



Latin American Journal of Energy Research – Lajer (2024) v. 11, n. 1, pp. 167–179  
<https://doi.org/10.21712/lajer.2024.v11.n1.p166-179>

**Avaliação dos mecanismos de previsão energética com fontes renováveis para maximização da matriz energética brasileira utilizando *machine learning***  
*Evaluation of energy forecasting mechanisms with renewable sources for maximizing the brazilian energy matrix using machine learning*

Laís Gomes Barreto Abreu<sup>1,\*</sup>, Diego Mauricio Yepes Maya<sup>2</sup>, Vladimir Rafael Melian Cobas<sup>2</sup>,  
Juan José Garcia Pabón<sup>2</sup>, Nestor Proenza Perez<sup>3</sup>

<sup>1</sup> Aluno do Programa de Pós-Graduação em Engenharia de Energia, Universidade Federal de Itajubá – Unifei, campus Itajubá, MG, Brasil

<sup>2</sup> Professor do Programa de Pós-Graduação em Engenharia de Energia, Universidade Federal de Itajubá – Unifei, campus Itajubá, MG, Brasil

<sup>3</sup> Professor do Departamento de Química e Energia, Universidade Estadual Paulista – Unesp/FEG, campus Guaratinguetá, SP, Brasil

\*Autor para correspondência, E-mail: engproducao.lais@gmail.com

Received: 17 April 2024 | Accepted: 27 May 2024 | Published online: 5 June 2024

**Resumo:** O presente estudo concentra-se no cenário energético brasileiro e destaca o aumento progressivo do uso de fontes renováveis na matriz de energia elétrica do país. O principal objetivo deste trabalho é contribuir para a busca de soluções e impulsionar debates e reflexões sobre as ações futuras necessárias para o planejamento energético. Para isso, a pesquisa emprega ferramentas computacionais baseadas em *machine learning* e mineração de dados, utilizando fontes de dados governamentais e de mercado de energia. A metodologia abrange a utilização de ferramentas computacionais para projetar a previsão do mercado de energia elétrica no país, inclui a execução de modelos de previsão, destacando o comportamento do mercado energético ao longo do tempo, a partir dos métodos Redes Neurais *Multilayer Perceptron* (MLP), Regressão do Processo Gaussiano (GPR) e Regressão Linear para projetar a geração elétrica por fonte no Brasil. Os resultados indicam um crescimento considerável das fontes renováveis no mercado energético nacional até o ano de 2030, aproximando-se do objetivo do Plano Decenal de Expansão de Energia de atingir 90% de renovabilidade, abrangendo fontes como hidrelétrica, biomassa, eólica e solar. O método de Regressão Linear alcançou 86% de renovabilidade, enquanto o método de GPR atingiu 90%, e o método das MLP chegou a 88%. A projeção da previsão do mercado de energia elétrica, possibilitou a identificação dos padrões de comportamento mercadológico, permitindo antecipar as tendências e mudanças no mercado. Essas previsões têm o propósito de fornecer informações para apoiar o desenvolvimento de ações no processo de planejamento energético, contribuindo para a transição para fontes mais sustentáveis e renováveis de energia no Brasil.

Palavras-chave: Planejamento Energético, *machine learning*, mineração de dados, energia elétrica, métodos de previsão.

**Abstract:** The present study focuses on the Brazilian energy scenario and highlights the progressive increase in the use of renewable sources in the country's electricity matrix. The main objective of this study is to contribute to the search for solutions and to stimulate debates and reflections on the future actions necessary for energy planning. To achieve this, the research employs computational tools based on machine learning and data mining, using government and energy market data sources. The research methodology encompasses the use of computational tools to project the forecast of the electricity market in the country. The methodology employed includes the execution of forecasting models, highlighting the behavior of the energy market over time, using methods such as Multilayer Perceptron Neural Networks (MLP), Gaussian Process Regression (GPR), and Linear Regression to project electricity generation by source in Brazil. The results indicate a considerable growth of renewable sources in the national energy market by the year 2030, approaching the goal of the Ten-Year Energy Expansion Plan to achieve 90% renewability, covering

*sources such as hydroelectric, biomass, wind, and solar. The Linear Regression method achieved 86% renewability, while the GPR method reached 90%, and the MLP method reached 88%. The projection of the electricity market forecast allowed for the identification of market behavior patterns, enabling the anticipation of trends and changes in the market. These forecasts aim to provide information to support the development of actions in the energy planning process, contributing to the transition to more sustainable and renewable energy sources in Brazil.*

*Keywords: Energy Planning, Machine Learning, Data Mining, Electric Power, Forecasting Methods.*

## 1 Introdução

A população e economia mundial se propagaram, a partir da década de 70, ocasionando um crescimento exponencial na demanda energética global, orientado, principalmente, pelo uso do petróleo. E, nos últimos anos, a dependência da energia determina o processo da economia a partir do petróleo, gás natural, carvão, energia nuclear e hidroeletricidade (Barros, 2007).

