

ARTIGO

Análise preditiva do consumo mensal de energia elétrica em unidades consumidoras da SEPM baseada em dados

Predictive analysis of monthly electricity consumption in SEPM consumer units based on data

DOI: doi.org/10.5935/2178-4590.20220009

3º SGT PM Ubiratan da Silva Tavares

ust1973@gmail.com



Data de submissão: 11/05/2022

Data de aceite: 07/07/2022



RESUMO

O atual cenário do setor elétrico brasileiro vem exigindo dos consumidores residenciais, comerciais e industriais um consumo consciente de energia elétrica. Esta conscientização da importância da redução do consumo de energia elétrica e da implantação de medidas de eficiência energética, também têm sido foco na administração pública. Na administração pública do Estado do Rio de Janeiro, a percepção da necessidade de um consumo consciente e eficiente de energia elétrica nos órgãos públicos vêm sendo notada em algumas referências normativas publicadas entre 2011 e 2015. Visando contribuir com os gestores das faturas de energia elétrica das unidades consumidoras da SEPM, com relação a necessidade de otimização do custo relacionado ao consumo de energia. Este trabalho tem como objetivo apresentar uma metodologia baseada em modelo de aprendizado de máquina (modelo preditivo) para o problema de previsão do consumo mensal de energia em unidades consumidoras. Os resultados obtidos com as previsões mostraram uma diferença percentual menor do que 7% em comparação com os valores reais e a métrica de avaliação revelou que o modelo preditivo é de boa qualidade. O modelo preditivo desenvolvido foi capaz de prever a ocorrência de aumento ou redução no consumo mensal de energia.

Palavras-chave: previsão; consumo de energia; aprendizado de máquina.

ABSTRACT

The current scenario of the Brazilian electricity sector has been demanding from residential, commercial and industrial consumers a conscious consumption of electricity. This awareness of the importance of reducing electricity consumption and implementing energy efficiency measures has also been a focus in the public administration. In the public administration of the State of Rio de Janeiro, the perception of the need for conscious and efficient consumption of electricity in public agencies has been noted in some normative references published between 2011 and 2015. Aiming to contribute with the managers of electric energy bills of the consumer units of SEPM, regarding the need for optimization of the cost related to energy consumption. This work aims to present a methodology based on machine learning model (predictive model) for the problem of forecasting monthly energy consumption in consumer units. The results obtained with the predictions showed a percentage difference of less than 7% compared to the real values and the evaluation metric revealed that the predictive model is of good quality. The developed predictive model was able to predict the occurrence of increase or decrease in the monthly energy consumption.

Keywords: forecasting; energy consumption; machine learning.

INTRODUÇÃO

A importância da redução do consumo de energia e implantação de medidas de eficiência energética vêm aumentando na administração pública. Esta percepção vem sendo notada em algumas referências normativas no âmbito da administração pública estadual no Rio de Janeiro. Abaixo é apresentado as referências normativas, em ordem cronológica, publicadas no diário oficial do Estado do Rio de Janeiro, no período de 2011 a 2015:

O Decreto nº 43.216, de 30 de setembro de 2011, dispõe sobre Política Estadual sobre Mudança Global do Clima e Desenvolvimento Sustentável, estabelecendo a redução de 30% das emissões de gases de efeito estufa na atmosfera, tendo como base de referência as emissões produzidas no ano de 2005, até o ano de 2030, provindas do consumo de energia no setor público,, mediante ações de eficiência energética.

O Decreto nº 43.629, de 05 de junho de 2012, dispõe sobre os critérios de sustentabilidade ambiental na aquisição de bens, contratação de serviços e

obras pela Administração Pública Direta e Indireta, onde foi considerado a urgência de se implementar ações de consumo sustentável, tendo a economia no consumo de água e energia como um dos critérios de sustentabilidade.

O Decreto nº 45.109, de 05 de janeiro de 2015, dispõe sobre a reavaliação das contratações dos órgãos e entidades da administração pública do Estado do Rio de Janeiro, estabelecendo que todos os órgãos e entidades da administração pública estadual deverão reduzir o consumo das despesas correntes em, pelo menos, 20% (vinte por cento), entre as quais se destacam o consumo de energia elétrica.

O Decreto nº 45.421, de 20 de outubro de 2015, instituiu as Comissões Internas de Conservação de Insumos de Energia, Água e Telefonia (CINCONSERV), no âmbito dos órgãos e entidades da administração pública do estado, dando-lhes atribuições, tais como a elaboração de análises periódicas dos dados de consumo, estabelecimento de metas de redução de consumo, projeção do consumo, descrição das medidas a serem implementadas para o atingimento de metas de consumo estabelecidas, identificando eventuais desvios das metas específicas estabelecidas, justificando-as e sugerindo medidas para a sua correção, buscando por melhorias de eficiência energética nas 2012 edificações sob sua responsabilidade.

