MODELO DE PREDIÇÃO DE INTERRUPÇÕES NA OPERAÇÃO DE TURBINAS DE USINAS HIDRELÉTRICAS BASEADO EM SEUS CICLOS DE CARGA: Uma Abordagem Por Redes Neurais

Daniel Farina Moraes (IC) e Prof. Arnaldo Rabello de Aguiar Vallim Filho (Orientador)

Apoio:PIBIC Mackenzie

RESUMO

Turbinas hidrelétricas estão propensas a sofrerem danos, causando falhas de equipamento e trazendo prejuízo para a empresa. E o modo em que a maioria das hidrelétricas lidam com esse problema é realizando manutenções periódicas. Desse modo há um desperdício de dinheiro por causa de manutenções desnecessárias, e ainda assim há paradas não programadas por causa de falhas nos equipamentos. O objetivo deste estudo é prever as falhas de equipamento para que o número de manutenções desnecessárias diminua e o número de manutenções necessárias aumente, de modo a reduzir os custos de manutenção, paradas operacionais e a quantidade de falhas. Neste projeto foi desenvolvido uma ferramenta de monitoramento de variáveis relevantes dos ciclos de carga (tempo de carga de jatos de água nas pás) dessas turbinas. Esses dados dos ciclos de carga foram utilizados junto com os dados de alarmes, que indicam quando ocorrem falhas de equipamento, para alimentar uma rede neural artificial capaz de prever em um período de 24 horas se ocorrerá uma falha de equipamento ou não. Os resultados dos experimentos desenvolvidos se mostraram promissores e com a rede neural artificial implantada na operação de uma usina hidrelétrica, trará uma redução de custos significativa.

Palavras-chave: Redes Neurais Artificiais. Manutenção Preditiva. Aprendizado de Máquina.

ABSTRACT

Hydroelectric turbines are always subject to damage, causing equipment failures and raising the company’s maintenance costs. The way that most hydroelectric power plants deal with this problem is to carry out periodic maintenance. However, in this way, there is a waste of money due to unnecessary maintenance, and even though they perform extra maintenance, sometimes failures still happen. The purpose of this study is to predict equipment failures so that the number of unnecessary maintenances decreases, and the number of necessary maintenances increases, in order to reduce costs, operational downtime and the number of future failures. In this project, a monitoring dashboard is developed to monitor relevant data of

the water jets cycles (duration of the water jets hitting the impulse blades) of those turbines. This data from the water jets cycles was used together with alarm data, which indicates when an equipment failure happens, to feed an artificial neural network capable of predicting, within a 24-hour period, whether an equipment failure will occur or not. The experiment results were promising and with the artificial neural network installed in a hydroelectric power plant's system, it will bring a significative cost reduction.

Keywords: Artificial Neural Networks. Predictive Maintenance. Machine Learning.

1. INTRODUÇÃO

Este projeto de iniciação cientifica fez parte de outro projeto maior denominado “Análise Preditiva Baseada em Inteligência Artificial para Sistemas Supervisórios de Usinas Hidrelétricas”, envolvendo análises estatísticas e Inteligência Artificial aplicadas à área de Energia, em particular, na operação de uma usina hidrelétrica. Esse estudo maior tem um conjunto amplo de objetivos, mas seu objetivo final é a construção de um modelo preditivo que seja capaz de estimar a probabilidade de paradas súbitas de turbinas hidrelétricas em um dado instante do tempo. Possibilitando, assim, com que as manutenções corretivas sejam reduzidas e que as manutenções preventivas obrigatórias sejam distribuídas de forma otimizada, reduzindo o número de paradas e assim reduzindo os custos gerados da operação.

No caso específico deste projeto de iniciação cientifica, o objetivo final é semelhante, mas os meios de se chegar ao objetivo serão outros. Este projeto tem o intuito de utilizar dados referentes aos sensores presentes nos equipamentos da usina relacionados aos ciclos de carga das turbinas. Um ciclo de carga é o tempo decorrido entre a abertura e fechamento de válvulas que liberam jatos d’água que atingem as pás da turbina, e foi com base em variáveis derivadas desses ciclos que se buscou prever ocorrências que levariam a paradas dos equipamentos.

1.1 Problema de Pesquisa

Para entender melhor a questão é necessário compreender um pouco sobre a função de uma turbina em uma usina hidrelétrica. Existe mais de um tipo de turbina hidrelétrica, a mais comum atualmente é a turbina Pelton, que é a turbina usada como referência para esse projeto. Em uma usina hidrelétrica, as turbinas são responsáveis pela transformação de energia cinética em energia elétrica. A energia cinética é nada mais do que a pressão com que a água atinge as pás da turbina, fazendo com que a turbina gire muito rápido. Com o giro da turbina, o eixo do gerador, que é acoplado, faz com que o gerador também gire, passando a gerar energia elétrica.

Durante o funcionamento de uma turbina, as pás por receberem frequentemente jatos de água, acabam se desgastando fazendo com que outros componentes auxiliares sejam danificados. Conforme colocado acima, um ciclo de carga é o tempo decorrido entre a abertura e fechamento de válvulas que liberam esses jatos d’água. Entende-se que quanto maior o volume de água e o tempo que a turbina fica submetida a esses jatos, maior o desgaste que ela sofrerá.

O dano a esses equipamentos faz com que uma parada para manutenção seja necessária em um dado momento do tempo. Se fosse possível prever o instante exato

em que uma manutenção é necessária, isso seria um instrumento muito importante para o planejamento e para a redução de perdas por paradas não programadas.

Assim, este projeto buscou desenvolver um modelo preditivo para prever essas paradas. Assim, a partir de dados reais de uma usina hidrelétrica, este projeto se propõe a estudar o comportamento das diversas variáveis presentes em um ciclo de carga de uma turbina, para que seja feita uma análise exploratória dos dados, identificando possíveis tendências para interrupções na operação dos equipamentos. Após identificar algumas tendências nos dados, um modelo preditivo pode ser proposto para estimar o momento em que uma manutenção seria necessária.