Com a diminuição das reservas de petróleo, associado ao grande consumo e a sua dependência, outras fontes de energia serão decisivas para o desenvolvimento da economia global. Ao avaliar as alternativas energéticas substitutas ao petróleo, as fontes de energias renováveis possuem custos de produção mais elevados e, em alguns casos, demandam uma quantidade de energia maior para serem produzidas, com relação à extração de petróleo (Castro e Lyra Filho, 2005; Barros, 2007).

As fontes renováveis estarão de forma complementar, na produção de energia, posto que não possuem disponibilidade suficiente, tecnologia desenvolvida e viabilidade ambiental comprovada para que ocorra a substituição dos combustíveis fósseis, de forma expressiva a suprir as demandas mundiais de energia (Ventura Filho, 2014).

Estudos desenvolvidos recentemente apontam que, quando na ausência de alternativas, os preços elevados dos combustíveis fósseis acarretam pobreza energética e perda de competitividade, assim, transcorre da vontade e determinação política moldar e determinar o horizonte para a transição que conduzirá para um mundo mais inclusivo, equitativo e estável (IRENA, 2022).

A eletricidade oriunda de fontes de energia renovável, atualmente, é considerada como a energia mais barata em grande parte das regiões, tendo como característica a opção de expansão da capacidade do setor energético em muitos países, dominando, assim, os investimentos atuais, pelo qual, o desenvolvimento necessita de políticas, investimentos e tecnologias para sua aplicabilidade e redução do impacto econômico e social (IRENA, 2022).

Portanto, o planejamento e a regulamentação do setor energético devem buscar mecanismos de suprimento compatíveis com as potencialidades e necessidades socioeconômicas, aproveitando estrategicamente cada fonte energética, garantindo a maximização dos benefícios e minimização dos impactos negativos gerados, onde, por meio dos resultados da previsão de demanda futura torna-se possível o planejamento do setor (Zarur, 2005; Moret, 2014; Gonzalez-Briones et al., 2019).

O Brasil é reconhecido como o terceiro país da América Latina em ter em sua matriz energética maior participação de energias renováveis, sua matriz energética diferencia-se da matriz mundial em decorrência da utilização em abundância de fontes renováveis, especialmente a biomassa, o vento, o sol e a hidroeletricidade (Pereira, 2014; Ventura Filho, 2014; Korzeniewicz, 2021).

Ademais, o Brasil detém um dos maiores e melhores potenciais energéticos do mundo, ainda que as reservas de combustíveis fósseis são consideravelmente reduzidas, os potenciais hidráulicos, solar, biomassa e eólico são suficientemente abundantes, garantindo a autossuficiência energética do país (Zarur, 2005; Gehrke, Goretti e Ávila, 2021; Oliveira, 2022).

A fonte hidráulica e o petróleo detêm grande importância para o setor elétrico brasileiro, entretanto, com o crescimento da demanda, insuficiência de oferta e limitações financeiras, socioeconômicas e ambientais à expansão, o suprimento de energia elétrica no futuro irá demandar o aproveitamento de fontes energéticas alternativas (Zarur, 2005; Basso, 2019; Noguchi e Nobre, 2023).

O estímulo ao desenvolvimento de energias renováveis no Brasil visa aumentar a variedade de fontes energéticas, além das hidrelétricas, objetivando a garantia da segurança no abastecimento de energia, reduzindo impactos negativos sobre as comunidades e o meio ambiente e diminuição da dependência excessiva das reservas hídricas do país (Noguchi e Nobre, 2023).

Mas, frente ao cenário atual do mercado de energia elétrica e de cara para uma transição energética sustentável, quais os mecanismos estratégicos a serem desenvolvidos para o crescimento das fontes renováveis na matriz energética?