Considerando a publicação em Boletim da PMERJ nº 187, de 09 de outubro de 2020, que trata sobre a necessidade de contenção de despesas no âmbito da SEPM, com relação a redução de gastos com o consumo de água e energia pelas OPM, visando o cumprimento dos Decretos Estaduais mencionados, este artigo visa apresentar uma metodologia baseada em modelo de aprendizado de máquina para o problema de previsão do consumo de energia em unidades consumidoras.

Neste sentido este trabalho visa a previsão do consumo mensal de energia em kWh, antes do fechamento do período de faturamento vigente da unidade consumidora com base na medida da variação da leitura do medidor por dia, possibilitando uma comparação entre o consumo de energia do mês do ano vigente com o consumo de energia do ano anterior, e prevendo se haverá uma redução ou aumento no consumo mensal de energia.

A aplicação da metodologia pode ser uma boa prática na estratégia do acompanhamento e avaliação do consumo de energia elétrica, diante da

necessidade de redução dos gastos com o consumo de energia nas unidades consumidoras da SEPM.

2 REFERENCIAL TEÓRICO

A potência elétrica representa uma grandeza física que mede a quantidade de trabalho realizado em determinado intervalo de tempo, ou seja, é a taxa de variação da energia, sendo o watt, equivalente a um joule por segundo, a unidade de potência no Sistema Internacional de Unidades (YOUNG et al., 2009 apud BOYLESTAD, 2004).

De acordo com Halliday et al. (2012), a variação instantânea de energia é dada por:

$$dE = U(t).I(t).dt = P(t).dt \quad (1)$$

Na equação (1), $U(t)$, $I(t)$ e $P(t)$ correspondem, respectivamente, tensão, corrente e potência instantânea, em um dado período de tempo infinitesimal.

Pode-se calcular a energia consumida a partir da integração temporal da potência instantânea entre os instantes de tempos t_1 e t_2 , a partir da equação (1) como:

$$E = \int_{t_1}^{t_2} P(t).dt \quad (2)$$

Como pode ser observado na equação (2), a energia consumida depende da potência instantânea dos equipamentos elétricos que se encontram em operação em um dado instante de tempo t . Portanto, para reduzir o consumo de energia elétrica é necessário reduzir a utilização simultânea de diversos equipamentos elétricos e utilizar equipamentos com menor potência possível ou reduzir o tempo de utilização dos equipamentos.

A energia ativa é aquela que pode ser convertida em outra forma de energia (energia térmica, mecânica, sonora, etc.), expressa em quilowatts-hora (ANEEL, 2010). A energia ativa é a grandeza física utilizada para dimensionar o consumo de energia elétrica e esta é a variável a ser prevista no próximo faturamento considerando o período de faturamento vigente. O período de faturamento consiste em um intervalo de tempo de aproximadamente de 30

dias, podendo variar de um intervalo mínimo de 27 dias e máximo de 33 dias (ANEEL, 2010).

3 DESENVOLVIMENTO

Nesta seção é descrito as etapas do processo de aprendizado de máquina utilizado na solução do problema de modelagem preditiva do consumo mensal de energia de uma unidade consumidora do SEPM.

Uma forma de categorizar um sistema de aprendizado de máquina é por meio da generalização. Isso significa que, dada uma série de dados de treinamento, o sistema precisa ser capaz de generalizar em dados de validação. O aprendizado de máquina, baseado em modelo, se generaliza a partir de um conjunto de dados de treinamento, através da construção, treinamento e utilização do modelo para a realização de previsões.

As etapas do processo de aprendizado de máquina utilizado no desenvolvimento do modelo são mencionados abaixo:

- Definir o problema
- Carregar os dados
- Analisar os dados
- Preparar os dados
- Selecionar recursos
- Selecionar o modelo
- Ajustar o modelo
- Avaliar o modelo
- Prever com o modelo

3.1 DEFINIÇÃO DO PROBLEMA

O problema foi definido utilizando os seguintes questionamentos:

1. Qual é o problema?
2. Por que resolver este problema?
3. Como resolver o problema?

Neste artigo, o problema é conseguir obter uma previsão do consumo mensal de energia elétrica do mês vigente de uma unidade consumidora, dado que o período de faturamento encontra-se em aberto. A motivação principal é a

possibilidade de se poder comparar os valores previsto e medido do mesmo mês do ano anterior, avaliando o percentual de aumento ou redução do consumo mensal de energia, trazendo como benefício a possibilidade de se efetuar ações corretivas para atingir o alvo de redução nos dias restantes para o fechamento do período de faturamento.

