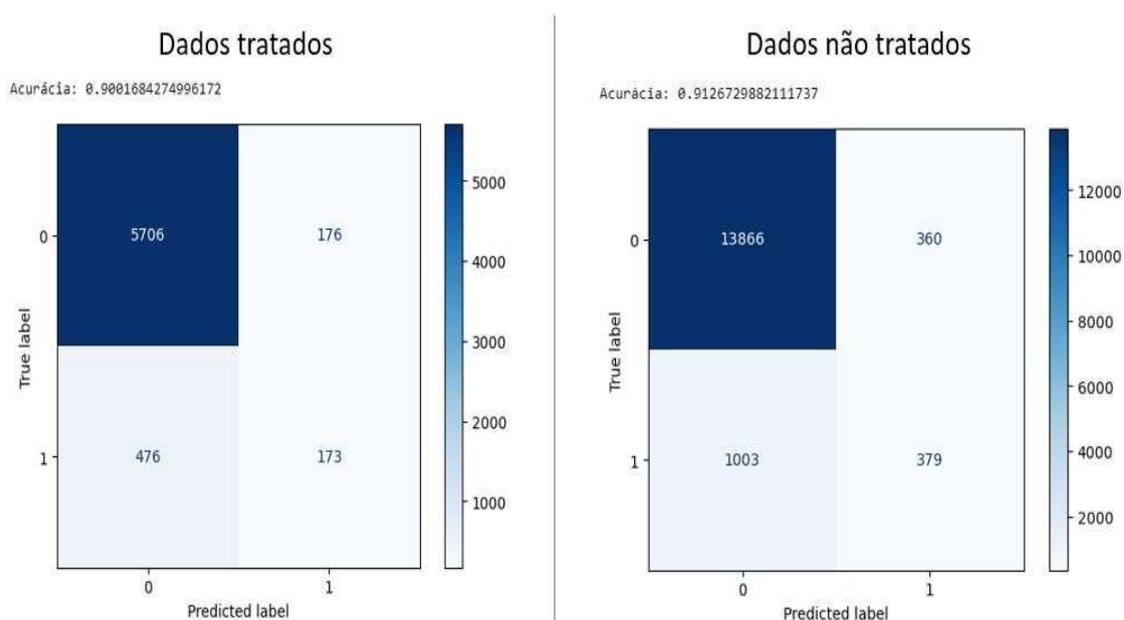
Resultado da previsão: [0.]
```

Fonte: Autor, 2023.

Outra amostragem foi a matriz de confusão, fundamental para uma interpretação mais granular dos resultados, indica que das 6.182 respostas categorizadas como 0, o modelo previu corretamente 5.706 e falhou em 467. Já para os resultados categorizados como 1, de um total de 349, o modelo acertou em 173.

Com base nesses resultados, alcançou-se uma acurácia de 90,01%. Uma visualização mais detalhada dessa performance pode ser encontrada na figura 7, ilustrando a habilidade do nosso modelo em fazer previsões de acurácia a partir dos dados fornecidos.

Figura 7 – Matriz de confusão



Fonte: Autor, 2023.

Da mesma maneira, realizou-se a aplicação do mesmo modelo nos dados não tratados tendo um comparativo de acurácia e matriz de confusão como indica a figura 7 sendo 14.869 respostas categorizadas como 0; o modelo previu corretamente 13.866 e falhou em 1003. Já para os resultados categorizados como 1, de um total de 739, o modelo acertou em 379 previsões, o que alcançou alcançamos uma acurácia de 91,26%. A figura 5 mostra a matriz de confusão de ambos os testes.

Observou-se que o modelo, ao receber os dados atuais, produz como saída os valores 0 ou 1. Estes indicam se a próxima leitura da elevatória 6 mostrará um nível superior, inferior ou igual ao atual, auxiliando assim na tomada de decisões. Com base nessa resposta, foi possível avaliar o nível subsequente e determinar as ações apropriadas a serem tomadas, como por exemplo, o acionamento da bomba na elevatória, visando prevenir extravasamentos na elevatória.