Mediante ao cenário energético brasileiro, o crescimento gradual do uso de fontes renováveis na matriz de energia elétrica e o potencial de crescimento, o intuito ao se esboçar diferentes perspectivas a um determinado período torna-se cabível a contribuição para possíveis soluções, impulsionando o debate e a reflexão sobre ações futuras, base do exercício de planejamento.

Assim, o estudo objetiva realizar a projeção da previsão do mercado de energia elétrica no Brasil, por meio da mineração de dados e ferramentas computacionais do tipo *machine learning*, tendo por propósito a criação de modelos de previsão que estimem como o mercado de energia elétrica no Brasil evoluirá ao longo do tempo por projeções futuras, a partir dos dados históricos das fontes de dados governamentais e do mercado de energia, com o intuito de fornecer informações fiáveis para o setor de planejamento energético.

## 2 Mineração de dados como ferramenta de análise

A mineração de dados aplicada à análise de dados é uma abordagem poderosa para descobrir informações e padrões ocultos em conjuntos de dados complexos. Caracterizada como uma ferramenta de descoberta de conhecimento em base de dados, pela extração de informações preliminarmente desconhecidas e possivelmente usuais em grandes bases de dados, pela utilização de diferentes abordagens técnicas (Zarur, 2005; Witten, Frank, Hall E Pal, 2016).

Pela mineração de dados alcança-se informações e conhecimentos de apoio a tomada de decisão, predição, previsão e estimação, de tal modo que pode ser aplicado em inúmeras áreas, sob duas formas, a orientada à verificação e à descoberta (Zarur, 2005).

A aplicação da mineração de dados divide-se em quatro etapas básicas, inicialmente, pela seleção de dados a serem minerados, segmentando-os sob alguns critérios; pela transformação de dados, ocorrendo a organização dos dados e definição de novos atributos; seguidamente pela mineração dos dados, na aplicação de técnicas de extração de informações; e, por fim, a interpretação dos resultados, desenvolvendo a análise das informações alcançadas para suporte à decisão e objetivos (Tan, Steinbach e Kumar, 2006).

### 2.1 Método de previsão

O método de previsão auxilia na identificação do que ainda não é conhecido, correspondendo a um método que se baseia na análise de dados históricos para identificar padrões e tendências que podem ser extrapolados para prever valores futuros (Zarur, 2005; Hyndman e Athanasopoulos, 2018).

Considerando a série temporal como uma descrição do passado, o procedimento lógico para realização de previsões futuras aplica-se ao uso dos dados históricos, uma vez que dados passados corresponde a indicativos do que se espera do futuro, por meio de modelo matemático representativo do processo (Zarur, 2005).

A extração de conhecimento das séries de dados históricas, por meio da aplicação de técnicas de mineração de dados proporcionam o descobrimento de padrões futuros para o exercício de planejamento, na qual, as principais etapas da previsão por extrapolação de tendências abrangem (Makridakis et al., 1998; Shumway e Stoffer, 2017; Hyndman e Athanasopoulos, 2018):

**Coleta de Dados Históricos:** A primeira etapa envolve a coleta de dados históricos relevantes para o fenômeno que está sendo previsto.

**Análise de Dados:** Os dados históricos são analisados para identificar tendências e padrões.

**Seleção de Modelo de Tendência:** Com base na análise, é selecionado um modelo de tendência que melhor se ajusta aos dados.

**Estimação de Parâmetros:** Os parâmetros do modelo de tendência pela definição dos coeficientes que melhor se ajustam aos dados históricos.

**Extrapolação Futura:** Com os parâmetros do modelo estimados, a extrapolação da tendência para o futuro envolve a projeção de valores futuros com base na equação do modelo de tendência.

**Avaliação do Erro:** A precisão da previsão é avaliada comparando-se os valores previstos com os valores reais que ocorrem no futuro, aplicando as métricas para avaliar o desempenho da previsão.

#### 2.1.1 Modelo de redes neurais *multilayer perception*

As Redes Neurais *Multilayer Perceptron* (MLP) são um tipo popular de rede neural artificial que consiste em múltiplas camadas de neurônios interconectados. Cada camada é composta por uma série de neurônios que processam informações e aprendem a partir dos dados (Aggarwal, 2018).

A arquitetura das MLPs consiste em uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada neurônio em uma camada está conectado a todos os neurônios na camada subsequente. As camadas ocultas permitem que as MLPs capturem relações complexas nos dados (Goodfellow, Bengio e Courville, 2016).