O problema pode ser tratado como um problema de regressão, onde o objetivo é obter um método capaz de prever um valor numérico, que neste caso seria a previsão do consumo mensal de energia elétrica.

3.2 CARREGANDO OS DADOS

Nesta etapa são carregados os dados da base de dados obtidos com a extração das faturas mensais de energia elétrica de uma unidade consumidora de energia do grupo A sobre gestão da SEPM, optante pelo faturamento com aplicação de tarifa do grupo B, tendo a Enel/Rio como o agente titular da concessão federal para prestar o serviço público de distribuição de energia, registrados em formato de arquivo CSV (Comma Separated Values), onde os valores dos atributos são separados por vírgula.

Abaixo são apresentados a descrição dos 7 (sete) atributos mantidos na base de dados, após o processo de seleção dos dados:

- **A_R**: ano de referência da medição.
- **M_R**: mês de referência da medição.
- **D_L_An**: data de leitura anterior.
- **D_L_At**: data de leitura atual.
- **L_An_kWh_FP**: leitura anterior do medidor para a energia ativa (fora ponta).
- **L_At_kWh_FP**: leitura atual do medidor para a energia ativa (fora ponta).
- **C_kWh_FP**: constante multiplicativa para a energia ativa (fora ponta)
- **kWh_FP**: consumo de energia ativa (fora ponta) em kWh.

3.3 ANALISANDO OS DADOS

Nesta etapa do desenvolvimento é necessário para se ter um bom entendimento do problema de aprendizado de máquina em que está sendo trabalhado.

O objetivo da etapa de análise de dados é aumentar a compreensão do problema por meio do entendimento mínimo das medições, contidas na base de dados utilizada, antes de se prosseguir nas etapas seguintes do desenvolvimento.

Isso envolve o entendimento de diferentes maneiras de se descrever os dados como uma oportunidade de revisar e capturar observações e suposições que podem ser testadas em experimentos posteriores.

Segundo Brownlee (2016), existem duas abordagens diferentes que podem ser usadas para criar perspectivas ou visões da base de dados, obtendo insights sobre os dados:

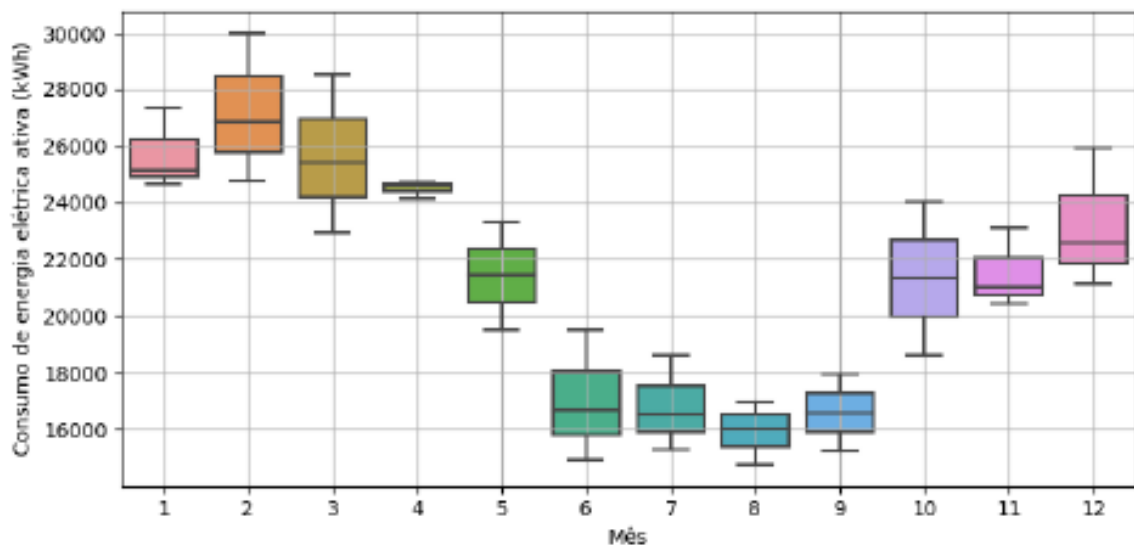
1. Resumir os dados
2. Visualizar os dados

Resumir os dados é descrever a estrutura dos dados, tais como o número de atributos e o tipo de dado de cada atributo. Este processo é importante para que seja avaliado a necessidade de transformações na etapa de preparação dos dados, como a conversão de atributos de um tipo para outro.