Dado esse quadro, a pergunta que se pode colocar nesta pesquisa seria:

"A partir de dados reais de uma usina hidrelétrica, seria possível se estabelecer correlações entre variáveis que caracterizam um ciclo de carga de uma turbina, com outras variáveis de outros componentes da usina, e com isto, possibilitar a construção de um modelo preditivo de paradas da turbina para manutenção?".

1.2 Objetivo

O objetivo deste projeto é confirmar se os dados dos sensores presentes nos equipamentos referentes aos ciclos de carga de uma turbina podem ser utilizados para prever uma falha equipamento. Para confirmar isso, foi desenvolvido um modelo preditivo que seja capaz de olhar para os dados em um determinado instante de tempo, e dizer se em até 24 horas haverá uma falha de equipamento ou não.

De modo mais específico, pretende-se:

- Identificar e relacionar ciclos de carga de uma turbina com interrupções dos equipamentos;
- Tentar identificar padrões quanto à quantidade e duração de ciclos de carga entre paradas das turbinas;
- Estudar o comportamento de variáveis físicas do sistema durante esses ciclos, por meio de análise exploratória, e relacioná-las a pontos notáveis desses ciclos: início, fim, média, quartis, entre outros;
- A partir dos resultados anteriores, desenvolver uma rede neural artificial que seja capaz de prever os momentos de ocorrências de falhas em uma turbina, associados a ciclos de carga.

2. REFERENCIAL TEÓRICO

Este capítulo tem como o objetivo apresentar uma revisão da literatura, envolvendo livros e artigos científicos, que tenham uma relação próxima com o tema deste projeto. As informações obtidas nessas obras auxiliaram na compreensão do tema e dos processos necessários para o desenvolvimento deste projeto.

2.1 Mineração de Dados

Um dos focos deste projeto é o uso de técnicas de Mineração de Dados (MD), assim, neste início faz-se uma breve revisão procurando-se dar uma visão geral deste tema.

De acordo com (SILVA et al., 2016), de forma simplificada, a mineração de da dos é um processo tanto automático quanto semiautomático para analisar de forma exploratória um conjunto de dados, de modo que possa se encontrar padrões relevantes sobre o conjunto, assim gerando conhecimento. Trata-se, portanto de aplicações de técnicas, implementadas por meio de algoritmos computacionais, capazes de receber dados referentes ao mundo real, e retornar um padrão de comportamento que pode ser útil para tomadas de decisões.

A MD engloba muitos métodos e ferramentas relacionadas a extração, processamento e apresentação de dados, como análise exploratória, análise preditiva, ETL (*Extraction, Transform and Loading*), estatística descritiva, entre outras. (SILVA et al., 2016)

Em MD busca-se fundamentalmente descobrir conhecimento em uma massa de dados. Essa descoberta de conhecimento em bases de dados possui algumas etapas, como a organização de um repositório único e o tratamento de “ruídos” no conjunto de dados, que são, entre outras coisas, amostras com informações faltando e amostras discrepantes; além disso, envolve ainda; a normalização dos valores, para que eles não entrem em desacordo e sejam igualmente considerados, Estas etapas são chamadas de pré-processamento.

Na sequência, tem-se a análise exploratória no conjunto de dados para que se compreenda o comportamento geral dos dados e, eventualmente, para se escolher as melhores amostras, as mais representativas e que fornecem uma identidade melhor do conjunto. Depois, inicia-se a análise preditiva, com tarefas como predição, agrupamento ou associação. Ao final do processo de MD, os resultados são validados, avaliados e geralmente apresentados em representações visuais de fácil entendimento. Os passos anteriores não necessariamente precisam ser executados

em uma sequência rígida, podendo ser realizados em ordens ligeiramente diferentes, dependendo da situação. (SILVA et al., 2016; DE CASTRO e Ferrari, 2016)

2.2 Pré-processamento de dados

Pré-processamento de dados foi uma das ferramentas principais utilizadas neste projeto. Segundo (HAN et al., 2011) os dados que encontramos hoje em dia são suscetíveis a dados inconsistentes e dados faltando, por causa do tamanho e da junção de dados de diferentes origens.

Existem diversas técnicas de pré-processamento de dados. *Data cleaning* é utilizada para remover inconsistência nos dados, como por exemplo eliminar linhas de um banco de dado nas quais estejam faltando um valor importante. *Data integration* junta dados de diferentes origens de uma maneira que os dados permaneçam coerentes. *Data reduction* reduz a quantidade de dados, podendo eliminar dados redundantes que não acrescentam muita ou até mesmo nenhuma informação. *Data transformation* transforma os dados para que eles possam ser mais coerentes tanto para quem irá utilizar os dados quanto para um algoritmo de aprendizado de máquina. A normalização dos dados é um exemplo de *Data transformation*, a normalização é utilizada para redimensionar os dados, por exemplo transformar um atributo representado em polegadas para um atributo representado em metros. Todas essas técnicas não são mutualmente exclusivas, elas podem ser utilizadas juntas. (HAN et al., 2011)

2.3 Análise Exploratória de Dados

Como dados são um dos recursos principais deste projeto, entender eles e tirar algum valor ou informação deles foram essenciais. Ai que entra a Análise Exploratória de Dados (AED). Não podemos mencionar AED sem citar John Tukey, popularmente conhecido como o fundador da AED. Uma das primeiras vezes que mencionou AED foi em sua obra *The Future of Data Analysis* (TUKEY, 1962), mas sua obra mais famosa foi *Exploratory Data Analysis* (TUKEY, 1977). Segundo (HOAGLIN, MOSTELLER e TUKEY, 1983; HOAGLIN, 2003), existem quatro temas principais em AED: resistência, resíduo, re-expressão e revelação.