A fórmula geral para calcular a saída de um neurônio em uma MLP é uma combinação linear das entradas ponderadas, passando por uma função de ativação. No qual, as fórmulas-chave para o cálculo da saída de um neurônio em uma camada oculta e na camada de saída de uma MLP corresponde a (Goodfellow, Bengio e Courville, 2016):

Para um neurônio na camada oculta de uma MLP, a fórmula para calcular a saída ( $z$ ) é dada pela Eq. (1)

$$z = \sum_{i=1}^n (x_i \cdot \omega_i) + b, \quad (1)$$

onde  $z$  é a soma ponderada das entradas,  $x_i$  são as entradas da camada anterior,  $\omega_i$  são os pesos associados às entradas e  $b$  é o viés (bias).

Posteriormente ao cálculo de  $z$ , a saída da camada oculta ( $a$ ) é obtida pela aplicação de uma função de ativação ( $f$ ), dada pela Eq. (2)

$$a = f(z), \quad (2)$$

Para a camada de saída de uma MLP, a fórmula é semelhante, mas a função de ativação pode variar dependendo do tipo de problema que a rede está resolvendo.

Nas redes MLP, o processo de treinamento envolve o ajuste dos pesos ( $\omega_i$ ) e do viés ( $b$ ) para minimizar o erro entre as saídas da rede e os valores desejados, onde, para redes complexas podem envolver várias camadas ocultas e diferentes funções de ativação, dependendo da tarefa em questão (Goodfellow, Bengio e Courville, 2016).

### 2.1.2 Modelo de regressão do processo gaussiano

A Regressão do Processo Gaussiano corresponde a um método de aprendizado de máquina não paramétrico que é usado para modelar relacionamentos não lineares entre variáveis. Ela se baseia em processos estocásticos e é uma técnica de regressão bayesiana (Rasmussen e Williams, 2006).

O GPR modela a relação entre uma variável de entrada  $X$  e uma variável de saída  $Y$ . No qual, supondo um conjunto de dados de treinamento  $(X_{train}, Y_{train})$ , onde  $X_{train}$  é uma matriz de entradas e  $Y_{train}$  é o vetor de saída correspondente, tem-se por objetivo prever  $Y$  para novos valores de  $X$  (Rasmussen e Williams, 2006).

A formulação básica do GPR, apresentada pela Eq. (3), envolve a construção de um modelo probabilístico da função desconhecida que mapeia  $X$  para  $Y$ . O GPR assume que a saída  $Y$  segue uma distribuição gaussiana condicional (Rasmussen e Williams, 2006)

$$Y|X \sim N ( f(X), \sigma^2 I), \quad (3)$$

onde  $f(X)$  é uma função desconhecida que modela a relação  $X$  e  $Y$ ,  $\sigma^2$  é a variância do ruído associado à observação de  $Y$  e  $I$  é a matriz identidade.

A função  $f(X)$  é modelada como um processo gaussiano, o que significa que qualquer subconjunto das variáveis aleatórias  $f(X_1), f(X_2), \dots$  segue uma distribuição gaussiana multivariada (Rasmussen e Williams, 2006).

### 2.1.3 Modelo de regressão linear

A regressão linear é um método estatístico amplamente utilizado para modelar a relação entre uma variável de resposta ou variável dependente e uma ou mais variáveis explicativas ou variáveis independentes (Moore, McCabe e Craig, 2014; Witten, Frank, Hall e Pal, 2016).

A regressão linear simples modela a relação entre uma variável de resposta  $Y$  e uma única variável explicativa  $X$ . A fórmula da regressão linear simples corresponde a representada pela Eq. (4) (Moore, McCabe e Craig, 2014)

$$Y = \beta_0 + \beta_1 X + \varepsilon, \quad (4)$$

onde  $Y$  é a variável de resposta,  $X$  é a variável explicativa,  $\beta_0$  é o coeficiente linear (*intercept*),  $\beta_1$  é o coeficiente de inclinação (*slope*) e  $\varepsilon$  é o erro aleatório.

A tarefa da regressão linear é encontrar os valores de  $\beta_0$  e  $\beta_1$  que minimizam a soma dos quadrados dos resíduos, ou seja, a diferença entre os valores observados e os valores previstos.