Na base de dados utilizada foram identificados 25 (vinte e cinco) atributos e 36 (trinta e seis) instâncias (registros) referente às medições do período de janeiro/2018 a dezembro de 2020, com dados faltantes nos meses de setembro/2019 e outubro/2019, em virtude de sua indisponibilidade, sendo identificados 2 (dois) atributos do tipo inteiro, 3 (três) atributos do tipo texto e 20 (vinte) atributos do tipo decimal.

Visualizar os dados é criar gráficos que resumem os dados, capturá-los e estudá-los para uma estrutura interessante que pode ser descrita, como histogramas e gráficos de dispersão e boxplot.

Na Figura 1 é apresentado o gráfico boxplot do consumo de energia ativa agrupado por mês, mostrando a variação das medições por meio de quartis.

Figura 1 – Gráfico boxplot do consumo de energia ativa agrupado por mês

Fonte: elaborado pelo autor.

Como pode ser observado na Figura 1, há uma tendência de aumento do consumo de energia ativa no período de verão e redução no período de inverno, mostrando que existe uma característica sazonal no atributo alvo do estudo (**kWh_FP**). Isto mostra que a análise de ocorrência de aumentos e reduções do consumo de energia ativa devem ser realizados comparando-se os consumos do ano vigente e anterior para um mês específico do ano.

3.4 Preparando os dados

Nesta etapa são preparados os dados para um algoritmo de aprendizado de máquina.

De acordo com Brownlee (2020), o processo de preparação de dados para um algoritmo de aprendizado de máquina pode ser resumido em três sub-processos:

- Selecionar os dados
- Pré-processar os dados
- Transformar os dados

Na seleção dos dados é considerado os dados que serão necessários para resolver o problema em questão. Neste sub-processo foram excluídos 17 (dezessete) dos 25 (vinte e cinco) atributos da base de dados, que foram

considerados desnecessários para a resolução do problema, sendo mantidos 7 (sete) atributos mencionados anteriormente.

Depois de selecionar os dados, a base de dados foi submetida ao processo de pré-processamento, tais como formatação, limpeza e amostragem. Neste sub-processo foram efetuados adequação do formato dos atributos **D_L_An** e **D_L_At** transformando-os em dados do tipo data, assim como suprir os atributos faltantes do mês de setembro/2019 com o cálculo da média dos meses de setembro de 2018 e 2020; e do mês de outubro/2019 com o cálculo da média dos meses de outubro de 2018 e 2020.

Para finalizar a etapa de preparação dos dados, a base de dados foi submetida ao processo de transformação dos dados, tais como dimensionamento, decomposições de atributos e agregações de atributos. Este processo também é conhecido como engenharia de recursos. Neste sub-processo foram criados novos atributos (extração de recursos) na base de dados descritos a seguir:

- dias: número de dias (período de faturamento) definido como sendo a diferença entre os atributos **D_L_At** e **D_L_An**.
- **d_l_kWh_dia**: variação média diária da leitura do medidor para a energia ativa (fora ponta) definido como sendo a média da diferença entre os atributos **L_At_kWh_FP** e **L_An_kWh_FP** pelo atributo dias.

3.5 Selecionando o(s) recurso(s)

A seleção de recurso(s) é o processo de identificação e seleção de um subconjunto de atributos de entrada que são mais relevantes para o atributo de destino (alvo).

O caso mais simples de seleção de recurso(s) é o caso em que existem atributos de entradas numéricas e uma resposta numérica para a modelagem preditiva de regressão. Isso ocorre porque a força de relação entre cada atributo de entrada e o destino pode ser calculada e comparada entre si.

De acordo com Brownlee (2020), existem duas técnicas populares de seleção de recurso(s) que podem ser usadas para dados de entrada de atributo numérico e um atributo de destino numérico:

- Estatísticas de correlação
- Estatísticas de informações mútuas

A correlação é o resumo estatístico do relacionamento entre as variáveis e pode ser calculado para diferentes tipos de variáveis e relacionamentos. Talvez a medida de correlação mais comum seja a correlação de Pearson, que assume uma distribuição gaussiana para cada variável e revela uma relação linear entre as variáveis.

Uma correlação pode ser positiva, significando que ambas as variáveis se movem na mesma direção, ou negativa, significando que quando o valor de uma variável aumenta, os valores das outras variáveis diminuem. A correlação também pode ser neutra, o que significa que as variáveis não estão relacionadas.

Segundo Kuhn et al. (2013), a abordagem clássica para quantificar cada relação com o resultado usa a estatística de correlação de amostra quando os preditores são numéricos.

Segundo Brownlee (2019), o coeficiente de correlação de Pearson pode ser usado para resumir a força da relação linear entre duas amostras de dados, sendo calculado como a covariância das duas variáveis dividida pelo produto do desvio padrão de cada amostra de dados. Trata-se da normalização da covariância entre as duas variáveis para dar uma pontuação interpretável.

