

MODELO DE PREVISÃO DE DEMANDA DE ORDENS DE MANUTENÇÃO EMERGENCIAIS BASEADA EM FATORES CLIMÁTICOS EM UM SISTEMA DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA

FORECASTING DEMAND MODEL FOR EMERGENCY MAINTENANCE ORDERS BASED ON CLIMATIC FACTORS IN AN ELECTRICITY DISTRIBUTION SYSTEM

Lucas Machado Basso* E-mail: lucas_mbasso@hotmail.com
Jéssica de Assis Dornelles** E-mail: jessica.dornelles@ufrgs.br
Verônica Maurer Tabim** E-mail: veronica.tabim@ufrgs.br
Vinícius Jaques Garcia* E-mail: viniciusjg@ufsm.br

*Universidade Federal de Santa Maria (UFSM), Santa Maria, RS, Brasil

**Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brasil

Resumo: A previsão da demanda de ordens de manutenção é fundamental para o planejamento dos recursos de mão-de-obra, materiais e infraestrutura necessários para atender aos chamados e garantir a disponibilidade do sistema de distribuição de energia elétrica. O objetivo deste estudo é desenvolver um modelo de previsão de demanda para ordens de manutenção emergencial em um sistema de distribuição de energia elétrica considerando fatores climáticos que possam afetar essa demanda. Este estudo apresenta uma pesquisa aplicada quanto a sua natureza, exploratória quanto aos seus objetivos e com abordagem quantitativa. O método de trabalho apresenta quatro etapas principais: (i) mineração dos dados das ordens de manutenção emergenciais em relação aos fatores climáticos; (ii) determinação do método de previsão de demanda adequado; (iii) elaboração de um modelo de previsão de demanda para as ordens de emergência considerando fatores climáticos; e (iv) validação do modelo de previsão da demanda a partir de dados coletados. Como resultado obteve-se o modelo de previsão de demanda de ordens de manutenção emergenciais baseado em fatores climáticos para um sistema de distribuição de energia elétrica.

Palavras-chave: Previsão de demanda. Mineração de dados. Manutenção emergencial. Regressão múltipla. Distribuidora de energia elétrica.

Abstract: Forecasting the demand for maintenance orders is essential for planning the manpower, materials, and infrastructure resources needed to answer calls and ensure the availability of the electricity distribution system. The objective of this study is to develop a demand forecasting model for emergency maintenance orders in an electricity distribution system considering climatic factors that may affect this demand. This study presents applied research as to its nature, exploratory as to its objectives, and with a quantitative approach. The working method has four main stages: (i) mining of emergency maintenance order data in relation to climatic factors; (ii) determining the appropriate demand forecasting method; (iii) elaborating a demand forecast model for emergency orders considering climatic factors; and (iv) validation of the demand forecast model based on collected data. As a result, the demand forecasting model for emergency maintenance orders based on climatic factors for an electricity distribution system was obtained.

Keywords: Demand forecast. Data mining. Emergency maintenance. Multiple regression. Electricity distributor.

1 INTRODUÇÃO

As concessionárias de distribuição de energia elétrica devem oferecer serviços de fornecimento de energia que tenham qualidade, níveis de tensão apropriados e uma baixa taxa de interrupção. Essas empresas geralmente realizam programas de manutenção preventiva e corretiva, para melhorar a confiabilidade do sistema, estabelecendo melhores condições de trabalho para prolongar a vida útil de seus equipamentos (BERTLING *et al.*, 2005).

A previsão de demanda de energia elétrica é essencial para a gestão de energia, programação de manutenção e decisões de investimento nos mercados de energia e na flutuação dos preços dos combustíveis. A demanda por energia elétrica depende de variáveis econômicas, sociais e das condições climáticas (Mirasgedis *et al.*, 2006). A previsão da demanda de eletricidade desempenha um papel fundamental no gerenciamento de sistemas de energia, especialmente para fins de operação e manutenção (GONZÁLEZ-ROMERA *et al.*, 2006).

Segundo Sittithumwat *et al.* (2004), os clientes atendidos por uma rede de distribuição de energia elétrica podem sofrer paralisações frequentes por diversos motivos. Contudo, ainda segundo os autores, quando essas falhas ocorrem é necessário o atendimento e reestabelecimento rápido do fornecimento de energia.

Para o rápido atendimento dos chamados de manutenção emergenciais, a empresa de energia deve manter capacidade de recursos humanos e materiais disponíveis para atender a demanda (FROGER *et al.*, 2016). A quantidade de capacidade a se ter disponível depende da quantidade e complexidade dos chamados de manutenção emergenciais. Acredita-se que parte desses problemas emergenciais são decorrentes de condições climáticas como ventos e chuvas (NGUYEN *et al.*, 2018).

Diante disso, o objetivo do presente trabalho é elaborar um modelo de previsão de demanda de ordens de manutenção emergenciais com a finalidade de auxiliar uma distribuidora de energia localizada no estado do Rio Grande do Sul no planejamento de suas equipes de manutenção, tanto em termos de quantidade como alocação das mesmas, considerando fatores climáticos.

Para alcançar este objetivo, foi realizada uma revisão teórica sobre previsão de demanda e ordens de manutenção para compreender o funcionamento do

gerenciamento de sistemas de distribuição de energia elétrica e, também sobre técnicas de modelagem para previsão de demanda. Em seguida, foi realizada a mineração dos dados brutos da distribuidora e utilizada a técnica de regressão múltipla para elaborar o modelo de previsão de demanda para as ordens emergenciais considerando os fatores climáticos significativos. Como resultado, determinou-se o modelo para a previsão da demanda baseada em fatores climáticos significativos que, por meio do teste estatístico MAPE (erro absoluto médio percentual), foi validado.

2 REFERENCIAL TEÓRICO

A revisão teórica sobre modelos de previsão de demanda e sistemas de distribuição de energia elétrica tem o objetivo de dar suporte ao desenvolvimento do presente estudo. Nesta etapa foram pesquisados os seguintes tópicos: previsão de demanda de energia elétrica e modelos de previsão de demanda de energia elétrica, além do método de regressão linear múltipla e mineração de dados.

2.1 Previsão de demanda de energia elétrica

No contexto atual, a mudança climática tornou-se um fenômeno inequívoco. Embora envolva principalmente a mudança de temperatura, outras variáveis climáticas, como chuva, velocidade do vento, evaporação e umidade, também podem ser afetadas como resultado da mudança climática. Abordar os impactos das mudanças climáticas na demanda de energia elétrica é essencial para prever a demanda futura. Por exemplo, os requisitos de refrigeração e aquecimento mudam significativamente em relação às mudanças climáticas que podem resultar na mudança na demanda de carga de eletricidade (AHMED *et al.*, 2018).

Está bem documentado que fatores relacionados ao clima desempenham um papel importante no consumo de energia elétrica. Por muitos anos, empresas de serviços públicos e a indústria de energia elétrica se interessaram pela relação entre consumo de energia elétrica e clima, e desenvolveram algoritmos empíricos de normalização de clima, visando melhorar a previsão de carga sujeita às variações no clima regional (SAILOR, 2000).

A seleção de variáveis climáticas adequadas para previsão da demanda de eletricidade é crítica, pois afeta a precisão da previsão. Diferentes variáveis climáticas podem ter diferentes impactos na demanda de energia elétrica devido as diferentes condições geográficas (VU *et al.*, 2015). Os elementos climáticos que influenciam a demanda de eletricidade são, em ordem decrescente de importância, temperatura, umidade, vento e precipitação (MIRASGEDIS *et al.*, 2007).

A previsão de demanda de energia elétrica tem sido uma área ativa nas pesquisas há muitos anos, bem como um tópico fundamental para operações e planejamento de sistemas de energia. A falta de precisão das previsões de demanda e carga tem impacto negativo no custo operacional das empresas de energia (HAIDA; MUTO, 1994).

No novo ambiente liberalizado dos mercados de eletricidade, é importante entender e ser capaz de prever os efeitos das variáveis naturais sobre necessidade de resfriamento. Espera-se que o resfriamento contínuo nos setores de processamento de alimentos, armazenamento, industrial, comercial e doméstico e aquecimento contínuo nos processos industriais funcione o tempo todo, independentemente das variações de temperatura ao ar livre. Como o sistema de processamento e armazenamento de alimentos troca calor com os arredores, qualquer alteração na temperatura externa pode alterar a quantidade de eletricidade necessária (HEKKENBERG *et al.*, 2009).

O aumento da temperatura global devido às mudanças climáticas tornou-se uma grande preocupação para as estratégias sobre demanda de energia elétrica. Como o aumento da temperatura tem um impacto significativo na demanda de eletricidade, a expansão da rede elétrica e o plano de aumento relacionado das indústrias de energia serão afetados devido à mudança nas condições climáticas (AHMED *et al.*, 2012). A mudança climática pode atribuir a maior demanda de eletricidade per capita em algumas regiões, enquanto a demanda em outras regiões pode diminuir levando a um maior resfriamento e menores requisitos de aquecimento. Portanto, é necessário saber como as mudanças climáticas desempenharão um papel no crescimento da demanda futura de eletricidade e afetarão o planejamento da expansão do sistema de energia (SALOMÃO *et al.*, 2007).

Sailor (2000) aplicou vários modelos de regressão para vários estados nos EUA e destacou que a sensibilidade da demanda de eletricidade com relação à temperatura é positiva em alguns estados, enquanto é negativa em outros estados. Estudos realizados por Mirasgedis *et al.* (2007) para a Grécia mostraram que a mudança climática faz com que a demanda de eletricidade aumente significativamente. Ruth e Lin (2006) implementaram modelos de regressão múltipla considerando as mudanças climáticas e atividade socioeconômicas e encontraram relações significativas entre a demanda de eletricidade e a mudança climática.