## 2.2 Métricas

As métricas RMSE (*Root Mean Square Error*), MAE (*Mean Absolute Error*) e MAPE (*Mean Absolute Percentage Error*) são comumente usadas para avaliar a qualidade de modelos de previsão, como modelos de regressão ou previsão de séries temporais. Elas fornecem insights sobre o quão bem um modelo se ajusta aos dados observados e o quão precisa é a previsão (Hyndman e Koehler, 2006; Montgomery, Peck e Vining, 2021).

O RMSE corresponde a uma métrica que calcula a raiz quadrada da média dos erros quadrados entre as previsões do modelo e os valores observados, onde quanto menor o RMSE, melhor o modelo está se ajustando aos dados, resultando pela Eq. (5) (Montgomery, Peck e Vining, 2021)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (5)$$

onde  $n$  é o número de observações,  $y_i$  é o valor observado e  $\hat{y}_i$  é o valor previsto pelo modelo.

O MAE refere-se a uma métrica que calcula a média dos valores absolutos dos erros entre as previsões do modelo e os valores observados, fornecendo uma medida simples da magnitude dos erros do modelo. Ele é menos sensível a outliers em comparação com o RMSE. Tendo para o MAE a Eq. (6) (Montgomery, Peck e Vining, 2021)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (6)$$

onde  $n$  é o número de observações,  $y_i$  é o valor observado e  $\hat{y}_i$  é o valor previsto pelo modelo.

O MAPE corresponde a métrica que calcula a média das porcentagens absolutas dos erros em relação aos valores observados. Ele expressa os erros como uma porcentagem dos valores reais, o que pode ser útil para compreender o erro relativo do modelo, sendo especialmente útil quando se deseja avaliar a precisão do modelo em termos percentuais em relação aos valores reais. O MAPE representa pela Eq. (7) (Hyndman e Koehler, 2006)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left( \frac{|y_i - \hat{y}_i|}{|y_i|} \right) \cdot 100\%, \quad (7)$$

onde  $n$  é o número de observações,  $y_i$  é o valor observado e  $\hat{y}_i$  é o valor previsto pelo modelo.

## 3 Materiais e métodos

O objeto de estudo refere-se a Matriz Energética e Elétrica brasileira, constituindo as inúmeras formas de geração de energia, objetivando a aplicação de instrumentos para previsão da intensificação de energia renováveis mediante ao alcance do ambiente de expectativa para uma transição energética confiável.

O Anuário Estatístico de Energia Elétrica apresenta as informações relativas à cadeia de energia elétrica do Brasil, sendo extraído os dados históricos referentes a Geração elétrica por fonte no Brasil (GWh), conforme Tabela 1.

A definição da aplicação dos modelos de Redes Neurais *Multilayer Perceptron* (MLP), Regressão do Processo Gaussiano (GPR) e Regressão Linear resultarão na estimativa de previsão futura, proporcionando deste modo a leitura e análise dos dados acerca do mercado de energia nacional. Assim como, a definição dos limites a serem extrapolados considera o horizonte no Plano Decenal de Expansão de Energia 2030, para se seja factível, a efeito de comparação, sendo realizado o alinhamento ou discrepância entre as previsões do modelo e as expectativas do plano.

Tabela 1. Geração elétrica por fonte no Brasil (GWh). Fonte: EPE (2021).

Fonte	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Total	<b>552.498</b>	<b>570.835</b>	<b>590.542</b>	<b>581.486</b>	<b>578.898</b>	<b>587.962</b>	<b>601.396</b>	<b>626.324</b>	<b>621.250</b>	<b>656.108</b>	<b>677.173</b>
Hidráulica (i)	415.342	390.992	373.439	359.743	380.911	370.906	388.971	397.877	396.381	362.818	427.114
Gás Natural	46.679	68.942	81.073	79.503	56.550	65.591	54.295	60.188	53.515	86.861	42.035
Derivados de Petróleo (ii)	16.293	22.223	31.529	25.708	12.207	12.911	10.293	7.846	8.556	18.244	7.816
Carvão	8.422	14.801	18.385	19.096	17.001	16.257	14.204	15.327	11.946	17.585	7.988
Nuclear	16.038	15.450	15.378	14.734	15.864	15.739	15.674	16.129	14.053	14.705	14.559
Biomassa (iii)	34.706	39.684	44.987	47.394	49.236	49.385	51.876	52.111	55.613	51.711	51.783
Eólica	5.050	6.578	12.210	21.626	33.489	42.373	48.475	55.986	57.051	72.286	81.632
Solar	2	5	16	59	85	831	3.461	6.651	10.748	16.752	30.126
Outras (iv)	9.966	12.160	13.524	13.623	13.554	13.968	14.147	14.210	13.387	15.146	14.121

Notas:

- (1) Inclui PCH, CGH e autoprodução;
- (2) Derivados de petróleo: óleo diesel e óleo combustível;
- (3) Biomassa: lenha, bagaço de cana e lixo;
- (4) Outras: gás de coqueria, outras não renováveis e outras renováveis.