O coeficiente retorna um valor entre -1 e 1 que representa os limites de correlação de uma correlação totalmente negativa para uma correlação totalmente positiva. Um valor de 0 significa que não há correlação.

O resultado do cálculo do coeficiente de correlação pode ser interpretado para entender a relação entre os atributos. valor do coeficiente abaixo de -0,5 ou acima de 0,5 indica uma correlação notável, e valores abaixo desses valores sugerem uma correlação menos notável.

Para a seleção de recursos, geralmente estamos interessados em uma pontuação positiva, onde quanto maior o valor positivo, maior o relacionamento e, mais provavelmente, o recurso deve ser selecionado para modelagem.

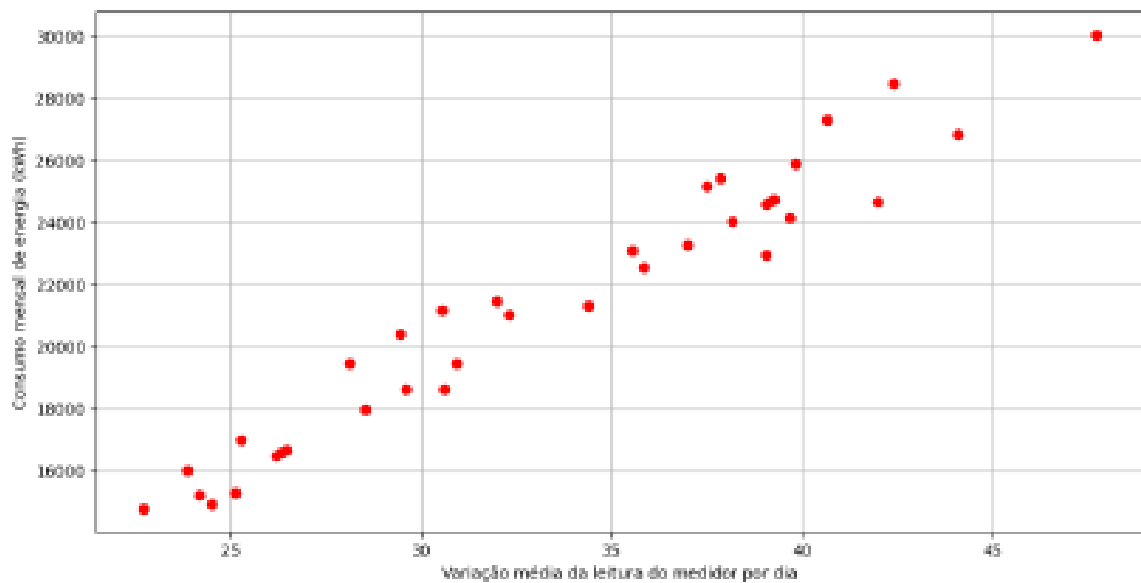
Obter uma visão sobre quais variáveis podem ou não ser relevantes como entrada para o desenvolvimento de um modelo é o ponto focal da análise para

verificação da existência de correlação entre os atributos de entrada e saída (atributo alvo da previsão).

Neste artigo foi calculado o coeficiente de correlação de Pearson entre os atributos **d_I_kWh_dia** e **kWh_FP**, obtendo-se uma correlação de 0,97, mostrando que ambos atributos estão positivamente correlacionados, sugerindo assim que há um alto nível de correlação ($0,5 < \text{correlação} < 1,0$).

Na Figura 2 é apresentado o gráfico de dispersão entre os atributos **d_I_kWh_dia** e **kWh_FP**.

Figura 2 – Gráfico de dispersão entre os atributos **d_I_kWh_dia** e **kWh_FP**



Fonte: elaborado pelo autor.

3.6 SELECIONANDO O MODELO

A seleção do modelo baseou-se na análise da Figura 2 que confirma a existência de uma forte correlação entre os atributos **d_I_kWh_dia** e **kWh_FP**. Isto sugere o desenvolvimento de um modelo preditivo baseado no modelo de regressão linear, onde a variável predita é o atributo **kWh_F** (consumo mensal de energia ativa em kWh) e a variável preditiva é o atributo **d_I_kWh_dia** (a variação média da leitura do medidor por dia).

O modelo de regressão linear é um método para modelar a relação entre uma ou mais variáveis independentes e uma variável dependente. É um elemento básico da estatística e costuma ser considerado um bom método introdutório de aprendizado de máquina.

A regressão linear é um método para modelar a relação entre dois valores escalares: a variável de entrada (**d_I_kWh_dia**) e a variável de saída (**kWh_F**).