- Resistencia: insensibilidade a *outliers*.
- Resíduo: diferença entre os dados reais e o modelo.
- Re-expressão: transformação dos dados para uma escala que facilite a análise.
- Revelação: mostrar o comportamento dos dados ou o resultado da análise, através de gráficos.

AED pode ser as vezes representada como uma caixa de ferramentas, mas esse aspecto não é sua essência, é apenas uma consequência da atitude exploratória assumida pelo analista de dados. AED não apresenta nenhuma restrição quanto aos procedimentos a serem usados e é comum que analistas de dados apareçam com novas maneiras de olhar para os dados. (MORGENTHALER, 2009)

2.3.1 Conceitos de Estatística Descritiva

Uma revisão de Estatística Descritiva foi feita neste projeto, pois esta deverá ser uma das ferramentas que serão empregadas para o desenvolvimento de Análises Exploratórias dos dados. A Estatística Descritiva costuma ser definida como sendo um conjunto de técnicas capazes de descrever e resumir dados, revelando aspectos importantes do conjunto de dados, como o tipo de distribuição e os valores mais representativos, permitindo uma melhor visualização do conjunto. (SILVA et al., 2016)

Em estatística descritiva, é possível calcular medidas de posição e separatrizes. As medidas de posição orientam a análise de dados quanto a sua localização e ajudam a identificar o comportamento da distribuição dos dados dentro do conjunto principal. Exemplos de medidas de posição, do tipo central, seriam, média, mediana e moda. Já as separatrizes subdividem o conjunto de dados em partes, não necessariamente iguais. Exemplos de separatrizes são medianas, quartis e percentis. (SILVA et al., 2016; DEVORE, 2014).

Outro tipo de medida fundamental em Análise Exploratória, são as medidas de dispersão. Estas medidas, como a variância, amplitude desvio padrão e coeficiente de dispersão, são utilizadas para calcular o quanto um valor está distante de uma medida central como por exemplo a medida de posição média. As medidas de dispersão servem assim, para descrever a variação e a dispersão dos valores de um conjunto de dados. (DEVORE, 2014)

2.3.2 Boxplots

Boxplot foi uma das ferramentas de visualização de AED utilizadas neste projeto. Segundo (WICKHAM e STRYJEWSKI, 2011), o *boxplot* é usado para mostrar um resumo da distribuição dos dados. Ele mostra menos detalhes que um histograma, mas ocupa menos espaço também. O *boxplot* utiliza estatísticas descritivas de dados reais e é rápido de se fazer, originalmente feito à mão, e não tem nenhum parâmetro ajustável. Ele é particularmente útil para comparar a

distribuição de diferentes grupos. Abaixo na figura 2-1, podemos ver a representação de um *boxplot*.

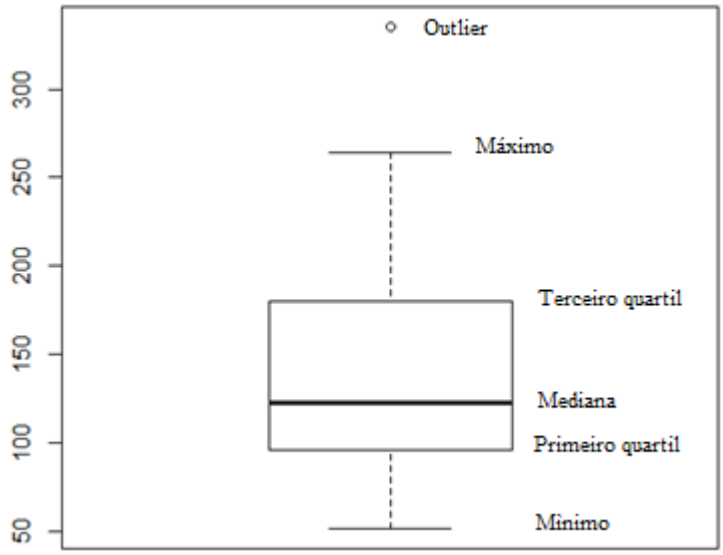

Figura 2-1: Exemplo de um *boxplot* comum.

De acordo com (MORGENTHALER, 2009; WICKHAM e STRYJEWSKI, 2011) e como se pode ver na figura 2-1, um *boxplot* é composto por esses seis valores; O mínimo é calculado através da seguinte formula onde Q1 e Q3 representam primeiro quartil e terceiro quartil respectivamente, Q1 - 1,5 * (Q3 – Q1); O máximo é calculado através desta outra formula, Q3 + 1,5*(Q3-Q1); O primeiro, segundo e terceiro quartil, sendo que o segundo quartil é chamado de mediana; E os *outliers* que são valores acima do Máximo e abaixo do Mínimo.

2.4 Análise Preditiva

Neste projeto é desenvolvido um modelo *Multilayer Perceptron* (MLP), que pode ser considerado uma ferramenta de Análise Preditiva.

Análise preditiva é um processo que permite descobrir se existe um relacionamento entre diferentes atributos em um conjunto de dados, e possivelmente um rotulo caso possam ser classificados. Um atributo do conjunto de dados pode assumir duas formas, uma classe ou um valor contínuo. Em geral, o relacionamento descoberto entre os atributos e classes são representados através de um modelo de predição no qual pode ser apresentado por funções ou organizado em estruturas de dados. O processo de construção do modelo preditivo se dá por meio da escolha e alterações de parâmetros utilizados no algoritmo. Quando esse algoritmo é um algoritmo de aprendizado de máquina, o processo é chamado de treinamento e após treinado, o algoritmo deverá ser capaz de predizer a classe ou um valor contínuo,

depende do atributo a se prever, para dados que não fizeram parte do conjunto de treinamento. (SILVA et al, 2016)

Convém salientar ainda, que a análise exploratória muitas vezes pode ser confundida com análise preditiva, mas elas não são equivalentes. Segundo (SILVA et al., 2016), a analise preditiva é um processo que permite descobrir relacionamentos entre amostras de um conjunto de dados através de uma série de características (atributos descritivos ou variáveis), e o rótulo, utilizado para classificar ou estimar um valor que diz respeito a uma ou mais características. A análise preditiva está ligada à inferência estatística, heurísticas e *machine learning*, uma área da Inteligência Artificial. Enquanto isso a análise exploratória de dados é utilizada para criar visões que representam melhor o conjunto, utilizando métodos presentes na estatística descritiva. A análise exploratória é muito importante para o pré-processamento e o pôs processamento, permitindo um resultado mais preciso e de boa visualização.