Ang *et al.* (1992) descobriram que o consumo residencial de eletricidade em Cingapura foi sensível a pequenas mudanças nas variáveis climáticas, particularmente a temperatura, que estava intimamente ligada à crescente difusão de aparelhos elétricos para controles ambientais, mas uma grande parte do crescimento futuro na demanda de eletricidade surge da crescente necessidade de ar condicionado, o que levará a variações sazonais cada vez maiores no consumo de eletricidade.

Os principais estudos indicam que a relação entre demanda de eletricidade e a temperatura é claramente não linear. Esta não linearidade refere-se ao fato de que ambos os aumentos e diminuições de temperatura, ligados a passagem de certas temperaturas “limiares”, aumentam a demanda por eletricidade. Segundo Henley e Peirson (1997), essa resposta é causada pelas diferenças entre a temperatura ambiente ou a temperatura externa e o conforto ou a temperatura interna. Quando o diferencial entre a temperatura interior e exterior aumenta, o arranque do equipamento de aquecimento ou resfriamento correspondente aumenta imediatamente à procura de energia. Naturalmente, a curva da resposta da demanda às temperaturas depende especialmente das características climáticas da área geográfica a que os dados de demanda se referem, que condicionam o tipo de equipamento instalado e as condições sob as quais ele começará a operar.

2.1.1 Modelos de previsão de demanda de energia elétrica

Singh *et al.* (2012) realizam uma revisão abrangente de métodos e proposições de modelos aplicados à previsão de demanda, incluindo técnicas como: regressão simples, regressão múltipla, suavização exponencial, mínimos quadrados

ordinários, previsão de carga adaptativa, séries temporais estocásticas auto regressivas, modelo ARMA, modelo ARIMA, máquinas de vetores de suporte, modelos baseados em *soft computing*, algoritmos genéticos, lógica *fuzzy*, redes neurais e sistemas especialistas baseados em conhecimento. Podemos analisar que são vários os métodos relatados em trabalhos anteriores que podem ser utilizados para se realizar uma previsão de demanda.

Alguns trabalhos recentes nesta área de estudo apresentam a combinação de dois ou mais métodos acima mencionados. Nie *et al.* (2012) descreve a combinação de um modelo de Média Móvel Integrada Autoregressiva (ARIMA) para a previsão da demanda diária de energia elétrica, combinado com *Support Vector Machines* (SVM) para corrigir o desvio da tendência básica de previsão anterior. A principal desvantagem do uso de modelos ARIMA é que a relação importante com as variáveis meteorológicas não é explicada (APADULA *et al.*, 2012).

Ahmed *et al.* (2012) utilizaram a análise de regressão linear múltipla para as variáveis climáticas históricas para estabelecer uma correlação entre a demanda de eletricidade e variáveis-chave associadas. Na sequência, os autores utilizaram a análise de séries temporais para prever a temperatura futura e os graus de refrigeração e aquecimento correspondente em *New South Wales*, local do estudo realizado.

Âmbar *et al.* (2018) comparam as capacidades e a eficiência de previsão de cinco modelos diferentes, prevendo o consumo de eletricidade de um edifício administrativo localizado em Londres, Reino Unido. Esses modelos são: Regressão Múltipla, Programação Genética, Rede Neural Artificial, Rede Neural Profunda e Máquina de vetor de Suporte. Os modelos de previsão são desenvolvidos com base em cinco parâmetros diferentes, como radiação solar, temperatura, velocidade do vento, umidade e índice de dias uteis. O consumo de eletricidade previsto pelos modelos é comparado com o consumo real, sendo assim demonstrando que a Rede Neural Artificial tem o melhor desempenho, apresentando o menor erro médio percentual absoluto (MAPE) entre os modelos.

Son e Kim (2017) realizaram um estudo para fornecer um modelo preciso para previsão de um mês da demanda de eletricidade no setor residencial. O estudo utilizou 14 variáveis meteorológicas e cinco variáveis sociais. Com base na regressão vetorial de suporte e na seleção de características *fuzzy-rough* com

algoritmos de otimização de enxame de partículas, o método proposto estabeleceu um modelo com variáveis que se relacionam com a previsão ignorando variáveis que inevitavelmente levam a erros de previsão. O modelo de previsão proposto foi validado usando dados históricos da Coréia do Sul entre janeiro de 1991 e dezembro de 2012. Este modelo também foi comparado com o obtido a partir de rede neural artificial, média móvel integrada auto regressiva (ARIMA) e modelos de regressão linear múltipla, mostrando ser o melhor modelo para previsão de curto prazo.

Ao analisar os estudos encontrados na literatura utilizando os modelos citados, temos exemplos como Apadula *et al.* (2012), onde um modelo de regressão linear múltipla baseado em variáveis relacionadas ao calendário e ao clima foi desenvolvido para estudar as relações entre as variáveis meteorológicas e a demanda de eletricidade, bem como para prever a demanda mensal de eletricidade um mês à frente. Além disso, no estudo de Ahmed *et al.* (2018) também foi utilizada a abordagem de regressão múltipla baseada em eliminação regressiva para analisar a influência das variáveis climáticas na previsão de carga elétrica. Uma análise de correlação foi realizada usando o coeficiente de *Pearson* para examinar a interdependência entre diferentes variáveis climáticas no contexto de *Sydney*. Assim, a regressão múltipla baseada na eliminação para trás foi utilizada para excluir as variáveis climáticas não significativas e avaliar a sensibilidade de variáveis significativas relacionadas à demanda de carga elétrica.

Nesse sentido, o presente estudo considerou a modelagem por meio da regressão múltipla a mais indicada ao contexto estudado, uma vez que buscou-se elaborar um modelo de previsão de demanda de energia elétrica com base nos fatores climáticos. E, de acordo com a literatura, outros estudos utilizaram esta mesma abordagem.

2.2 Regressão linear múltipla

Segundo Ghani e Ahmad (2010), existem seis tipos de análises de regressão disponíveis em estatísticas, sendo essas: regressão linear simples, regressões lineares múltiplas, regressão logística, regressão ordinal, regressão multinomial e análise discriminante. Regressões lineares múltiplas (MLR) é o método de

estatística em regressão que analisa a relação entre a variável de resposta única (variável dependente), com duas ou mais variáveis controladas (variáveis independentes).

Segundo Oliveira *et al.* (2011), modelos de regressão múltipla envolvem três ou mais variáveis ou estimadores. Isto significa que uma variável é função de duas ou mais variáveis independentes (explicativas). O propósito do uso de variáveis independentes adicionais é melhorar a capacidade preditiva do modelo, em comparação com a regressão linear simples. Mesmo quando há interesse no efeito de apenas uma variável, é importante incluir variáveis que possam afetar a variável principal, realizando uma análise de regressão múltipla. Assim, pode-se reduzir o desperdício estocástico, pois reduzir a variância residual aumenta o poder dos testes de significância. Por meio de regressões múltiplas, também podemos eliminar a tendência que poderia resultar se fosse simplesmente ignorada uma das variáveis que afeta diretamente a variável dependente.

De acordo com Hadi e Homayoon (2017), o objetivo geral da regressão múltipla é aprender mais sobre a relação entre diversas variáveis independentes e uma variável dependente. Em geral, os procedimentos de regressão múltipla estimam uma equação linear é apresentada pela Equação 1 a seguir:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_nX_n + \varepsilon \quad (1)$$

Onde ε é a taxa de erro do modelo. Os coeficientes de regressão ou "termos b" representam as contribuições independentes de cada variável independente para a previsão da variável dependente. A linha de regressão expressa a melhor previsão da variável dependente (Y), dadas as variáveis independentes (X). No entanto, geralmente há variação substancial dos pontos observados em torno da linha de regressão ajustada. O desvio de um ponto particular da linha de regressão (seu valor previsto) é chamado de valor residual (HADI; HOMAYOON, 2017).

Dentro da regressão linear múltipla, a análise de correlação é importante pois, conforme apresentado em Lira (2004), o coeficiente de correlação $\hat{\rho}$ é interpretado como um indicador que descreve a interdependência entre as variáveis X e Y. Outra forma de interpretar o coeficiente de correlação é em termos de $\hat{\rho}^2 = R^2$, denominado

coeficiente de determinação ou de explicação. Quando multiplicado por 100, o $\rho^2 = R^2$ fornece a porcentagem da variação em Y (variável dependente), que pode ser explicada pela variação em X (variável independente), ou seja, o quanto de variação é comum às duas variáveis.

2.2.1 Mineração de dados

A mineração de dados pode descobrir tendências e padrões ocultos que não aparecem na análise de consulta simples, para que possa ter uma parte significativa em termos de descoberta de conhecimento e tomada de decisão. Tais tarefas podem ser preditivas, como classificação e regressão, ou descritivas, como *Clustering* e associação (TAN, 2007). De acordo com Liu *et al.* (2019) a mineração de dados é um acoplamento de descobertas e práticas científicas, cujo assunto é coletar, gerenciar, processar, analisar e visualizar a grande quantidade de dados estruturados ou não estruturados.

Segundo Brown (2008) o termo “mineração de dados” não significa somente minerar dados em si, mas em minerar informações a partir dos dados. Segundo o autor, vivemos em tempos ricos em dados e, a cada dia que passa, mais dados são coletados e armazenados em bancos de dados. Com isso o autor afirma que o desejo de processar e usar dados para ajudar a responder ou entender questões importantes impulsionou o desenvolvimento de técnicas de mineração de dados, ou seja, o objetivo dessas técnicas é encontrar informações nos dados, sendo informação, padrões persistentes e significativos. Então, o autor explica que a mineração de dados por vezes é entendida como a descoberta de conhecimento, que atribui o objetivo mais profundo de extração de conhecimento a partir de dados.

Os estudos de Olkin e Sampson (2001) relatam que a mineração de dados se refere a um conjunto de abordagens e técnicas que permitem extrair informações valiosas de vastas e vagamente estruturadas bases de dados. Como exemplo, os autores relatam o caso de um fabricante de produtos de consumo que pode usar a mineração de dados para entender melhor o relacionamento das vendas de um produto específico com as estratégias promocionais, as características da loja de vendas e os dados demográficos regionais.