O modelo assume que o atributo **kWh_F** é uma função linear do atributo **d_I_kWh_dia** definida por:

$$\mathbf{kWh_FP} = a * \mathbf{d_I_kWh_dia} + b \quad (3)$$

O modelo de regressão linear definido na equação (3) utiliza os coeficientes a e b. O objetivo neste ponto da modelagem é encontrar os valores dos coeficientes que minimizam o erro na previsão da variável de saída **kWh_F**.

A maneira de se encontrar os valores dos coeficientes é encontrar a solução em que os valores dos coeficientes do modelo minimizem o erro quadrático, e isto pode ser obtido com o método de aproximações de mínimos quadrados.

Segundo Strang (2016) nem sempre é possível obtermos o erro próximo de zero, mas quando o erro é o menor possível, a variável de entrada é uma solução de mínimos quadrados.

3.7 AJUSTANDO O MODELO

O modelo de regressão linear foi ajustado utilizando o conjunto de dados de treinamento, contendo 28 (vinte e oito) instâncias, correspondendo a 80% do total de amostras da base de dados. O modelo ajusta um modelo linear com os coeficientes, minimizando a soma residual dos quadrados entre alvos observados do conjunto de dados de treinamento e os alvos previstos pela aproximação linear por meio da otimização pelo métodos dos mínimos quadrados, sendo a abordagem mais eficiente para encontrar os coeficientes que minimizam o erro do modelo.

3.8 AVALIANDO O MODELO

Segundo Brownlee (2016), é comum a utilização de 3 (três) métricas para avaliar as previsões sobre problemas de aprendizado de máquina de regressão:

- Erro médio absoluto
- Erro médio quadrático
- Coeficiente de determinação (R2)

Neste trabalho foi utilizado o coeficiente de determinação (R^2) como métrica de avaliação do modelo preditivo. Esta métrica fornece um valor entre 0 (não ajuste) e 1 (ajuste perfeito), indicando a qualidade do ajuste de um conjunto de previsões aos valores reais.

A definição mais geral do coeficiente de determinação é:

$$R^2 = 1 - u/v \quad (4)$$

Onde u é a soma dos quadrados residuais e v é a soma total dos quadrados.

A soma dos quadrados residuais é dado por:

$$u = \sum_{i=1}^n (y_i - f_i)^2 \quad (4)$$

Onde y_i é o i -ésimo valor medido (real) do atributo **kWh_F** e f_i é o i -ésimo valor predito do atributo **kWh_F** pela Equação 3.

A soma total dos quadrados é dado por:

$$v = \sum_{i=1}^n (y_i - y)^2$$

Onde y_i é o i -ésimo valor medido (real) do atributo **kWh_F** e y é média dos dados observados do atributo **kWh_F**.

A qualidade do ajuste do modelo de regressão linear foi avaliada utilizando o conjunto de dados de validação, contendo 8 (oito) instâncias, não utilizadas na etapa de ajuste do modelo, correspondendo a 20% do total de amostras da base de dados.

3.9 PREVENDO COM O MODELO

Esta etapa do desenvolvimento baseia-se na utilização do modelo de regressão linear após a conclusão das etapas de ajuste e avaliação.

As previsões do consumo mensal de energia (kWh) foram obtidas utilizando-se as 8 (oito) instâncias do conjunto de dados de validação como entrada no modelo de regressão linear.

4 METODOLOGIA

A metodologia adotada foi baseada nas etapas do processo de aprendizado de máquina apresentadas na seção de desenvolvimento, onde foi implementado códigos na linguagem Python dentro do ambiente de programação de código aberto e gratuito, PyCharm Community versão 2020.3. O PyCharm é um ambiente de desenvolvimento integrado usado em programação de computadores, especificamente para a linguagem Python.

A linguagem Python possui uma vasta comunidade, que disponibiliza bibliotecas Python para diversas áreas da tecnologia. As bibliotecas Python utilizadas para implementação dos códigos Python em cada etapa do processo de aprendizado de máquina no PyCharm foram:

- **Numpy:** pacote básico da linguagem Python que permite trabalhar com arranjos, vetores e matrizes de N dimensões, de uma forma comparável e com uma sintaxe semelhante ao software proprietário Matlab
- **Pandas:** biblioteca da linguagem Python para manipulação e análise de dados.
- **Matplotlib:** biblioteca da linguagem Python para criação de gráficos e visualizações de dados em geral.
- **Scipy:** biblioteca da linguagem Python que implementa diversas técnicas úteis na computação científica.
- **scikit-learn:** biblioteca da linguagem Python desenvolvida especificamente para aplicação prática de machine learning. Esta biblioteca dispõe de ferramentas simples e eficientes para análise preditiva de dados.