2.5 Redes Neurais Artificiais

Outra ferramenta importante usada neste projeto foram as Redes Neurais Artificiais (RNA), mais precisamente os *Multilayer Perceptrons* (MLP), que são uma versão de RNA que implementa *deep learning*.

RNAs são comumente usadas para solucionar problemas complexos que envolvem muitas variáveis nas quais suas relações não sejam muito conhecidas. Uma de suas características mais importantes é a capacidade de aprender por meio de exemplos e de generalizar o que foi aprendido. Após o processo de aprendizado, é gerado um modelo não-linear, que pode ser utilizado para testes em dados nunca vistos no processo de aprendizagem. (SPÖRL et al., 2011)

HAYKIN (2008) menciona que as redes neurais artificiais são uma tentativa de imitar o cérebro humano. Redes neurais, artificiais ou não, são compostas por neurônios, os neurônios são os responsáveis pelo processo de pensamento do nosso cérebro. Um neurônio é composto por três partes, os dendritos, os axônios e o corpo celular, identificados na Figura 2-2. Quando o axônio de um neurônio se aproxima do dendrito de outro neurônio, o axônio emite impulsos nervosos através de processos chamados sinapses. O trabalho do dendrito é receber as sinapses e enviá-las ao corpo celular, também conhecido como soma, a soma por sua vez irá “somar” todos os impulsos recebidos dos dendritos formando um sinal. Dependendo da amplitude do sinal, o axônio irá ou não dar continuidade ao ciclo de informação. (CHURCHLAND e SEJNOWSKI, 2016; HAYKIN, 2008)

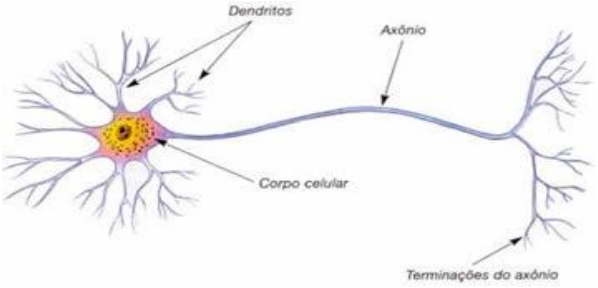

Figura 2-2: Fisiologia básica de um neurônio biológico.

Segundo (HAYKIN, 2008), uma rede neural artificial é formada por neurônios artificiais, como podemos ver, a Figura 2-3 mostra o modelo de um tipo específico de neurônio artificial, o *Perceptron*. Nessa figura podemos identificar três componentes importantes desse neurônio:

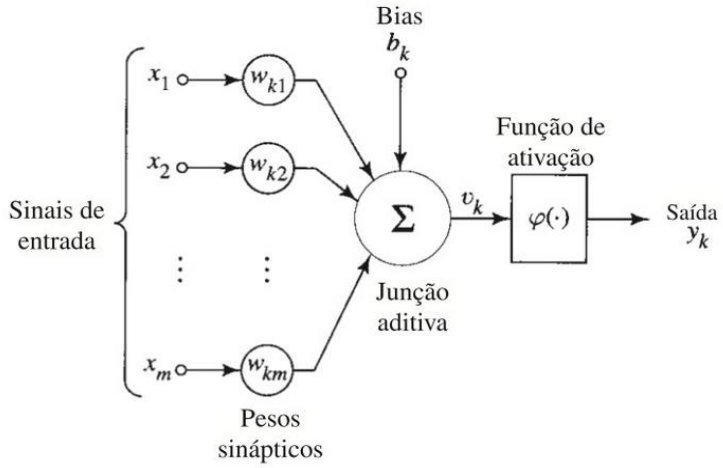

Figura 2-3: Estrutura de um neurônio artificial do tipo *Perceptron*.

O primeiro componente são as sinapses, elas são os sinais x_j que estão sendo transmitidos para suas respectivas portas, sendo multiplicados pelo valor W referente a cada porta e levados a soma. É importante sobressaltar que os valores após a multiplicação podem ser tanto negativos quanto positivos.

O segundo componente é um combinador linear que soma as entradas recebidas incluindo um valor bias, que se refere a um valor que irá aumentar ou diminuir a entrada liquida da função de ativação, dependendo de seu valor se for negativo ou positivo.

O terceiro componente é uma função de ativação, que restringe a amplitude da saída de um neurônio. A função de ativação limita um intervalo que irá permitir que o sinal passe adiante ou não.

Segundo (SILVA et al., 2016) para se utilizar um neurônio artificial na resolução de um problema é necessário determinar cada uma das variáveis que compõem o

modelo neuronal conforme as especificações do problema. Cada uma das sinapses associadas às portas do neurônio pode representar uma informação referente ao problema em questão. Cabe ao projetista do neurônio descrever a função de ativação de modo que esteja de acordo com o problema apresentado. Então o peso de cada porta precisa ser ajustado de acordo com um algoritmo de treinamento, de tal forma que a saída do neurônio seja equivalente a resposta esperada pelo problema.