A primeira etapa da mineração de dados, consiste em obter os dados referentes à série histórica a ser analisada da geração elétrica por fonte, reescrevendo as grandezas para facilitar a visualização das informações a serem extraídas, subdividindo-as quatro frentes, as usinas hidrelétricas, as fontes renováveis, as não renováveis e outras fontes.

As usinas hidrelétricas correspondem as fontes hidráulica, incluindo PCH, CGH e autoprodução, as não renováveis abrangem as fontes de gás natural, carvão, nuclear e derivados de petróleo, como óleo diesel e óleo combustível. Para as renováveis, são consideradas as fontes eólica, solar e biomassa, como lenha, bagaço de cana e lixo e a categoria outras corresponde a fontes de gás de coqueria, outras não renováveis e outras renováveis.

Posteriormente, mediante a adequação para entrada no software WEKA, os dados foram convertidos para o formato ARFF (*Attribute-Relation File Format*), objetivando, delimitar o conjunto de observações analisadas, sendo a estrutura reescrita para definição dos atributos analisados. O WEKA, corresponde a uma suíte de software de mineração de dados e aprendizado de máquina desenvolvida pela Universidade de Waikato, na Nova Zelândia. Criado em Java, é uma ferramenta de código aberto amplamente utilizada em pesquisa e aplicações práticas de análise de dados, utilizado para mineração de dados e aprendizado de máquina, oferecendo uma variedade de ferramentas e algoritmos que permitem aos pesquisadores e profissionais analisarem dados, criar modelos preditivos e tomar decisões baseadas em dados.

A terceira etapa, consiste no início do uso do software WEKA, pela abertura e carregamento do arquivo ARFF na aba *Preprocess*. E quarta etapa refere-se à configuração dos parâmetros, abrangendo a seleção de variáveis objetivo, o número de unidades de tempo a ser previsto, a variável de referência temporal e a periodicidade da série temporal, assim como, a seleção das métricas a serem apresentadas, o intervalo de defasagem e a seleção das métricas de avaliação, conforme Tabela 2.

Tabela 2. Parâmetros dos softwares.

Parâmetro	Dado
Seleção de variáveis objetivo	Usinas hidrelétricas, Não renováveis, Renováveis, Outras
Número de unidades de tempo a ser previsto	8
Variável de referência temporal	ano
Periodicidade da série temporal	anual
Intervalo de confiança	95%
Intervalo de defasagem	5
Métricas de avaliação	RMSE, MAE, MAPE

Por conseguinte, após definição dos parâmetros, foi realizada a previsão dos modelos de Redes *Neurais Multilayer Perceptron* (MLP), Regressão do Processo Gaussiano (GPR) e Regressão Linear, para análise dos padrões de comportamento futuro da geração elétrica por fonte no Brasil.

O diagrama de blocos para a projeção da previsão da geração elétrica por fonte no Brasil, por meio da mineração de dados e ferramentas computacionais do tipo *machine learning* ocorreu conforme a Figura 1.