A mineração de dados e a descoberta de conhecimento são um processo iterativo que envolve diversas etapas, incluindo seleção, limpeza, pré-processamento e transformação de dados; incorporação de conhecimentos prévios, análise com algoritmos computacionais e/ou abordagens visuais, interpretação e avaliação de resultados; formulação ou modificação de hipótese e teorias e, também, ajuste aos dados e métodos de análises (MENNIS; GUO, 2009).

As técnicas de mineração de dados incluem métodos para agrupar elementos de dados que possuem recursos semelhantes (descoberta de padrões, regras de associação e *clustering*) e métodos para determinar relacionamentos preditivos entre grupos de elementos de dados, desenvolvendo modelos a partir de técnicas de regressão, classificação, redes neurais, máquinas de vetores, árvores de decisão entre outros (HARRISON, 2008).

Para Gordan *et al.* (2020) a mineração de dados é o processo de identificação de informações valiosas de diferentes bancos de dados, que podem lidar com a análise qualitativa do comportamento complexo do mundo real. Segundo os autores, o conhecimento obtido dos dados brutos pelo processo de mineração de dados é caracterizado principalmente como modelos ou padrões e é muito significativo que esse conhecimento deva ser válido, autêntico, inovador e compreensível, como relacionamentos significativos de grandes quantidades de dados.

De acordo com Pang-Ning *et al.* (2006), a mineração de dados geralmente pode ser categorizada em três classes:

- Mineração preditiva: na mineração de dados preditiva, a chave principal é prever um valor contínuo;
- Mineração descritiva: usada para encontrar características relevantes com ocorrência frequente em um grupo;
- Mineração prescritiva: utiliza algoritmos de otimização e simulação para encontrar o melhor resultado.

No caso do presente estudo, a técnica de mineração adequada é a descritiva uma vez que se busca determinar características que seguem um determinado padrão dentro da base bruta de dados.

3 MÉTODO DA PESQUISA

Este estudo foi realizado a fim de atender às necessidades de uma distribuidora de energia localizada no estado do Rio Grande do Sul. Esta distribuidora realiza manutenções emergenciais na rede de distribuição de energia elétrica da região central do estado e necessitava determinar a quantidade de mão de obra para realizar as manutenções em tempo apropriado, para isso foi proposta a elaboração de um modelo de previsão de demanda para as ordens de manutenção emergenciais.

A condução de pesquisas científicas deve estar balizada em pressupostos metodológicos para que possa ser considerada válida e apresentar resultados coerentes (MARCONI; LAKATOS, 2010). Dessa forma, esta pesquisa trata-se de um estudo de caso aplicado quanto a sua natureza, exploratório quanto aos seus objetivos e com abordagem quantitativa, o método de trabalho utilizado para atingir o objetivo proposto será descrito nesta seção, e abrange as técnicas, etapas e instrumentos utilizados.

Deste modo, a coleta de dados foi realizada a partir do acesso ao banco de dados de registro de ordens de manutenção emergencial disponibilizado pela empresa. A sequência de etapas que compuseram o desenvolvimento desta pesquisa foi delineada segundo os pressupostos referentes à previsão de demanda de energia elétrica e a técnica de regressão linear múltipla.

3.1 Mineração dos dados das ordens de manutenção emergenciais em relação aos fatores climáticos

Esta etapa teve início após o recebimento dos dados da distribuidora que eram referentes a todas as ordens de chamados de manutenção da rede elétrica, nos períodos de 2013, 2014, 2015, 2016 e 2017, na região de Santa Maria e alguns municípios do estado do Rio Grande do Sul. Como o enfoque do presente estudo se limitava em realizar a previsão de demanda para um local específico, os dados que não eram referentes ao município de Santa Maria não foram utilizados. Sendo assim, em razão de problemas de falta de dados em alguns anos, a pesquisa foi conduzida por meio dos dados do ano de 2017 que estavam completos.

As informações contidas nos dados brutos recebidos eram: número da equipe designada para a realização da manutenção, tempo de deslocamento das equipes, localização, tempo de manutenção, horário em que foi realizado o serviço, causa ocorrente para ser realizada a chamada, detalhe da causa, componente danificado, previsão climática no momento da manutenção e várias outras questões relatadas sobre o serviço realizado.

Após a compreensão do funcionamento do processo para a realização das ordens de manutenção, foi realizada a mineração dos dados brutos para posteriormente serem tratados na regressão linear múltipla. A mineração buscou eliminar dados repetidos e dados que não eram classificados como manutenção emergencial, por exemplo, serviços de manutenção comercial na rede elétrica.

Depois de realizada a primeira mineração de dados, ou seja, dados classificados como manutenção emergencial, foi realizada uma nova separação, classificando os dados em manutenção emergencial causada por fatores climáticos e manutenção emergencial geral. A manutenção emergencial causada por fatores climáticos se refere às ordens de chamada de manutenção com o detalhe da causa sendo inundação, descarga atmosférica, vendaval ou temporal. Enquanto que os dados referentes a manutenção emergencial geral, se referem a dados que tem como detalhe da causa a falha de componente, poste com problema, defeito transitório, problemas com a vegetação (árvores), animais diversos, transformadores e outras falhas que causam chamadas emergenciais.

Como o objetivo do presente trabalho era a elaboração de um modelo de previsão de demanda de ordens de manutenção emergenciais relacionados aos fatores climáticos, os dados utilizados foram referentes a manutenção emergencial causada por fatores climáticos que são traduzidos em quantidades de horas diárias utilizadas para a realização desses serviços de manutenções emergenciais do ano de 2017 no município de Santa Maria - RS.

3.2 Determinação do método de previsão de demanda adequado

Com base na análise dos dados coletados e de acordo com as aplicações observadas na literatura, foi determinada a utilização do método de regressão linear múltipla para o grupo de dados de manutenção emergencial causadas por fatores

climáticos a fim de elaborar um modelo de previsão de demanda para esse tipo de ordem.

O método de regressão linear múltipla utilizou variáveis climáticas para explicar a demanda de ordens de manutenção emergências, que foram: máxima rajada no dia, soma da precipitação, insolação, evaporação, umidade relativa média, velocidade média do vento, *Cooling Degree Days* e *Heating Degree Days*. Esses dados referentes às variáveis climáticas foram obtidos a partir do Instituto Nacional de Meteorologia (INMET) do ano de 2018.

Para executar esse método foram utilizadas ferramentas para análises estatísticas, tais como correlação entre a variável dependente “Y”, que representa as horas de manutenção emergencial, com as variáveis climáticas já citadas, esta correlação serviu para medir a associação entre as variáveis. Além disso, alguns testes de hipótese foram realizados para definir se os dados atendiam às premissas para a utilização do método de regressão linear múltipla, dentre eles: o teste ANOVA, que tem como objetivo comparar a distribuição das variáveis do problema; o teste de hipótese para analisar homocedasticidade, com a finalidade de analisar os resíduos da regressão, a fim de verificar se os resíduos estavam correlacionados com alguma variável do modelo; o teste de hipótese de *Durbin-Watson*, para analisar a presença de auto correlação (dependência) nos resíduos do modelo de regressão múltipla; o teste de hipóteses de *Kolmogorov-Smirnov*, para medir a normalidade da distribuição; o teste de hipótese de fator de inflação da variância, para analisar a existência de multicolineariedade entre variáveis. E, com a intenção de resolver problemas com os resíduos obtidos a partir da Regressão múltipla, foi realizada a transformação logarítmica nos dados para novas análises.

3.3 Elaboração de um modelo de previsão de demanda para as ordens de emergência considerando fatores climáticos

O método de regressão linear múltipla foi utilizado para desenvolver um modelo de previsão de demanda para os dados referentes a manutenção emergencial causadas por fatores climáticos, com o objetivo de explicar a partir de variáveis climáticas o comportamento da demanda.

Os ajustes dos dados para o desenvolvimento deste método provêm da mineração dos dados e de ferramentas de análises estatísticas.

O desenvolvimento do modelo foi realizado por meio do software STATISTICA®.

3.4 Validação do modelo de previsão da demanda a partir dos dados coletados

Para realizar validação do modelo de previsão de demanda diária de horas de manutenção emergenciais foram utilizados dos dados fornecidos pela empresa de um período de 7 dias de horas de manutenção emergenciais, denominado período de teste.

A validação do modelo elaborado foi realizada a partir do teste MAPE (erro absoluto médio percentual) em que é calculado o erro absoluto em cada período, e então calculada a média desses erros absolutos percentuais. Por meio desse teste, é possível fornecer uma indicação de quão grandes os erros de previsão estão em comparação com os valores dos dados reais (HANKE *et al.*, 2001).

4 RESULTADOS E DISCUSSÕES

De acordo com a metodologia apresentada, os resultados do presente estudo de caso se concentram na execução das etapas do método de trabalho, ou seja, na mineração dos dados das ordens de manutenção emergenciais em relação aos fatores climáticos, na realização da regressão linear múltipla e consequente elaboração do modelo de previsão de demanda para as ordens de emergência conforme os dados minerados e, por fim, na validação do modelo de previsão da demanda a partir dos dados coletados.

4.1 Mineração dos dados das ordens de manutenção

A etapa de mineração dos dados consistiu na separação dos dados brutos fornecidos pela empresa distribuidora de energia elétrica. Esses dados eram referentes às horas de atividades de manutenção emergenciais que a empresa realizou na cidade de Santa Maria no ano de 2017. A amostra era proveniente de

365 dias, dos quais 7 dias foram utilizados para a etapa de treino e validação do modelo.

Deste modo, os dados foram minerados em horas de ordens de manutenção emergencial com causas relacionadas a fatores climáticos (descarga elétrica, vento, etc.), classificados na base de dados da empresa como causa “Meio ambiente”. As demais ordens de manutenção, ou seja, provenientes de causas gerais (desgaste de equipamentos, acidentes, etc.), não foram consideradas para a análise e regressão, uma vez que não eram o objeto deste estudo.