2.5.1 MLP

A associação de *Perceptrons* permite que se crie uma rede MLP que se trata de uma rede neural de múltiplas camadas, com essa associação é possível obter estruturas mais complexas. Geralmente, redes MLP são compostas por uma camada de entrada, camadas escondidas e uma camada de saída. A camada de entrada é responsável apenas pelo recebimento dos valores, não possuindo nenhum processamento de dados dentro dela. As camadas escondidas são responsáveis pelo processamento da informação, implementando mapas especiais que possibilitam a resolução do problema. A camada de saída é responsável por receber a informação processada provinda das camadas escondidas e fornecer uma resposta. (FLECK et al., 2016; SILVA et al., 2016).

3. METODOLOGIA

A pesquisa desenvolvida neste projeto é aplicada, uma vez que sua utilização prática se dá de modo imediato. O estudo adotou uma abordagem quantitativa, e sua finalidade foi o desenvolvimento de uma metodologia de previsão de paradas de uma turbina. Sobre os meios utilizados no estudo, foi feita uma revisão bibliográfica do tema e foram usadas bases de dados reais disponíveis e a documentação técnica do projeto de pesquisa denominado “Análise Preditiva Baseada em Inteligência Artificial para Sistemas Supervisórios de Usinas Hidrelétricas”.

Durante o processo de desenvolvimento, foram revisados e empregados métodos de ciência de dados, mineração de dados, pré-processamento, análise exploratória e análise preditiva.

Todas as análises e resultados foram tratados com técnicas de visualização de dados e com procedimentos estatísticos formais. Os resultados dos experimentos indicam a viabilidade do modelo proposto.

De forma geral, foram seguidas as etapas abaixo:

1. Revisão bibliográfica

Uma revisão na literatura foi necessária para o desenvolvimento deste projeto. A revisão auxiliou na compreensão do tema e dos métodos utilizados no projeto, e deu início ao processo de desenvolvimento.

2. Levantamento de Dados

O levantamento dos dados reais foi feito em uma usina hidrelétrica. Os dados são provindos de sensores presentes nos equipamentos ligados com os ciclos de carga das turbinas e de alarmes que demonstram falha de equipamento. O período que esses dados foram coletados são de um pouco mais de um ano e eles não vem previamente processados.

3. Pré-processamento

O pré-processamento tem duas etapas, a curadoria dos dados e a normalização dos dados. Na curadoria dos dados, eles são limpos de modo que valores inconsistentes sejam removidos. Esses valores inconsistentes podem ser tanto valores nulos quanto valores fora do padrão que aquele atributo segue. Na normalização dos dados é utilizado um algoritmo que transforma a escala dos dados para -1 e 1, fazendo com que o modelo não tenha problemas, como perda de informação por causa de números muito grandes.

4. Análise exploratória dos Dados

Durante a análise exploratória dos dados, é observado o comportamento dos dados e a relação entre os atributos. Essa análise é feita através de técnicas de visualização de dados e estatísticas descritivas, de modo que as relações sejam mais compreensíveis. Além do comportamento ser observado, também é feito uma segunda curadoria dos dados para remover novas inconsistências encontradas. E por último, a partir dos dados dos sensores, são criados novos dados sobre os ciclos de carga, dados esses que estão implícitos no meio dos dados levantados.

5. Desenvolvimento de Ferramenta de Monitoramento

Nessa etapa, uma ferramenta gráfica que funciona como um painel de controle foi criada. Através dessa ferramenta é possível observar em tempo real os dados transformados durante o processo da análise exploratória, proporcionando um auxílio visual durante o projeto e maior controle sobre os ciclos.

6. Análise preditiva

A fase da análise preditiva se refere ao objetivo final que é a construção de um modelo preditivo capaz de prever as falhas de equipamentos. Durante essa fase, os dados, tanto os dos sensores quanto os dados adquiridos durante a análise exploratória, são inseridos dentro de um modelo *Multilayer Perceptron* criado através de uma biblioteca em Python chamada H2o, para treiná-lo com o intuito de predizer se um alarme irá disparar em 24 horas ou não. Após o treino, o modelo é alimentado com outros dados previamente separados que não fizeram parte do treino, chamados de dados de teste. A partir da classificação dos dados de teste, os alarmes reais são comparados com os alarmes previstos, de modo a calcular o quão bem o modelo preditivo se saiu para dados nunca vistos.

4. RESULTADO E DISCUSSÃO

A pesquisa realizada neste projeto pode ser separada em quatro etapas principais, conforme a metodologia: pré-processamento, onde os dados são limpos e normalizados; analise exploratória de dados, onde os dados são analisados e transformados, o pré-processamento e a analise exploratória foram feitas simultaneamente; monitoramento, onde os dados foram monitorados através de uma ferramenta desenvolvida; e por último o desenvolvimento do modelo MLP que prediz caso um alarme irá disparar em até 24 horas ou não.

Os dados utilizados nesse projeto são provindos de sensores presentes em diferentes equipamentos que estão ligados com os ciclos de carga da turbina e dados referentes aos alarmes que indicam uma parada não intencional, totalizando 24 atributos. Esses dados são coletados de trinta em trinta segundos. E para este projeto foram utilizados um pouco mais de um ano de dados coletados.

4.1 Pré-processamento

Nessa etapa os dados tiveram que ser limpos por causa de algumas inconsistências encontradas. Um exemplo é o caso em que a variável que determina a pressão da válvula injetora do lado A cai para -229 em alguns momentos. Isso é um comportamento muito fora do normal e acaba prejudicando na hora da normalização dos dados por se um outlier.

Outra inconsistência é quando as capturas dos dados que são feitas de trinta em trinta segundos, acabam vindo zeradas por algum motivo desconhecido. Não apenas alguns atributos, mas sim todos. Então é preciso retirar esses momentos em

que as variáveis são zeradas da base de dados, pois além de não agregarem valor, acabam atrapalhando o modelo MLP.