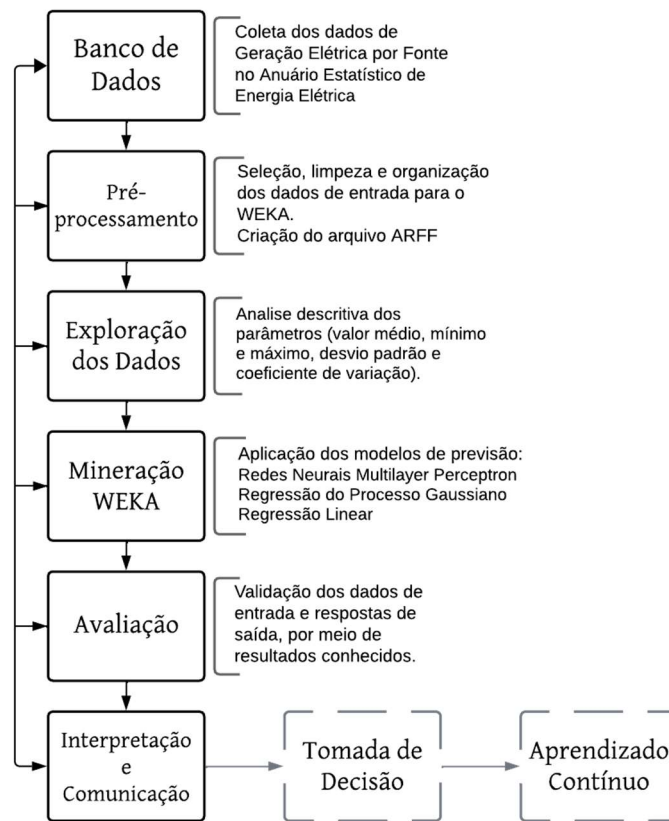


Figura 1. Diagrama de blocos para a projeção da previsão da geração elétrica por fonte no Brasil.

## 4 Resultados e discussões

### 4.1 Validação dos Modelos

A análise dos dados de treinamento que abrangem a série histórica da geração elétrica por fonte entre os anos de 2012 a 2017 tem como intuito compreender o desempenho dos modelos de previsão em relação ao ano de 2018, conforme indicado pelo Anuário Estatístico de Energia Elétrica, onde apresenta a geração elétrica por fonte para o ano de 2018 de 601.396 GWh. Entre os modelos utilizados, o *Multilayer Perceptron* obteve uma geração elétrica estimada para 2018 de 581.156 GWh, com um MAPE de 3,37%, no qual o resultado indica uma precisão relativamente boa, mostrando a capacidade do modelo neural em ajustar-se à complexidade dos dados de geração elétrica.

Além disso, a Regressão do Processo Gaussiano apresentou um desempenho competitivo, gerando uma estimativa de 593.330 GWh para 2018, com um MAPE de 1,34%, demonstrando ser eficaz na captura de padrões de variação nos dados históricos. Por outro lado, a Regressão Linear, gerou uma estimativa de 605.864 GWh para o ano 2018, com o menor MAPE entre os modelos, com apenas 0,74%, no qual, apesar de sua abordagem menos complexa, revelou-se robusto na previsão da geração elétrica, indicando que, em certos contextos, simplicidade pode ser uma vantagem.

Por conseguinte, a análise dos modelos de previsão destaca a eficácia dos algoritmos utilizados na projeção da geração elétrica para o ano de 2018, no qual a escolha do modelo apropriado dependerá da natureza dos dados e dos requisitos específicos, considerando a complexidade do sistema elétrico e a necessidade de precisão nas estimativas.

### 4.1 Análise dos dados

A análise descritiva dos dados baseia-se na análise dos parâmetros, avaliando o valor médio, o valor mínimo, o valor máximo, o desvio padrão e o coeficiente de variação referentes à série histórica da geração elétrica por fonte, entre os anos de 2012 a 2022, conforme Tabela 3.

Tabela 3. Parâmetros dos dados.

Parâmetro	Usinas hidrelétricas	Não renováveis	Renováveis	Outras
Média (GWh)	387.681	108.927	93.668	13.437
Mínimo (GWh)	359.743	72.398	39.758	9.966
Máximo (GWh)	427.114	146.356	163.541	15.146
Desvio padrão (GWh)	21.074	24.208	39.668	1.36
Coefficiente de variação	5,436	22,224	42,350	1,012

Sabendo que o coeficiente de variação corresponde a uma medida da dispersão dos dados em relação à média, considera-se que os dados são homogêneos em relação à média quando considerados os dados históricos das usinas hidrelétricas e outras fontes e considerados heterogêneos para as renováveis e não renováveis, indicando maior variabilidade na produção de energia em relação à média.

Assim, as usinas hidrelétricas correspondem a principal fonte de geração de energia em média, e têm a maior produção mínima. As fontes renováveis têm o maior desvio padrão, significando que a produção de energia pode variar consideravelmente ao longo do tempo, frente as usinas hidrelétricas que têm o menor desvio padrão, indicando uma produção mais consistente.

Por conseguinte, as usinas hidrelétricas caracterizam-se por ser a fonte mais estável de produção de energia, com a menor variabilidade. As fontes renováveis, embora possam ter a capacidade de produzir grandes quantidades de energia em momentos ideais, também apresentam a maior variabilidade em relação à média e as fontes não renováveis têm um desempenho intermediário.