4.2 Regressão linear múltipla

A regressão linear múltipla foi realizada considerando os dados da demanda de ordens emergenciais como variável dependente “Y”, e os fatores climáticos como variáveis independentes “x1, x2, ..., xn. Como variáveis independentes utilizaram-se: máxima rajada no dia, soma da precipitação, insolação, evaporação, umidade relativa média, velocidade média do vento, *Cooling Degree Days* e *Heating Degree Days*.

Para os dados de clima de Santa Maria, utilizou-se a base de dados do INMET que possui uma estação meteorológica na cidade. Sendo assim, para a realização das análises estatísticas dos dados obtidos, foi utilizado o software STATISTICA®.

Primeiramente foi analisada a relação das correlações entre todas as variáveis envolvidas no modelo conforme apresentado na Figura 1.

Figura 1 – Correlação entre todas as variáveis envolvidas no modelo

Variable	Correlations (Spreadsheet1)								
	Max rajada	precipitação	Insolação	Evaporação	Umidade	vel med vento	HDD	CDD	emerg
Max rajada	1,000000	0,320192	-0,107289	0,418020	-0,231476	0,683151	-0,061600	0,043665	0,407594
precipitação	0,320192	1,000000	-0,378024	-0,090149	0,338943	0,061631	-0,239886	0,331246	0,184849
Insolação	-0,107289	-0,378024	1,000000	0,159254	-0,661201	0,000735	0,541793	-0,684710	-0,052640
Evaporação	0,418020	-0,090149	0,159254	1,000000	-0,481609	0,372643	0,399415	-0,247537	-0,016146
Umidade	-0,231476	0,338943	-0,661201	-0,481609	1,000000	-0,376206	-0,375479	0,632224	0,030977
vel med vento	0,683151	0,061631	0,000735	0,372643	-0,376206	1,000000	-0,203591	0,020035	0,099748
HDD	-0,061600	-0,239886	0,541793	0,399415	-0,375479	-0,203591	1,000000	-0,643429	-0,084346
CDD	0,043665	0,331246	-0,684710	-0,247537	0,632224	0,020035	-0,643429	1,000000	0,079204
emerg	0,407594	0,184849	-0,052640	-0,016146	0,030977	0,099748	-0,084346	0,079204	1,000000

Fonte: Elaborado pelos autores (2020) com STATISTICA®.

A correlação é realizada para medir quanto a variável dependente é explicada pela variável independente, além disso, mensura o grau de variância comum entre

as variáveis (GUJARATI; PORTER, 2011). Deste modo, por meio da Figura 1 podemos analisar que a variável dependente “Y” do modelo é “emerg”, e assim podemos analisar o valor da correlação dela com as variáveis independentes.

Na sequência, realizou-se a regressão linear múltipla entre a variável dependente horas emergenciais e as demais variáveis climáticas. O modelo de regressão múltipla foi ajustado e os coeficientes calculados, conforme apresentado na Figura 2. As variáveis em vermelho são as variáveis que resultaram significativas no modelo.

Figura 2 - Coeficientes do modelo de regressão múltipla

Regression Summary for Dependent Variable: emerg (Spreadsheet1) R= ,51747919 R²= ,26778471 Adjusted R²= ,25133044 F(8,356)=16,274 p<0,0000 Std.Error of estimate: 234,81						
N=365	Beta	Std.Err. of Beta	B	Std.Err. of B	t(356)	p-level
Intercept			-224,837	198,9819	-1,12994	0,259263
Max rajada	0,744166	0,071473	55,572	5,3374	10,41187	0,000000
precipitação	-0,042383	0,054441	-0,805	1,0342	-0,77851	0,436786
Insolação	0,145155	0,075789	9,838	5,1369	1,91524	0,056262
Evaporação	-0,152349	0,064494	-16,605	7,0295	-2,36224	0,018703
Umidade	0,048611	0,084013	1,083	1,8717	0,57861	0,563220
vel med vento	-0,351336	0,072716	-113,469	23,4848	-4,83161	0,000002
HDD	-0,095930	0,073345	-11,669	8,9218	-1,30793	0,191740
CDD	0,037008	0,076002	4,522	9,2858	0,48694	0,626603

Fonte: Elaborado pelos autores (2020).

Pode-se verificar que poucas variáveis resultaram significativas para o modelo, além de um R² relativamente baixo (0,26) ou seja, a variável dependente é pouco explicada pelo modelo selecionado. Ao analisar os resíduos do modelo de regressão linear múltipla, verificou-se que ele era inapropriado, pois seus resíduos eram heterocedásticos, ou seja, a variância entre eles não é constante. Essa conclusão baseou-se nos testes de hipótese de Breusch-Pagan, que segundo Pesaran et al. (2008), pressupõe que existe uma relação linear entre o quadrado dos resíduos padronizados e as variáveis explicativas do modelo. O teste de hipótese de Breusch-Pagan examina se os resíduos estimados estão correlacionados com alguma variável do modelo. O teste é apresentado por meio da Figura 3 e considera que a hipótese nula é de homocedasticidade, ou seja, tem-se um p valor < 0,05, logo rejeita-se a hipótese nula de que os resíduos sejam homocedásticos.

Figura 3 - Teste de hipótese para homocedasticidade

```

studentized Breusch-Pagan test

data: regsomeioamb
BP = 53.889, df = 8, p-value = 7.253e-09

> car::ncvTest(regsomeioamb)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 4405.891, Df = 1, p = < 2.22e-16
    
```

Fonte: Elaborado pelos autores (2020).

A partir disto, foi realizada uma nova seleção nos dados, sendo estes selecionados pelo detalhamento da causa, que foram: descarga atmosférica, ventos e inundação. Outra questão foi a implementação de uma nova variável independente, representada pela variável dependente com uma defasagem. Essa defasagem é representada por uma diferença de tempo nos dados observados, sendo assim, essa variável entrou no modelo para ajudar na explicação, com a intenção de investigar se o dia anterior causou impacto no dia posterior. A equação que explica o conceito de defasagem de uma variável segundo Mattos (2018), é apresentada na Equação 2:

$$\Delta X_t = X_t - X_{t-1} \tag{2}$$

Deste modo, os coeficientes calculados e o modelo de regressão linear múltipla ajustado são apresentados na Figura 4, onde as variáveis em vermelho são consideradas significativas para o modelo.

Figura 4 - Coeficientes do modelo de regressão múltipla

Regression Summary for Dependent Variable: emerg (Spreadsheet)						
R= ,55002896 R²= ,30253185 Adjusted R²= ,28479961						
F(9,354)=17,061 p<0,0000 Std.Error of estimate: 187,62						
	Beta	Std.Err. of Beta	B	Std.Err. of B	t(354)	p-level
N=364						
Intercept			-175,246	159,3047	-1,10007	0,272049
Max rajada	0,764526	0,070850	46,773	4,3345	10,79080	0,000000
precipitação	-0,056038	0,053397	-0,869	0,8281	-1,04946	0,294681
Insolação	0,133442	0,074659	7,392	4,1359	1,78736	0,074735
Evaporação	-0,159230	0,063134	-14,169	5,6179	-2,52210	0,012103
Umidade	0,036491	0,082503	0,664	1,5005	0,44230	0,658542
vel med vento	-0,377287	0,071715	-99,560	18,9243	-5,26095	0,000000
HDD	-0,098858	0,071927	-9,822	7,1463	-1,37443	0,170179
CDD	0,019085	0,074375	1,904	7,4211	0,25661	0,797629
y defasado	0,175432	0,044694	0,175	0,0447	3,92519	0,000104

Fonte: Elaborado pelos autores (2020).

Novamente foi obtido um R^2 relativamente baixo, de aproximadamente 30% e apenas quatro das variáveis foram significativas para o modelo proposto, de acordo com a Figura 4. Ao analisar os resíduos do modelo de regressão linear múltipla, verificou-se que o modelo era inapropriado, pois seus resíduos são heterocedásticos, ou seja, a variância entre eles não é constante. Essa conclusão baseou-se novamente nos testes de hipótese de *Breusch-Pagan*, conforme a Figura 5.

Figura 5 - Teste de hipótese para homocedasticidade

```
studentized Breusch-Pagan test
data: regmaisset
BP = 53.535, df = 9, p-value = 2.314e-08
> car::ncvTest(regmaisset)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 4256.894, Df = 1, p = < 2.22e-16
```

Fonte: Elaborado pelos autores (2020).

A fim de resolver o atual problema de homocedasticidade dos resíduos da regressão, foi utilizado o método de transformação logarítmica. Para a transformação logarítmica foi necessária a exclusão dos dados com valores iguais a zero da variável dependente. Esses dados zerados representavam dias em que o sistema falhou ao coletar dados e dias em que ordens emergenciais não foram registradas. Além disso, foi realizada a exclusão dos dados *outliers*, para melhor ajuste do modelo, totalizando uma amostra de 345 dias para o desenvolvimento do modelo.

As variáveis foram reconsideradas a partir do método da defasagem, sendo elas: Y (tempo de manutenção), X1 (Y defasado), X2 (máxima rajada no dia), X3 (máxima rajada defasada), X4 (precipitação), X5 (precipitação defasada), X6 (insolação), X7 (evaporação), X8 (umidade) e X9 (velocidade média vento).

Foi possível analisar por meio da Figura 6, que os valores das correlações entre a variável dependente com as variáveis independentes possuem valores maiores, ou seja, correlação mais forte, ao comparar com a Figura 1.