Esse modelo tem potencial para ser integrado a outros sistemas de gestão e controle. Em relação à tomada de controle do CLP, o sistema de gestão pode ser programado para disparar alertas via SMS ou e-mail quando um determinado nível é atingido na elevatória. Adicionalmente, foi viável realizar uma comparação dos dados com a elevatória 10, que se situa logo após a elevatória 6. Tal comparação é fundamental, pois o nível da elevatória 10 pode impactar diretamente o nível da elevatória 6, especialmente se a tubulação estiver completamente cheia, causando acúmulo na elevatória 6.

Dadas as características do modelo proposto, surgem limitações. Nos dados, 0e 100 contém várias repetições sequencialmente, e valores repetidos são marcados como 0 na variável *target*. Recomenda-se revisar o modelo e adicionar um terceiro nível na variável *Target* para valores inalterados, melhorando a precisão. O modelo analisa uma elevatória por vez, exigindo repetição para múltiplas elevatórias, o que pode ser ineficiente com sistemas simultâneos e também se baseia apenas no nível da elevatória.

É importante notar que o modelo se baseia exclusivamente no nível da elevatória para suas previsões. Entretanto, a complexidade inerente ao extravasamento sugere a necessidade de considerar variáveis adicionais, como pressão interna da elevatória, tempo de operação da bomba e periodicidade das amostras. Contudo, é essencial destacar que mesmo com a incorporação de variáveis adicionais, isso não necessariamente mitiga completamente as limitações do modelo. Quando se aborda

elevatórias desprovidas de ligações, o modelo demonstra limitações, uma vez que, a ausência de variação nos dados de entrada dificulta a realização de previsões consistentes.

Contudo o tempo de resposta do modelo se dá pela quantidade de dados disponível para treinamento e teste, uma vez em que existe também as limitações de hardware, se torna uma limitação o domínio do modelo de aprendizagem de máquinas e a familiaridade com outras tecnologias empregadas no modelo apresentado e vale lembrar que deverá ser compatível com o CLP das elevatórias.

4. CONSIDERAÇÕES FINAIS

Conclui-se que o avanço da tecnologia no estudo das elevatórias oferece vantagens significativas em benefício da sociedade. A aplicação da aprendizagem de máquinas na previsão do próximo nível das elevatórias se destaca como um avanço notável demonstrando a viabilidade do uso da tecnologia para prevenir extravasamentos, o que tem benefícios diretos para a comunidade e o meio ambiente. A previsão precisa dos níveis das elevatórias promove a eficiência operacional, evitando transbordamentos e evitando o extravasamento.

Como destaque, verificou-se que o modelo de previsão apresentou um grau elevado de acurácia, sendo perfeitamente utilizável para obter dados confiáveis e tomar decisões assertivas. Como resultados, observou-se que, a partir do recebimento de dados atualizados, o modelo produz como saída valores 0 ou 1, sendo capaz de indicar se a leitura seguinte da elevatória irá apresentar um nível mais elevado, mais baixo ou igual ao atual, o que promove uma tomada de decisão mais assertiva.

O *Python* apresentou alto nível de confiabilidade no processamento dos dados e um tempo curto de resposta, mostrando sua capacidade análise e eficiência com as novas tecnologias de aprendizagem de máquinas, também pode-se notar a facilidade de empregar o modelo com tal linguagem de programação. Sabe-se que é cabível de melhoria o modelo proposto, para âmbito industrial e comercial, fazendo ele se tornar viável em campo.

Recomenda-se a integração de dados operacionais com a parte física do sistema de elevatórias para projetos futuros, visando aprimorar ainda mais a precisão do modelo de previsão, assim, possibilitando outras análises como, esquema de vazamento entre linhas, monitoramento e controle de tubulação com interconexão de

dados com análise em tempo real, tornando-se fundamentais em direção a um sistema mais robusto e confiável.