Além de limpar os dados, nessa etapa também é feita a normalização dos dados, de modo que eles sejam escalonados dentro de um limite de -1 a 1, onde -1 é referente ao menor valor encontrado para aquele atributo e 1 é referente ao maior valor. A normalização é muito importante por causa de valores muito grandes, caso um valor seja muito grande o peso dele tem que ser muito pequeno e o contrário também acontece. Pode acontecer desses pesos atingirem valores tão pequenos ou tão grandes que podem consumir mais memória do que está alocada para eles, fazendo com que o modelo falhe.

Não são muitos alarmes que disparam durante esse tempo de um ano e pouco, então os dados acabam ficando desbalanceados. Ou seja, a quantidade de instantes que não há alarme é muito maior que a quantidade de instantes que alarmes são disparados. Por esse motivo, a variável alarme é estendida, de modo que caso haja um alarme no instante x, esse alarme é estendido em 24 horas determinando assim um período em que o modelo tem que prever esse alarme. Então o modelo não prevê o alarme apenas no instante que ele é acionado, mas sim até 24 horas antes.

4.2 Análise exploratória dos Dados

Na etapa da análise exploratória dos dados é onde é feita uma análise de como os dados se comportam e onde são identificadas mais inconsistências. Na tabela 4.1 é possível ver algumas estatísticas descritivas como a média, o valor máximo e mínimo, quartis e desvio padrão dos valores de um atributo. Esses valores ajudam na hora de extrair conhecimento a partir dos dados. Como por exemplo na Tabela 4.1 é possível perceber que a pressão da válvula injetora do lado A está na maioria do tempo ligada e que os valores variam de -24 até 75, estando na maioria de seu tempo em volta de 70.

Tabela 4.1: Estatísticas descritivas de alguns sensores

	Pressao_ValvPrincipal_InjLadoA	Pressao_ValvPrincipal_InjLadoB	Cond_Valvula_Borboleta
count	1.406734e+06	1.406734e+06	1.406734e+06
mean	6.161913e+01	6.306867e+01	8.558967e-01
std	2.188966e+01	2.059757e+01	3.511946e-01
min	-2.492000e+01	-2.537000e+01	0.000000e+00
25%	6.519000e+01	6.626000e+01	1.000000e+00
50%	7.067000e+01	7.101000e+01	1.000000e+00
75%	7.076000e+01	7.110000e+01	1.000000e+00
max	7.526000e+01	7.565000e+01	1.000000e+00

Alguns processos de visualização de dados também foram realizados, onde são gerados gráficos que facilitam a compressão, como é o exemplo da figura 4-1. Podemos ver que durante um ciclo de carga normal, o regulador de velocidade sofre poucas variações e se mantem bem perto de 60.

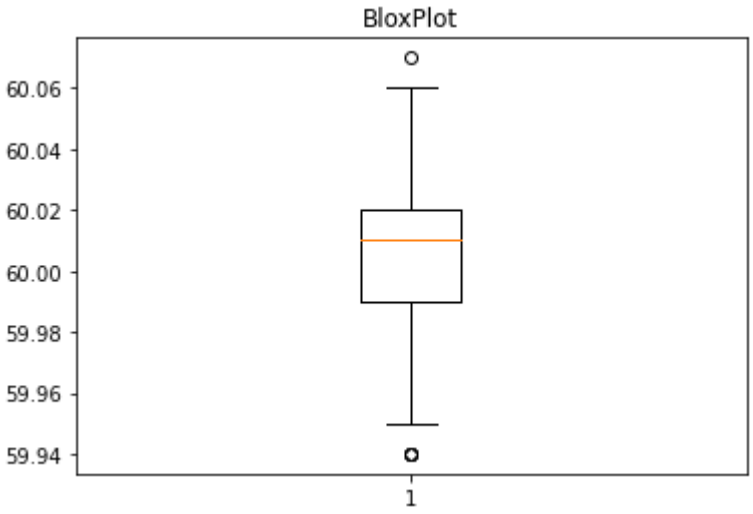

Figura 4-1: *Boxplot* do regulador de velocidade durante um ciclo de carga.

Ainda nesta etapa, por causa que o comportamento e a relação dos dados foram analisados nas etapas anteriores, e com um pouco mais de conhecimento sobre turbinas, novos dados foram criados a partir dos dados dos sensores e dos alarmes. Esses dados estavam implícitos dentro dos outros dados e servem para definir melhor um ciclo de carga. Dentre esses dados estão a quantidade de ciclos da turbina A e B, o tempo desde o último alarme, qual foi o último alarme, entre outros. Esses novos atributos servem mais para representar tempo e desgaste, algo que os sensores não conseguem mostrar com tanta acurácia. Durante a fase de análise preditiva será possível ver o quão importante esses novos atributos foram para o treinamento do modelo MLP.

4.3 Monitoramento

Nesta etapa, uma ferramenta gráfica foi desenvolvida através da biblioteca *shiny* em R, para o auxílio visual durante o projeto. Essa ferramenta que pode ser vista na figura 4-2, serve como um painel de controle onde é possível ver algumas das variáveis criadas durante o processo de análise exploratória dos dados.

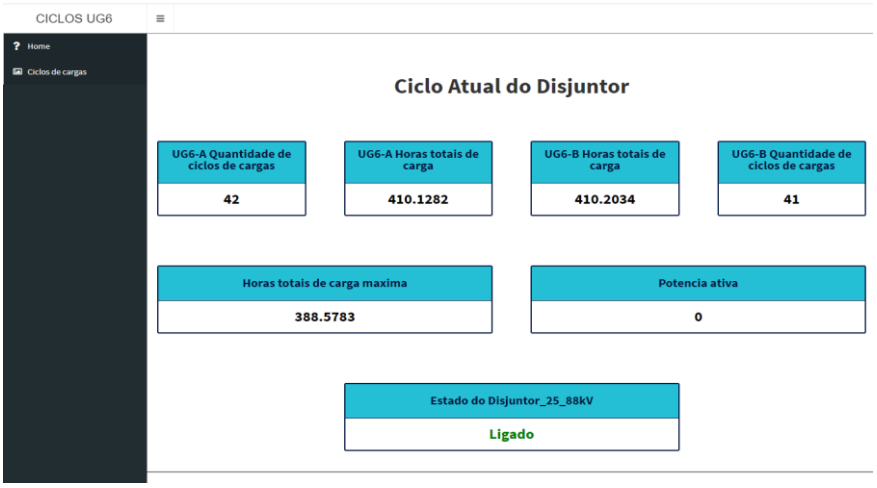

Figura 4-2: Painel de controle mostrando o ciclo atual do disjuntor.