Figura 6 - Correlação entre variáveis modelo de regressão

Variable	Correlations (Spreadsheet2)									
	X1	X2	X3	X4	X5	X6	X7	X8	X9	Y
X1	1,000000	-0,046146	0,374960	0,038793	0,272119	-0,082473	-0,086939	0,104987	0,031240	0,395409
X2	-0,046146	1,000000	0,237313	0,169961	-0,186391	-0,106595	0,489514	-0,272839	0,747131	0,410111
X3	0,374960	0,237313	1,000000	0,003873	0,154221	-0,174689	0,266788	0,038399	0,253170	0,156676
X4	0,038793	0,169961	0,003873	1,000000	0,235020	-0,416778	-0,154833	0,586058	-0,094771	0,274325
X5	0,272119	-0,186391	0,154221	0,235020	1,000000	-0,179096	-0,581381	0,351288	-0,158991	0,093081
X6	-0,082473	-0,106595	-0,174689	-0,416778	-0,179096	1,000000	0,182161	-0,435307	-0,105757	-0,111283
X7	-0,086939	0,489514	0,266788	-0,154833	-0,581381	0,182161	1,000000	-0,458348	0,388737	0,121765
X8	0,104987	-0,272839	0,038399	0,586058	0,351288	-0,435307	-0,458348	1,000000	-0,344263	-0,073431
X9	0,031240	0,747131	0,253170	-0,094771	-0,158991	-0,105757	0,388737	-0,344263	1,000000	0,203864
Y	0,395409	0,410111	0,156676	0,274325	0,093081	-0,111283	0,121765	-0,073431	0,203864	1,000000

Fonte: Elaborado pelos autores (2020).

Por meio da Figura 7 é possível analisar os valores da regressão linear múltipla para o modelo ajustado. As variáveis em vermelho, ou seja, com valor-p < 0,05 foram consideradas significativas pelo modelo de regressão.

Afim de analisar as variáveis que obtiveram um p-valor significativo para o atual modelo desenvolvido, foram elaborados gráficos de dispersão entre estas variáveis independentes significativas com a variável dependente do modelo de regressão. As Figuras 8, 9, 10, 11 e 12 apresentam os gráficos de dispersão da variável dependente Y com as variáveis significativas X1, X2, X4, X8 e X9, respectivamente.

Figura 7 - Valores dos coeficientes do modelo de regressão

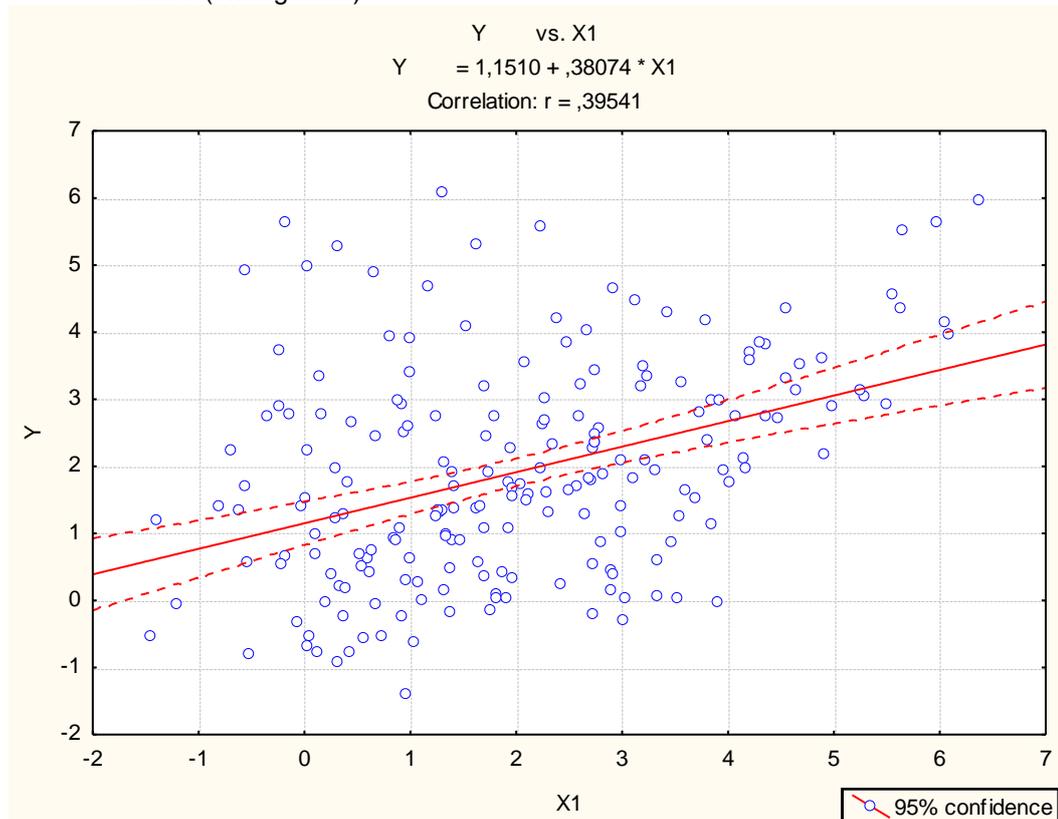
Regression Summary for Dependent Variable: Y (Spreadsheet2)						
R= ,65789960 R²= ,43283188 Adjusted R²= ,40624587						
F(9,192)=16,280 p<0,0000 Std.Error of estimate: 1,2431						
N=202	Beta	Std.Err. of Beta	B	Std.Err. of B	t(192)	p-level
Intercept			7,46790	3,623685	2,06086	0,040665
X1	0,454561	0,060879	0,43770	0,058622	7,46659	0,000000
X2	0,528335	0,098584	2,48847	0,464333	5,35924	0,000000
X3	-0,079473	0,069061	-0,36777	0,319585	-1,15077	0,251258
X4	0,260686	0,079940	0,06462	0,019815	3,26104	0,001313
X5	0,048726	0,074935	0,01186	0,018237	0,65025	0,516306
X6	-0,046103	0,065701	-0,01588	0,022626	-0,70170	0,483715
X7	-0,024905	0,087502	-0,04613	0,162076	-0,28463	0,776238
X8	-0,256586	0,084332	-2,31177	0,759812	-3,04256	0,002674
X9	-0,236026	0,089759	-1,02301	0,389045	-2,62954	0,009243

Fonte: Elaborado pelos autores (2020).

A Figura 8 representa o resultado gerado pela combinação da variável dependente Y (emergência) representada pela série histórica de horas de manutenção emergencial por dia vs. a variável X1 (emergência defasada). O coeficiente de correlação resultou em $r = 0,39$, ou seja, uma correlação positiva, de

média intensidade representando a grande dispersão dos pontos representados no gráfico.

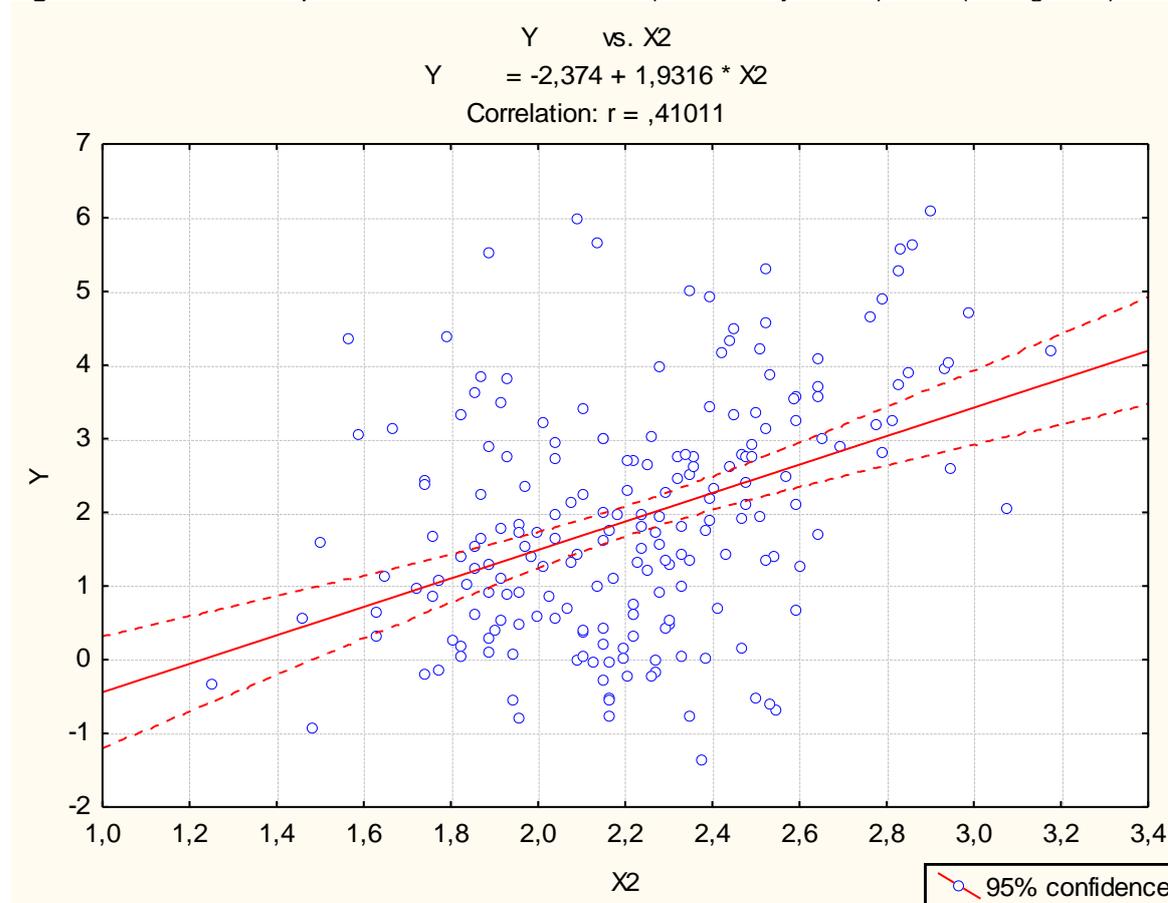
Figura 8 - Gráfico de dispersão entre as variáveis X1 (emergência defasada) vs. Y (emergência)



Fonte: Elaborado pelos autores (2020).

A Figura 9 apresenta a correlação entre a combinação da variável dependente Y (emergência) vs. X2 (máxima rajada dia). O coeficiente de correlação resultante foi de $r = 0,41$, ou seja, uma correlação positiva e de média intensidade, representando a grande dispersão dos pontos representados no gráfico.