No Anexo A é apresentado o código principal na linguagem Python que implementa as etapas do processo de aprendizado de máquina utilizado na solução do problema de modelagem preditiva do consumo mensal de energia de uma unidade consumidora do SEPM.

5 RESULTADOS

O modelo de regressão linear que ajustou-se o mais próximo possível do conjunto de dados de treinamento é determinado pela equação (3), onde os

coeficientes a e b , foram determinados com a otimização pelos métodos dos mínimos quadrados, sendo obtidos os valores 594, 16 e 1.398,46, respectivamente.

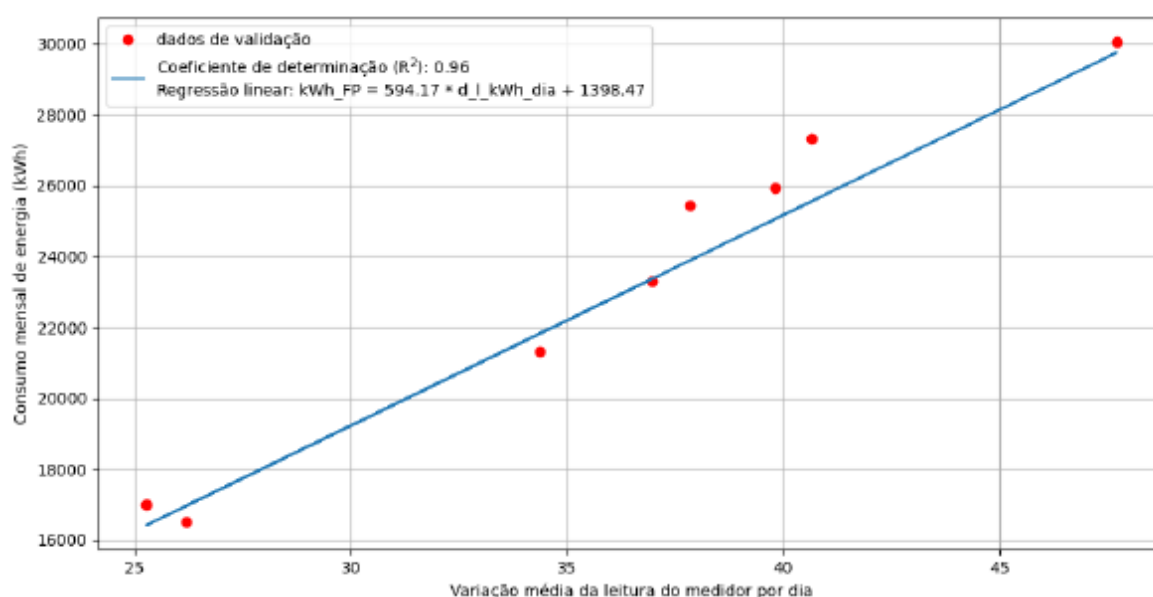
O coeficiente de determinação (R^2) obtido foi de 0,96 (96%), sendo considerado um percentual que reflete a boa qualidade do ajuste do modelo preditivo, indicando que o modelo é capaz de generalizar para novos dados a serem fornecidos

A Figura 3 apresenta o gráfico de dispersão entre os atributos **d_l_kWh_dia** e **kWh_FP** (pontos vermelhos) e o gráfico de linha do modelo de regressão linear (linha azul).

Figura 3 – Gráfico do modelo de regressão linear e dados de validação

Fonte: elaborado pelo autor.

A Tabela 1 apresenta a diferença percentual entre os dados de validação (valores reais medidos) e os valores previstos do consumo mensal de energia



ativa (kWh). Observa-se que as diferenças percentuais negativas demonstram que os valores previstos foram menores do que os valores reais medidos, não sendo menor do que 7% e as diferenças percentuais positivas demonstram que os valores previstos foram maiores do que os valores reais medidos, não sendo maiores do que 3%.

A Tabela 1 apresenta a diferença percentual entre os dados de validação (valores reais medidos) e os valores previstos do consumo mensal de energia ativa (kWh). Observa-se que as diferenças percentuais negativas demonstram que os valores previstos foram menores do que os valores reais medidos, não sendo menor do que 7% e as diferenças percentuais positivas demonstram que os valores previstos foram maiores do que os valores reais medidos, não sendo maiores do que 3%.

Tabela 1 – Diferença percentual entre os dados de validação e os valores previstos

Mês/Ano	Consumo de energia (kWh)		Diferença (%)
	Valor medido	Valor previsto	
Jan/2019	27.321	25.556	-6,5
Fev/2019	30.051	29.741	-1,0
Mai/2019	23.289	23.363	0,3
Out/2019	21.336	21.824	2,3
Mar/2020	25.431	23.884	-6,1
Jul/2020	16.506	16.966	2,8
Ago/2020	16.989	16.420	-3,3
Dez/2020	25.914	25.051	-3,3

Fonte: elaborado pelo autor.