A ferramenta é composta por duas partes. A primeira parte pode ser vista na figura 4-2, nessa parte temos algumas caixas que contêm alguns dos atributos criados. Ela representa o ciclo atual do disjuntor, que pode ser definido como o tempo desde que o disjuntor liga até o momento que ele é desligado. Apenas durante esse ciclo que as turbinas podem são ligadas e energia é gerada. A segunda parte pode ser vista na figura 4-3, nessa parte dados referentes a ciclos passados do disjuntor são mostrados.

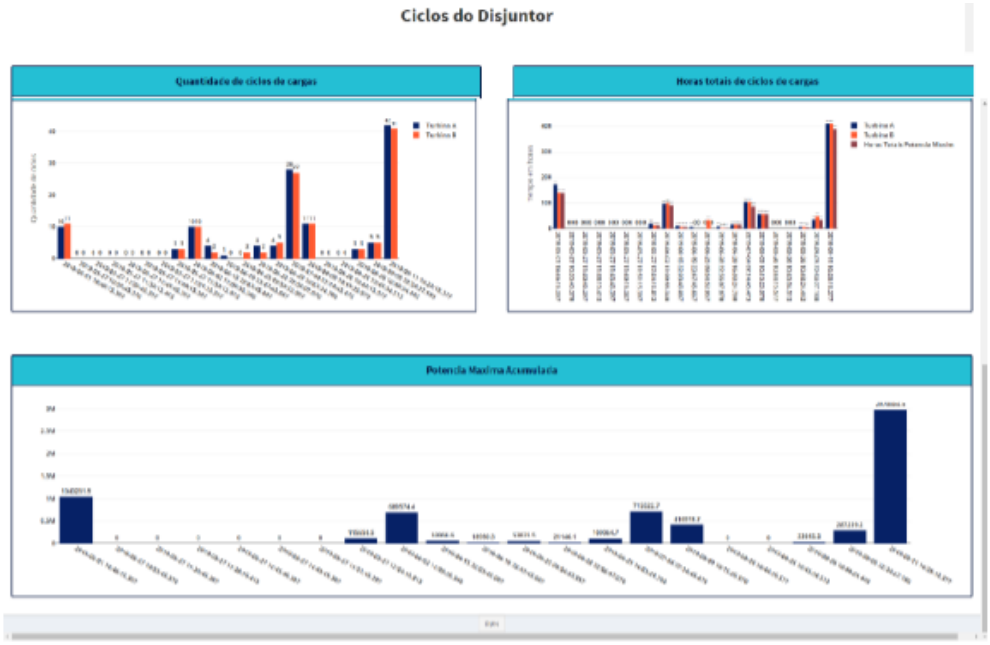

Figura 4-3: Painel de controle mostrando dados sobre os ciclos passados do disjuntor.

4.4 Análise preditiva e desenvolvimento do modelo

O modelo preditivo é um modelo *Multilayer Perceptron* (MLP) que contém duas camadas escondidas. Esse modelo foi criado através da biblioteca em *python* chamada H2o. Na tabela 4.2 é possível ver a arquitetura da MLP utilizada. A saída desse modelo é uma probabilidade, 0 caso não irá tocar um alarme em até 24 horas ou 1 caso irá tocar um alarme em até 24 horas.

Tabela 4.2: Dados da Arquitetura do modelo MLP utilizado

layer	units	type	dropout	l1	l2
1	34	Input	0		
2	200	Rectifier	0	0	0
3	200	Rectifier	0	0	0
4	2	Softmax		0	0

Para o treino do modelo, os dados das turbinas junto com os atributos criados durante a etapa de analise exploratória são separados em dados de treino e teste, onde 87,5% dos dados vão para o treino e o restante para o teste.

As medidas utilizadas durante o treino para medir o desempenho do modelo são mostradas na figura 4-4 junto com seus valores finais. Depois que o modelo foi treinado, ele é então testado e assim é calculada as mesmas medidas calculadas durante o treino para o teste. Por fim o resultado do teste pode ser visto na figura 4-5.

```
MSE: 0.007797188361131185
RMSE: 0.08830168945796668
LogLoss: 0.09005203281165272
Mean Per-Class Error: 0.007895432377232225
AUC: 0.9978162766051768
AUCPR: 0.998255904881267
Gini: 0.9956325532103536
```

Figura 4-4: Medidas de desempenho do modelo durante o treino.

```
MSE: 0.0069263784225824755
RMSE: 0.08322486661198368
LogLoss: 0.057033599568264245
Mean Per-Class Error: 0.008877010692370346
AUC: 0.9968671114349656
AUCPR: 0.9742887946335668
Gini: 0.9937342228699313
```

Figura 4-5: Medidas de desempenho do modelo durante o teste.

Das medidas encontradas nas figuras 4-4 e 4-5, à medida que mais foi considerada para medir o desempenho do modelo, por se tratar de um problema de regressão logística, foi o valor Gini que quanto mais perto de 1, menos erros de classificação estão sendo cometidos.