Figura 9 - Gráfico de dispersão entre as variáveis X2 (máxima rajada dia) vs. Y (emergência)

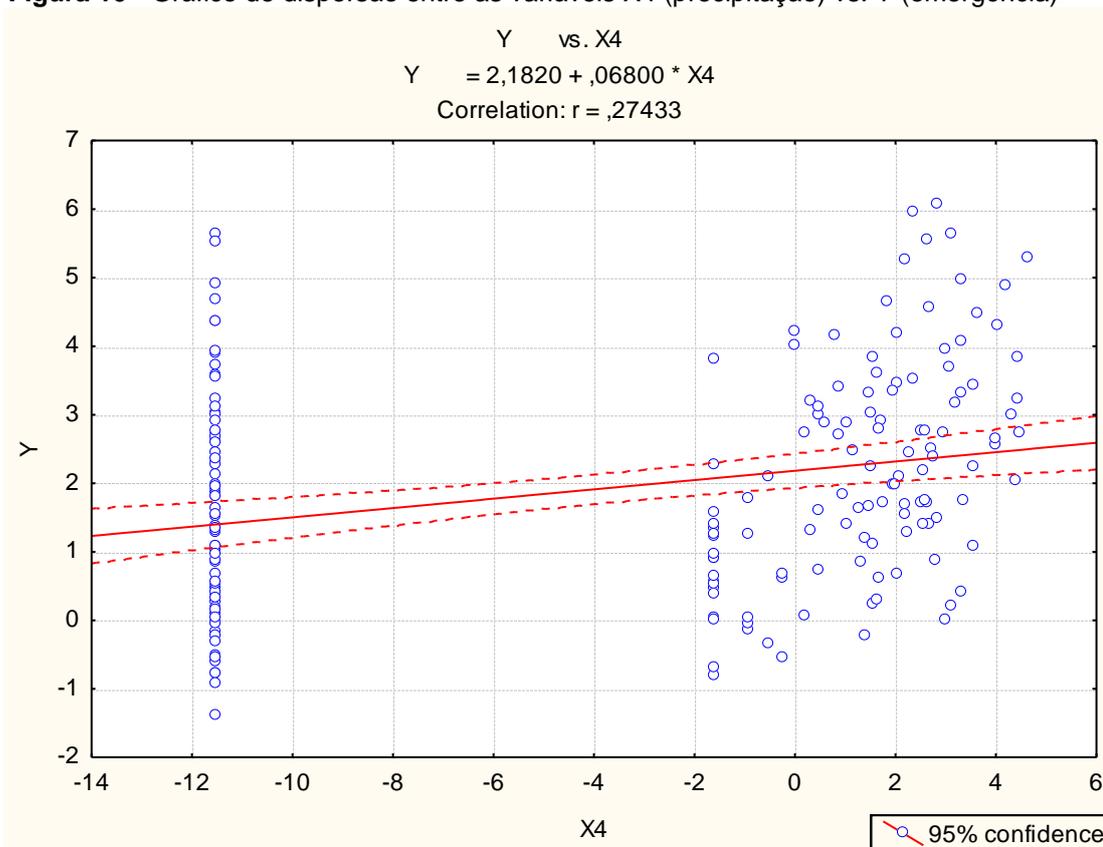


Fonte: Elaborado pelos autores (2020).

A Figura 10 demonstra o resultado gerado pela combinação da variável dependente Y (emergência) vs. X4 (precipitação). O coeficiente de correlação resultante foi $r = 0,27$, ou seja, uma correlação positiva e de baixa intensidade, a explicação para que exista uma linha de pontos separada da nuvem de pontos clássicos de um gráfico de dispersão é referente aos dias em que não houve registro de valor (precipitação).

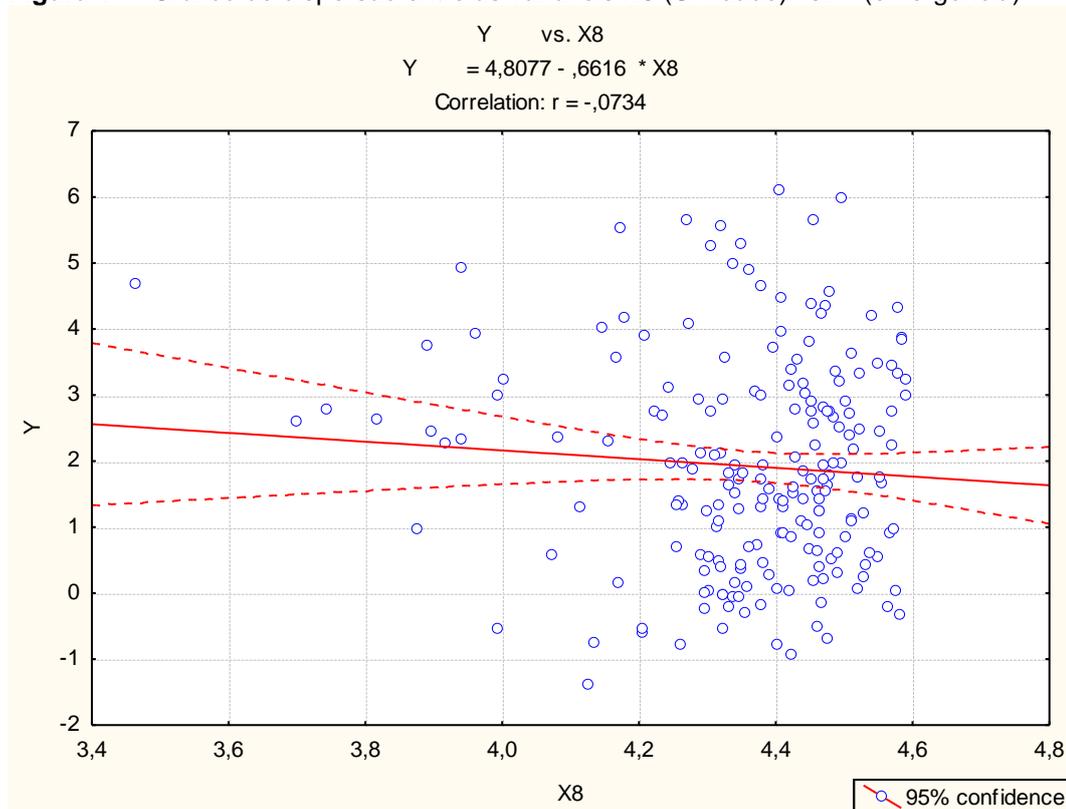
A Figura 11 demonstra o resultado gerado pela combinação da variável dependente Y (emergência) vs. X8 (umidade). O coeficiente de correlação resultante foi $r = -0,07$, ou seja, uma correlação negativa e de baixa intensidade, o que ajuda a entender a grande dispersão da nuvem de pontos identificadas no gráfico.

Figura 10 - Gráfico de dispersão entre as variáveis X4 (precipitação) vs. Y (emergência)



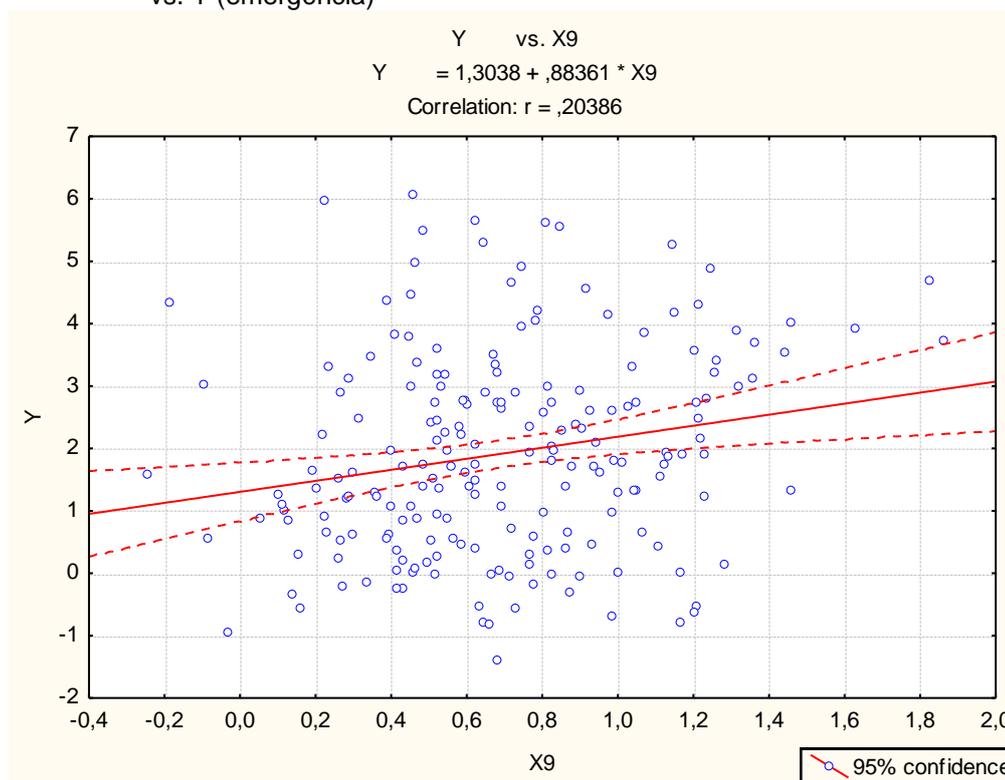
Fonte: Elaborado pelos autores (2020).

Figura 11 - Gráfico de dispersão entre as variáveis X8 (Umidade) vs. Y (emergência)



Fonte: Elaborado pelos autores (2020).

Figura 12 – Gráfico de dispersão entre as variáveis X9 (Velocidade média vento) vs. Y (emergência)



Fonte: Elaborado pelos autores (2020).