5. REFERÊNCIAS

ALLEGRETTA, Ignazio; MARANGONI, Bruno; MANZARI, Paola; PORFIDO, Carlo; TERZANO, Roberto; PASCALE, Olga De; SENESI, Giorgio S. Macro-classification of meteorites by portable energy dispersive X-ray fluorescence spectroscopy (pED-XRF), principal component analysis (PCA) and machine learning algorithms. **Talanta**, Volume 212, 2020. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S003991402030076X?via%3Di> hub. Acesso em 03 nov. 2023.

AQUINO, L. M. **Automação de estação elevatória de água tratada via módulo de rádio de baixo custo**. TCC (Curso de graduação em Engenharia Elétrica) - Campus de Sobral, Universidade Federal do Ceará, Sobral, 2018.

BEUREN, I. M.; SANTOS, V. DOS; THEISS, V. **Análise da Formalidade dos Sistemas de Controle Gerencial em Empresas Incubadas**. Revista Mineira de Contabilidade, v. 19, n. 2, p. 19–30, 21 ago. 2018. Disponível em: <https://doi.org/10.21714/2446-9114RMC2018v19n2t02>. Acesso em 03 nov. 2023.

BREIMAN, L. **Random forests**. **Machine Learning** 45, 5–32, 2001, Disponível em: <https://doi.org/10.1023/A:1010933404324>. Acesso em 03 nov. 2023.

BRK. **Automação aumenta segurança operacional de Estação Elevatória de Esgoto de Santa Gertrudes**. 2023. Disponível em: <https://www.brkambiental.com.br/santa-gertrudes/automacao-aumenta-seguranca-operacional-de-estacao-elevatoria-de-esgoto-de-santa-gertrudes>. Acesso em 17 jun. 2023.

BURKART, N., HUBER, M.F., **A survey on the explainability of supervised machine learning**. Journal of Artificial Intelligence Research, Vol. 70, 2021 Disponível em: <https://doi.org/10.1613/jair.1.12228>. Acesso em 03 nov. 2023.

CAMARGO, Valter Luís Arlindo de. **Elementos de Automação** Editora Saraiva, 2014. E-book. ISBN 9788536518411. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788536518411/>. Acesso em: 03 nov. 2023.

COSTA FILHO, Isaias de Almeida da, et al. (org.). **Noções sobre operação de estações elevatórias**. Natal: CAERN - Companhia de Águas e Esgotos do Rio Grande do Norte, out. 2014.

DORF, Richard C.; BISHOP, Robert H. **Sistemas de Controle Modernos**, 13ª edição. São Paulo: Grupo GEN, 2018. E-book. ISBN 9788521635147. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788521635147/>. Acesso em: 22 out. 2023.

ITATIAIUÇU. **Estação Elevatória de Esgoto Bruto (EEEB) SEDE**. Prefeitura Municipal de Itatiaiuçu. 2020. Disponível em: https://www.itatiaiuçu.mg.gov.br/abrir_arquivo.aspx/Concorrencia_13_2020_Anexo_

XDMemorial_Descritivo_EEBB_Sede_?cdLocal=3&arquivo=%7BD3CE7AAD-CEAA-E2C1-57AA-701AA0B1D111%7D.pdf&cdLicitacaoArquivo=63580.

KAVESKI, I. D. S.; BEUREN, I. M. Influência dos sistemas de controle gerencial e da criatividade sobre o desempenho no trabalho. *Cadernos EBAPE.BR*, v. 18, p. 543–556, 16 out.2020. Disponível em: <https://www.scielo.br/j/cebape/a/cXPVDQ7c7hYCssf7C5H9z9F/?format=pdf&lang=pt>
LI, W.; ARUNDEL, S. T. GeoAI and the Future of Spatial Analytics. **Springer**. p. 151–158, 1 jan. 2022. Disponível em: https://doi.org/10.1007/978-981-19-3816-0_17
Acesso em: 16 jun. 2023.

MACHADO, Jéssica Roberta Maciel. **Otimização da malha de controle de uma estação elevatória de esgoto**. Trabalho de Conclusão de Curso – Universidade de Brasília – UnB. Orientador: Dr. Luís Filomeno de Jesus Fernandes. Faculdade UnB Gama – FGA, 2020. Acesso em 03 nov. 2023.