A biblioteca h2o ainda nos proporciona uma medida que mostra a importância relativa de cada atributo para a predição do alarme. Como é possível ver na tabela 4.3, alguns dos atributos que são mais importantes para prever caso haja um alarme ou não, são os atributos criados na etapa da análise exploratória dos dados, indicados pelos retângulos verdes. Suas importâncias relativas são determinadas pelo atributo "relative_importance"

Tabela 4.3: Importância das variáveis para a classificação

	variable	relative_importance	scaled_importance	percentage
0	Tempo_CicloMax	1.000000	1.000000	0.059905
1	Ultimo_Alarme	0.971785	0.971785	0.058215
2	Tempo_Ultimo_Alarme	0.856414	0.856414	0.051303
3	Vazao_CV	0.764851	0.764851	0.045818
4	Pressao_ValvPrincipal_InjLadoB	0.713174	0.713174	0.042723
5	Quant_CiclosA	0.601315	0.601315	0.036022
6	Tempo_CicloB	0.592556	0.592556	0.035497
7	Pot_Ativa	0.563559	0.563559	0.033760
8	Posicao_Defletor_A	0.547473	0.547473	0.032796
9	Quant_CiclosMax	0.533210	0.533210	0.031942
10	Posicao_Agulha_1A	0.532897	0.532897	0.031923

5. CONSIDERAÇÕES FINAIS

O valor Gini que o modelo MLP desenvolvido atinge para os dados de teste é de 0.993. Esse valor não é absoluto e pode ser melhorado. Mas esse resultado confirma que os dados referentes aos ciclos de cargas têm sim relação com paradas operacionais por falhas no equipamento embora essa relação não seja clara. O resultado também confirma que se os dados forem analisados e monitorados através da análise exploratória dos dados, é possível criar um modelo que prevê a maioria das falhas de equipamento a partir desses dados.

Em síntese, os resultados encontrados nesse estudo são promissores e podem, ser futuramente elaborados para que sua aplicação na indústria seja feita de modo a otimizar as manutenções de usinas hidrelétricas, podendo economizar reduzindo o número de manutenções desnecessárias e evitando danos com manutenções necessárias.

Como continuidade desta pesquisa, alguns dos meios para aprofundar ainda mais esse estudo seria acrescentar mais dados de mais sensores, adicionar dados referentes a manutenção dos equipamentos, testar diferentes arquiteturas para a rede neural e também criar novos atributos a partir dos dados adquiridos, além dos que já foram criados neste projeto.

REFERÊNCIAS

CHURCHLAND, P. S. e SEJNOWSKI, T. J. (2016) The computational brain. MIT press.

DE CASTRO, L. N.; FERRARI, D. G. (2016). Introdução a Mineração de Dados: Conceitos Básicos, Algoritmos e Aplicações, Saraiva.

DEVORE, J. L. (2014) Probability and Statistics for Engineering and the Sciences. 8° ed. Cengage Learning,

EMAE – Empresa Metropolitana de Águas e Energia (2019). Operação do Sistema. Documento Técnico. Disponível em http://www.emae.com.br/conteudo.asp?id=Operacao-do-Sistema. Acesso em: 15/03/2019

EMAE – Empresa Metropolitana de Águas e Energia (2019). Operação do Sistema. Documento Técnico. Disponível em http://www.emae.com.br/conteudo.asp?id=Usina-Hidroeletrica-Henry-Borden. Acesso em: 15/03/2019

FAYYAD, U., PIATETSKY-SHAPIRO, G., SMYTH, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37

FLECK, L., TAVARES, M. H., EYNG, E., HELMANN, A. C., e ANDRADE, M. A. (2016). Redes Neurais Artificiais: Princípios Básicos. *Revista Eletrônica Científica Inovação e Tecnologia*, *1*(13), 47-57.

GREIN, H.; ANGEHRN, R.; LORENZ, M. e BEZINGE, A. (1985) INSPECTION PERIODS FOR PELTON RUNNERS. 49-54, 56.

H2O (2020). The H2O Python Module. H2O Documentation. Disponível em: https://docs.h2o.ai/h2o/latest-stable/h2o-py/docs/intro.html. Acesso em: 10/07/2020

HAYKIN, S. (2008) Neural networks and learning machines. 3° ed, Prentice-Hall.

HAN, J., PEI, J., KAMBER, M. (2011) Data mining: concepts and techniques. Elsevier.

HOAGLIN, D. C. (2003). John W. Tukey and data analysis. *Statistical Science*, 311-318.

HOAGLIN, D. C., MOSTELLER, F., & TUKEY, J. W. (Eds.). (1983). *Understanding robust and exploratory data analysis* (Vol. 3). New York: Wiley.

MONTGOMERY, D. C.; RUNGER, G. C. (2013) Estatística Aplicada e Probabilidade para Engenheiros. 5 ed. Rio de Janeiro: Livros Técnicos e Científicos.

MORGENTHALER, S. (2009). Exploratory data analysis. *Wiley Interdisciplinary Reviews: WIREs Computational Statistics*, Vol *1, jul-aug*, 33-44.

R Project (2020). Shiny v1.5.0. R Foundation. Disponível em: https://www.rdocumentation.org/packages/shiny/versions/1.5.0. Acesso em 10/11/2019

SILVA, L. S.; PERES, S. M. e BOSCARIOLI, C. (2016) Introdução a mineração de dados: Com aplicações em R. 1° ed. Elsevier.

SPÖRL, C.; CASTRO, E. G.; LUCHIARI, A. Aplicação de Redes Neurais Artificiais na construção de modelos de fragilidade ambiental. Revista do Departamento de Geografia, v. 21, n.1, p. 113-135, 2011.

TUKEY, J. W. (1962). The future of data analysis. *The annals of mathematical statistics*, *33*(1), 1-67.

WICKHAM, H. e STRYJEWSKI, L. (2011). 40 years of boxplots. Selected publications. Disponível em: https://vita.had.co.nz/papers/boxplots.html Acesso em: 20/01/2020

Contatos: danielfarinam@gmail.com , aavallim@gmail.com