A Figura 12 demonstra o resultado gerado pela combinação da variável dependente Y (emergência) vs. X9 (velocidade média vento). O coeficiente de correlação resultante foi $r = 0,20$, ou seja, uma correlação positiva e de baixa intensidade, o que ajuda a entender a grande dispersão da nuvem de pontos identificadas no gráfico.

Para o melhor ajuste do modelo, as variáveis não significativas ($p\text{-valor} > 0,05$) foram excluídas. Esse procedimento foi realizado de forma iterativa, ou seja, com a exclusão de uma variável por vez. O modelo de regressão em que todas as variáveis foram significativas é apresentado na Figura 13.

As estimativas dos valores estatísticos do modelo foram de R^2 múltiplo igual a 0,423 e R^2 ajustado de 0,408. O coeficiente de determinação do modelo ajustado é de 0,4235, ou seja, 42,35% da variável dependente Y (tempo de manutenção) é explicada pelas variáveis independentes do modelo de regressão. Os testes de homocedasticidade foram satisfatórios, conforme apresentado na Figura 14.

Figura 13 - Valores dos coeficientes do modelo de regressão

Regression Summary for Dependent Variable: Y (Spreadsheet)						
R= ,65080506 R ² = ,42354722 Adjusted R ² = ,40884179						
F(5,196)=28,802 p<0,0000 Std.Error of estimate: 1,2404						
N=202	Beta	Std.Err. of Beta	B	Std.Err. of B	t(196)	p-level
Intercept			6,21611	3,416536	1,81942	0,070374
X1	0,438289	0,054967	0,42204	0,052929	7,97365	0,000000
X2	0,485211	0,092894	2,28536	0,437533	5,22328	0,000000
X4	0,291586	0,076597	0,07228	0,018986	3,80674	0,000188
X8	-0,235701	0,075952	-2,12360	0,684311	-3,10327	0,002197
X9	-0,225853	0,087368	-0,97892	0,378683	-2,58506	0,010462

Fonte: Elaborado pelos autores (2020).

Figura 14 - Teste de hipótese para homocedasticidade

```
> lmtest::bptest(regteste2) # Breusch-Pagan test
studentized Breusch-Pagan test
data: regteste2
BP = 9.6202, df = 5, p-value = 0.08674
> car::ncvTest(regteste2)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 2.380327, Df = 1, p = 0.12287
```

Fonte: Elaborado pelos autores (2020).

Ao interpretar o teste de hipótese para homocedasticidade, foi obtido o p-valor = 0,08674, ou seja, p-valor > 0,05, sendo assim, não foi possível rejeitar a hipótese nula, ou seja, não existe forte evidência que a variação dos resíduos não é constante, portanto, assume-se que eles sejam homocedásticos.

Foi realizado o teste de *Durbin-Watson*, para analisar a presença de auto correlação (dependência) nos resíduos do modelo de regressão múltipla (JOHNSTON, 1977). O resultado do teste é apresentado na Figura 15, o p-valor = 0,82, ou seja, p-valor > 0,05, sendo assim, não foi possível rejeitar a hipótese nula, portanto, assume-se que os resíduos são independentes.

Figura 15 - Teste de *Durbin-Watson*

```
> durbinwatsonTest(regteste2)
lag Autocorrelation D-w statistic p-value
1 -0.02181643 2.037078 0.82
Alternative hypothesis: rho != 0
```

Fonte: Elaborado pelos autores (2020).

A Figura 16 apresenta um resumo dos resultados dos resíduos da regressão. É possível analisar que a coluna *Standard Residual*, apresenta valores entre -3 e 3, o que significa que não existem mais *outliers* no modelo de regressão, além disso, a coluna *Cook's Distance* apresenta valores entre -1 e 1, o que significa que não há valores de *outliers* que possam ser influentes no modelo de regressão. Segundo Kim et al. (2001), a distância de *Cook* serve para detectar observações influentes, que podem levar a exclusão de uma dada observação.

Figura 16 - Resumo de resultados dos resíduos

Case No.	Predicted & Residual Values (Spreadsheet2)								
	Dependent variable: Y								
	Observed Value	Predicted Value	Residual	Standard Pred. v.	Standard Residual	Std.Err. Pred.Val	Mahalanobis Distance	Deleted Residual	Cook's Distance
Minimum	-1,38741	-0,854331	-3,18651	-2,64280	-2,56899	0,127060	1,11411	-3,25792	0,000000
Maximum	6,08757	4,544397	3,38703	2,49929	2,73065	0,507123	32,60326	3,54338	0,060014
Mean	1,92037	1,920372	-0,00000	-0,00000	0,00000	0,206571	4,97525	0,00577	0,005648
Median	1,74586	1,818038	-0,04623	-0,09747	-0,03727	0,195965	4,02198	-0,04719	0,001994

Fonte: Elaborado pelos autores (2020).

Para confirmar se os resíduos estão distribuídos normalmente, foi realizado o teste de hipóteses de *Kolmogorov-Smirnov*, apresentado na Figura 17.

Figura 17 - Teste *Kolmogorov-Smirnov*

```
> ks.test(residregteste2, "pnorm")

One-sample Kolmogorov-Smirnov test

data: residregteste2
D = 0.079946, p-value = 0.1512
alternative hypothesis: two-sided
```

Fonte: Elaborado pelos autores (2020).

Como resultado do teste obteve-se p-valor = 0,1512, ou seja, p-valor > 0,05, sendo assim, não é possível rejeitar a hipótese nula, portanto, assume-se que os resíduos seguem uma distribuição normal padrão.

Na Figura 18 é apresentado o valor do fator de inflação da variância, que representa a existência de multicolineariedade entre as variáveis.

Figura 18 - Valores de fator de inflação da variância

Current Status of Sweep Matrix; DV: Y (Spreadsheet2)						
	X1	X2	X4	X8	X9	Y
X1	-1,02731	-0,16546	0,01508	0,12249	0,19931	0,438289
X2	-0,16546	-2,93405	1,13448	-0,74257	2,04916	0,485211
X4	0,01508	1,13448	-1,99489	1,27061	-0,59971	0,291586
X8	0,12249	-0,74257	1,27061	-1,96144	-0,00386	-0,235701
X9	0,19931	2,04916	-0,59971	-0,00386	-2,59538	-0,225853
Y	0,43829	0,48521	0,29159	-0,23570	-0,22585	0,576453

Fonte: Elaborado pelos autores (2020).

Ao analisar a Figura 20, observa-se que os valores em módulo são menores que 4, o que confirma a não existência de multicolinearidade entre as variáveis.

4.3 Elaboração do modelo de previsão de demanda

Com as verificações a partir de ferramentas estatísticas e de mineração de dados foi concluído que todos os testes foram satisfatórios para confirmar que o modelo desenvolvido segue os padrões necessários para se realizar a previsão considerando as variáveis independentes e a variável dependente.

O modelo de regressão linear múltipla pode ser descrito a partir da Equação (3):

$$Y = 6,216 + 0,422 X1 + 2,285 X2 + 0,072 X4 - 2,123 X8 - 0,978 X9 \quad (3)$$

4.4 Validação do modelo de previsão da demanda

Para a validar o modelo de previsão a partir do modelo de regressão linear múltipla foi utilizado um período de treino de sete dias. Para analisar os erros gerados pela previsão utilizou-se o teste MAPE, erro absoluto médio percentual, que é calculado por meio da Equação 4.

$$\frac{|Real - Previsão|}{|Real|} * 100 \quad (4)$$

A realização dos cálculos do MAPE, utilizando os sete dias de treino é apresentado na Tabela 1.

Tabela 1 – MAPE, erro absoluto médio percentual

Previsão	Real	MAPE
2,67 horas	3,42 horas	22%
2,93 horas	3,90 horas	25%
3,12 horas	4,01 horas	22%
3,77 horas	3,13 horas	20%
2,10 horas	3,25 horas	35%
3,71 horas	2,63 horas	41%
2,09 horas	3,10 horas	33%
Total 20,42 horas	Total 23,47 horas	Total 28%

Fonte: Elaborado pelos autores (2020).

Considerando o resultado de 28% na comparação entre os erros dos dados previstos com os dados reais coletados, significa que o modelo pode ser utilizado para a previsão da demanda de ordens de manutenção com um erro de 28%.

5 CONSIDERAÇÕES FINAIS

A presente pesquisa teve como objetivo geral desenvolver um modelo de previsão de ordens de manutenção emergenciais para um sistema de distribuição de energia elétrica, baseado em fatores climáticos. O ponto de partida da pesquisa foi realizar um levantamento bibliográfico para identificar as técnicas de previsão de demanda realizadas nas áreas de energia elétrica e manutenção, bem como modelos preditivos.

O estudo na bibliografia identificou a utilização de variáveis climáticas para o desenvolvimento de modelos preditivos de consumo, demanda e manutenção em sistemas de distribuição de energia elétrica, sendo assim, o presente trabalho identificou quais variáveis climáticas poderiam ser utilizadas para o desenvolvimento do modelo de previsão de ordens de manutenção emergenciais.

Para solucionar o questionamento proposto e executar os objetivos do trabalho, foram realizadas análises para entender as características dos dados de horas de manutenção emergenciais, e com isso determinar o método de regressão linear múltipla para elaboração do modelo de previsão de demanda. A mineração dos dados foi realizada a fim de classificar quais variáveis climáticas foram utilizadas no modelo, utilizando testes de hipótese para ajustar os dados perante os pressupostos necessários.

Por meio da análise dos resultados obtidos a partir do modelo desenvolvido de regressão linear múltipla foi possível identificar a relação de variáveis climáticas

como máxima rajada do dia, quantidade de chuva, umidade e velocidade média do vento com as horas de manutenção emergências coletadas no projeto, sendo este modelo com um coeficiente de determinação (R^2) de 0,4235, ou seja, 42,35% dos dados coletados são explicados pelas variáveis. A previsão de demanda de ordens de manutenção emergenciais baseada em fatores climáticos foi realizada e a validação do modelo foi feita a partir do teste estatístico MAPE (erro absoluto médio percentual), que apresentou um resultado de 28%, comparando os erros gerados dos dados previstos com os dados reais coletados.