MARTINS, Alaide Barbosa. **Desenvolvimento de metodologia para análise de risco com base no método CORAS e avaliação quantitativa para aplicação de plantas de saneamento**. Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Sistemas Eletrônicos. -- versão corr. -- São Paulo, 2014. Acesso em 03 nov. 2023.

MINISTÉRIO DAS CIDADES. **Esgotamento sanitário: operação e manutenção de estações elevatórias de esgotos: guia do profissional em treinamento: nível 2 / Ministério das Cidades. Secretaria Nacional de Saneamento Ambiental (org.). – Brasília: Ministério das Cidades, 2008.**

MOHAMMED, S.A.T.A.; AL-ZAWI, S. Understanding of a convolutional neural network. 2017 **International Conference on Engineering and Technology (ICET)**, Antalya, Turkey, 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186. Acesso em 03 nov. 2023.

MONICO, João Francisco Galera; PÓZ, Aluir Porfírio Dal; GALO, Maurício; SANTOS, Marcelo Carvalho dos; OLIVEIRA, Leonardo Castro de. **Acurácia e precisão: revendo os conceitos de forma acurada**. Boletim de ciências geodésicas, vol. 15, núm. 3, pp. 469-483. 2009, Disponível em: <http://www.redalyc.org/articulo.oa?id=393937709010>. Acesso em 03 nov. 2023.

NETTO, José Martiniano de A.; FERNÁNDEZ, Miguel Fernández Y. **Manual de hidráulica**. São Paulo: Editora Blucher, 2015. E-book. ISBN 9788521208891. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788521208891/>. Acesso em: 02 nov. 2023.

NISE, Norman S. **Engenharia de Sistemas de Controle**. São Paulo: Grupo GEN, 2023. E-book. ISBN 9788521638285. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788521638285/>. Acesso em: 22 out. 2023.

OGATA, Katsuhiko. **Engenharia de controle moderno**. Trad. Heloísa Coimbra de Souza; revisor técnico Eduardo Aoun Tannuri. -- 5. ed. -- São Paulo: Pearson Prentice Hall, 2010.

PENEDO, Sérgio Ricardo M. **Sistemas de controle - matemática aplicada a projetos**

- 1ª edição - 2014. Disponível em: Minha Biblioteca, Editora Saraiva, 2014.

PETRUZELLA, Frank D. **Motores elétricos e acionamentos**. (Tekne). São Paulo: Grupo A, 2013. E-book. ISBN 9788580552584. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788580552584/>. Acesso em: 02 nov. 2023.

SAMAE. Serviço Autônomo Municipal de Água e Esgoto. **Apostila operador de estação de bombeamento**. Caxias do Sul. Disponível em: <https://www.samaecaxias.com.br/Concurso/DownloadArquivoConcurso/Apostila%20Operador%20ETAE.pdf/>. Acesso em 03 nov. 2023.

SANAL, M. G. et al. **Artificial Intelligence and Deep Learning: The Future of Medicine and Medical Practice**. The Journal of the Association of Physicians of India, v. 67, n. 4, p. 71–73, 1 abr. 2019. Disponível em: <https://europepmc.org/article/med/31309802>. Acesso em: 16 jun. 2023.

SEWWANDI, M.A.N.D.; LI, Y.; ZHANG, J. k-outlier removal based on contextual label information and cluster purity for continuous data classification, Expert Systems with Applications. **Expert Systems with Applications**. Volume 237, Part C, 2024, Disponível em: https://www.sciencedirect.com/science/article/pii/S0957417423018493?ssrnid=4214220&dgcid=SSRN_redirect_SD. Acesso em 03 nov. 2023.

TSUTIYA, M. T.; ALEM SOBRINHO, P. **Coleta e Transporte de Esgoto Sanitário**. 3ª edição. Rio de Janeiro: ABES – Associação Brasileira de Engenharia Sanitária e Ambiental, 2011.