A Tabela 2 apresenta a diferença percentual entre os valores reais dos meses do ano anterior e os valores preditos (Tabela 1) dos meses do ano vigente. Observa-se que as diferenças percentuais negativas preveem a ocorrência de redução do consumo e as diferenças percentuais positivas preveem a ocorrência de aumento do consumo.

Tabela 2 – Diferença percentual entre os valores medidos do ano anterior e previstos no ano vigente

Mês/Ano	Consumo de energia (kWh)				Diferença (%)
	Ano	Valor medido	Ano	Valor previsto	
Janeiro	2018	25.179	2019	25.556	1,5
Fevereiro	2018	24.728	2019	29.741	16,8
Mai	2018	21.483	2019	23.363	8,0
Outubro	2018	18.627	2019	21.824	14,6
Março	2019	28.476	2020	23.884	-19,3
Julho	2019	18.648	2020	16.966	-9,9
Agosto	2019	14.784	2020	16.420	9,9
Dezembro	2019	21.168	2020	25.051	15,5

Fonte: elaborado pelo autor.

A Tabela 3 apresenta a diferença percentual entre os valores reais dos meses do ano anterior e os valores medidos (Tabela 2) dos meses do ano vigente. Observa-se que as diferenças percentuais negativas confirmam a ocorrência de redução do consumo e as diferenças percentuais positivas confirmam a ocorrência de aumento do consumo, apesar das diferenças percentuais obtidas em ambas as tabelas terem valores ligeiramente diferentes.

Tabela 3 – Diferença percentual entre os valores medidos do ano anterior e do ano vigente

Mês/Ano	Consumo de energia (kWh)				Diferença (%)
	Ano	Valor medido	Ano	Valor previsto	
Janeiro	2018	25.179	2019	27.321	7,8
Fevereiro	2018	24.728	2019	30.051	17,7
Mai	2018	21.483	2019	23.289	7,7
Outubro	2018	18.627	2019	21.336	12,7
Março	2019	28.476	2020	25.431	-11,9
Julho	2019	18.648	2020	16.506	-12,9
Agosto	2019	14.784	2020	16.989	12,9
Dezembro	2019	21.168	2020	25.914	18,3

Fonte: elaborado pelo autor.

Comparando as diferenças percentuais da Tabela 2 e Tabela 3, pode-se observar que o modelo preditivo foi capaz de prever a ocorrência de aumento e redução do consumo.

6 CONCLUSÕES

Pode-se concluir que a partir das técnicas de análises exploratórias de dados pode-se extrair insights para uso posterior em técnicas de aprendizado de máquina para a previsão baseada em dados. O modelo regressão linear pode ser uma ferramenta eficaz no acompanhamento sistemático do consumo mensal de energia, obtendo uma previsão a partir da variação média da leitura do medidor por dia, e posteriormente, podendo se comparar o valor previsto do mês do ano vigente com o valor real do mês do ano anterior, prevendo a ocorrência de aumento ou redução no consumo, que em caso de previsão de aumento poderão ser adotadas medidas corretivas ainda no ciclo vigente ou posterior de faturamento.

REFERÊNCIAS BIBLIOGRÁFICAS

BOYLESTAD, Robert L. **Introdução à análise de circuitos**. 10ª ed. São Paulo: Pearson Education do Brasil, 2004.

YOUNG, Hugh D; Freedman, Roger A. **Física III: Eletromagnetismo**. 12ª ed. São Paulo: Addison Wesley, 2009.

HALLIDAY, David; Resnick, Robert; Walker, Jearl. **Fundamentos de Física 3**. 9ª ed. Rio de Janeiro: LTC, 2012.

BROWNLEE, Jason. **Machine Learning Mastery with Python: Understand Your Data, Create Accurate Models and Work Projects End-To-End**. Machine Learning Mastery, 2016.

BROWNLEE, Jason. **Statistical Methods for Machine Learning: Discover how to Transform Data into Knowledge with Python**. Machine Learning Mastery, 2019.

BROWNLEE, Jason. **Data Preparation for Machine Learning: Data Cleaning, Feature Selection and Data Transforms in Python**. Machine Learning Mastery, 2020.

KUHN M., Johnson K. **Applied Predictive Modeling**. 1st edition. Springer, 2013

STRANG, Gilbert. **Introduction to Linear Algebra**. 5th edition. Wellesley-Cambridge Press, 2016.