Como limitações do estudo, pode-se ressaltar que dos dados obtidos para análise, apenas os dados referentes ao ano de 2017 estavam aptos e foram utilizados, sendo estes uma amostra de 365 dias, dos quais 345 foram minerados para o modelo desenvolvido e 7 para a validação da previsão.

Com contribuição, pode-se destacar a utilização de modelos de regressão linear múltipla para a elaboração de modelos de previsão de demanda no setor de energia elétrica, esta prática pode ajudar as empresas deste segmento a otimizarem a utilização de seus recursos. Também acredita-se ser uma contribuição a descoberta dos fatores climáticos que de fato influenciam na previsão da demanda, podendo ser o ponto de partida para estudos futuros em empresas distribuidoras de energia elétrica.

Para pesquisas futuras sugere-se a determinação de outros métodos de previsão de demanda o que pode gerar modelos considerando variáveis independentes diferentes, esses métodos podem ser, por exemplo, redes neurais e árvores de decisão.

REFERÊNCIAS

AHMED, T., VU, D. H., MUTTAQI, K.M., AGALGAONKAR, A.P. Load forecasting under changing climatic conditions for the city of Sydney, Australia. **Energy**, 2018. <https://doi.org/10.1016/j.energy.2017.10.070>

AHMED, T., MUTTAQI, K.M., AGALGAONKAR, A.P. Climate change impacts on electricity demand in the State of New Wales, Australia. **Applied Energy**, 2012. <https://doi.org/10.1016/j.apenergy.2012.03.059>

ANG, B.W, GOH, T.N., LIU, X.Q. Demanda residencial de eletricidade em Cingapura. **Energy**, 1992.

- APADULA, F., BASSINI, A., ELLI, A., SCAPIN, S. Relationships between meteorological variables and monthly electricity demand. **Applied Energy**, 2012. <https://doi.org/10.1016/j.apenergy.2012.03.053>
- BERTLING, L. ALLAN, R. ERIKSSON, R. A reliability-centered asset maintenance method for assessing the impact of maintenance in power distribution systems. **IEEE Power System**, 2005. <https://doi.org/10.1109/TPWRS.2004.840433>
- BROWN, D.E. Introduction to Data Mining for Medical Informatics. **Clinics in Laboratory Medicine**, 2008. <https://doi.org/10.1016/j.cll.2007.10.008>
- FROGER, A. GENDREAU, M. MENDOZA, J. E. PINSON, E. ROUSSEAU, L.M. Maintenance scheduling in the electricity industry: a literature Review. **European Journal of Operational Research**, 2016. <https://doi.org/10.1016/j.ejor.2015.08.045>
- GHANI, I. M., AHMAD, S. Stepwise Multiple Regression Method to Forecast Fish Landing. *In: INTERNATIONAL CONFERENCE ON MATHEMATICS EDUCATION RESEARCH*, 2020. [Proceedings...]. 2010. <https://doi.org/10.1016/j.sbspro.2010.12.076>
- GONZÁLEZ-ROMERA, E., JARAMILLO-MORÁN, M. A., CARMONA-FERNÁNDEZ, D. Monthly electric energy demand forecasting based on trend extraction. **IEEE Transactions on Power Systems**, 2006. <https://doi.org/10.1109/TPWRS.2006.883666>
- GORDAN, M. ISMAIL, Z. RAZAK, H. A. GHAEDI, K. IBRAHIM, Z. TAN, Z.X. GHAYEB, H.H. Data mining-based damage identification of a slab-on-girder bridge using inverse analysis. **Measurement**, 2020. <https://doi.org/10.1016/j.measurement.2019.107175>
- GUJARATI, D.N.; PORTER, D.C. **Econometria básica** .5.ed. São Paulo: McGraw Hill Brasil, 2011.
- HADI, F., HOMAYOON, K. New empirical model to evaluate groundwater flow into circular tunnel using multiple regression analysis. **International Journal of Mining Science and Technology**, 2017. <https://doi.org/10.1109/59.331433>
- HAIDA, T., MUTO, S. Regression based peak load forecasting using a transformation technique. **IEEE Transactions**, 1994.
- HANKE, J. E.; WICHERN, D. W.; REITSCH, A. G. **Business forecasting**. 7. ed. New York: Prentice Hall, 2001.
- HARRISON, J. H. Introduction to the mining of clinical data. **Clinics in Laboratory Medicine**, 2008. <https://doi.org/10.1016/j.cll.2007.10.001>
- HEKKENBERG, M., MOLL, H.C. AJM Schoot Uiterkamp. Dynamic temperature dependence patterns in future energy demand models in the context of climate change. **Energy**, 2009. <https://doi.org/10.1016/j.energy.2009.07.037>
- HENLEY, A., PEIRSON, J. **Não linearidades na demanda de eletricidade e temperatura: Métodos paramétricos versus não paramétricos**. **Boletim Oxford de Economia e Estatística**, 1997.
- INMET. Instituto Nacional de Meteorologia. **Brasil**, 2018.
- JOHNSTON, J. **Métodos econométricos**. São Paulo: Atlas, 1977.

- KIM, C.YONJOO, L. PARK, B. U. Cook's distance in local polynomial regression. **Statistics & Probability Letters**, 2001. [https://doi.org/10.1016/S0167-7152\(01\)00031-1](https://doi.org/10.1016/S0167-7152(01)00031-1)
- LIRA, S. A. **Análise de correlação**: abordagem teórica e de construção dos coeficientes com aplicações". Curitiba, 2004. 196 p. Dissertação (Mestrado) - Setores de Ciências Exatas e de Tecnologia, UFPR.
- LIU, J. KONG, X. ZHOU, X. WANG, L. ZHANG, D. LEE, I. XU, B. XIA, F. Data Mining and Information Retrieval in the 21st century: a bibliographic review. **Computer Science Review**, 2019. <https://doi.org/10.1016/j.cosrev.2019.100193>
- MARCONI, M.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 7. ed. São Paulo: Atlas, 2010.
- MATTOS, R. S. **Tendências e raízes unitárias**. Economia, Universidade Federal de Juiz de Fora, 2018.
- MENNIS, J. GUO, D. Spatial data mining and geographic knowledge discovery: an Introductio. Computers. **Environment and Urban Systems**, 2009. <https://doi.org/10.1016/j.compenvurbsys.2009.11.001>
- MIRASGEDIS, S., SARAFIDIS, Y., GEORGOPOULOU, E., KOTRONI, V., LAGOUVARDOS, K. Modeling framework for estimating impacts of climate change on electricity demand at regional level: Case of Greece. **Energy Convers Manage**, 2007. <https://doi.org/10.1016/j.enconman.2006.10.022>
- NGUYEN, TUYET, T. A. CHOU, S. Y. Maintenance strategy selection for improving cost-effectiveness of offshore Wind systems. **Energy Conversion and Management**, 2018. <https://doi.org/10.1016/j.enconman.2017.11.090>
- OLIVEIRA, M. O., MARZEC, D. P., BORDIN, G., BRETAS, A. S., BERNARDON, D. Climate Change Effect on Very Short-Term Electric Load Forecasting. **IEEE Trondheim PowerTech**, 2011. <https://doi.org/10.1109/PTC.2011.6019249>
- OLKIN, I. SAMPSON, A.R. Multivariate analysis: overview. **International Encyclopedia of the Social & Behavioral Sciences**, 2001. <https://doi.org/10.1016/B0-08-043076-7/00472-1>
- PANG-NING, T. STEINBACH, M. KUMAR, V. **Introduction to data mining**. Boston: Pearson, 2006.
- PESARAN, M.H.; ULLAH, A., YAMAGATA, T. A Bias-adjusted LM test of error cross-section Independence. **Econometrics Journal**, 2008. <https://doi.org/10.1111/j.1368-423X.2007.00227.x>
- RUTH, M., LIN, D.C. Regional energy demand and adaptations to climate change: methodology and application to the state of Maryland, USA. **Política Energética**, 2006. <https://doi.org/10.1016/j.enpol.2005.04.016>
- SAILOR, D.J. Relacionando as cargas de eletricidade do setor residencial e comercial ao clima, avaliando as vulnerabilidades e vulnerabilidades do estado. **Energy**, 2000.
- SALOMÃO S., QIN D., M. MANNING, CHEN Z., MARQUÊS M., AVERYT KB et al. **Mudança climática 2007**: a base da ciência física. Grupo de Trabalho I ao Quarto Relatório de

Avaliação do Painel Intergovernamental sobre Mudança do Clima. Cambridge, UK, Nova Iorque, EUA, Cambridge University Press; 2007.

SINGH, A. K., IBRAHEEM, I., KHATOON, S., MUAZZAM, M., CHATUVERVEDI, D. Load forecasting techniques and methodologies: a review. Power, Control and Embedded Systems. *In: INTERNATIONAL CONFERENCE. [Anais...]*, 2012.

<https://doi.org/10.1109/ICPCES.2012.6508132>

SITTITHUMWAT, A. SOUDI, F. TOMSOVIC, K. Optimal allocation of distribution maintenance resources with limited information. **Electric Power Systems Research**, 2004.

<https://doi.org/10.1016/j.epsr.2003.07.001>

SON, H., KIM, C. Short-term forecasting of electricity demand for the residential sector using weather and social variables. Resources. **Conservation and Recycling**, 2017.

<https://doi.org/10.1016/j.resconrec.2016.01.016>

TAN, P-N. **Introduction data mining**. Omstructor's Solution Manual Pearson Education India, 2007.

VU, D. H., MUTTAQI, K. M., AGALGAONKAR, A. P. A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. **Applied Energy**, 2015. <https://doi.org/10.1016/j.apenergy.2014.12.011>



Artigo recebido em: 04/11/2020 e aceito para publicação em: 15/12/2020

DOI: <http://dx.doi.org/10.14488/1676-1901.v21i1.4163>