Ainda que, no contexto brasileiro, as hidrelétricas tenham historicamente desempenhado um papel crucial na matriz energética, observa-se um crescente interesse e investimento em fontes renováveis solar e eólica. Essa transição busca equilibrar a necessidade de fornecer energia de maneira eficiente com a consideração dos impactos ambientais e sociais.

Entretanto, as usinas hidrelétricas, ao aproveitarem a energia da água, oferecem uma produção estável e contínua de eletricidade, assim, a geração de energia hidrelétrica caracteriza-se por ser relativamente constante, proporcionando uma resposta flexível à demanda de eletricidade.

As fontes renováveis, como solar e eólica, são influenciadas por fatores climáticos e ambientais, no qual, a produção de energia solar está diretamente relacionada a quantidade de luz solar disponível, enquanto a geração eólica depende da velocidade e consistência do vento, essa dependência de fatores naturais resulta em variabilidade sazonal e diária.

E, ainda que as fontes renováveis possam atingir altos níveis de produção em momentos ideais, como dias ensolarados ou ventosos, sua variabilidade em relação à média significa que a produção não é constante ao longo do tempo, criando desafios para a integração eficiente dessas fontes na matriz energética, uma vez que se torna necessário lidar com flutuações na produção.

A variabilidade das fontes renováveis pode criar desafios para o gerenciamento da rede elétrica, exigindo soluções como o armazenamento de energia ou a implementação de tecnologias avançadas de previsão. A capacidade de prever e gerenciar efetivamente a variabilidade é crucial para otimizar a integração dessas fontes na matriz energética.

Em suma, o desafio reside em encontrar um equilíbrio entre a estabilidade oferecida pelas usinas hidrelétricas e a necessidade de integrar eficientemente as fontes renováveis, reconhecendo suas características variáveis, torna-se um papel crucial no processo de transição para uma matriz energética mais sustentável.

#### 4.2 Redes neurais *multilayer perceptron*

A aplicação das Redes Neurais *Multilayer Perceptron*, a partir de diferentes camadas de arquitetura e ajuste de hiperparâmetros, obteve como resultado os dados apresentados na Tabela 4, onde apresenta a previsão da geração elétrica por fonte para o período de 2023 a 2030.

Os hiperparâmetros foram delimitados pelo número de camadas ocultas “a” correspondendo ao número de atributos + número de classes, a taxa de aprendizagem de 0.3, o impulso de 0.2 e o número de épocas de 500.



Tabela 4. Previsão da Geração Elétrica por Fonte 2030 pelo método MLP (GWh).

Fontes	2023	2024	2025	2026	2027	2028	2029	2030
Usinas hidrelétricas	375.800	383.564	395.483	442.795	325.483	462.431	399.194	384.585
Não renováveis	115.018	93.164	119.871	56.678	207.132	55.141	104.574	78.165
Renováveis	157.577	187.646	206.991	219.090	226.458	286.524	249.830	289.596
Outras	14.112	15.454	17.651	12.346	17.749	15.557	14.729	15.469
Total	662.507	679.828	739.996	730.909	776.822	819.653	768.327	767.815

As usinas hidrelétricas apresentam valores variáveis ao longo dos anos, representam uma parte substancial da produção total de energia elétrica, entretanto pela previsão, ocorre flutuações consideráveis em sua contribuição ao longo do período. As fontes não renováveis, como o carvão, gás natural e petróleo, também se apresenta com variações significativas nas previsões, indicando uma produção de energia que pode ser altamente volátil.

Para as fontes renováveis, como energia solar, eólica e biomassa, apresentam um aumento constante na previsão de geração de energia ao longo dos anos, sendo considerada uma fonte de crescimento constante na produção de energia. E, a categoria de outras fontes, apresenta valores relativamente estáveis ao longo do período.

No geral, pelo método de previsão MLP, a produção total de energia elétrica varia de 662.507 GWh em 2023 a 819.653 GWh em 2028, no entanto, há uma diminuição na produção total em 2030, voltando a 767.815 GWh.

#### 4.3 Regressão do processo gaussiano

Pelo método de Regressão do Processo Gaussiano, a partir de processos estocásticos, obteve como resultado os dados apresentados na Tabela 5, onde apresenta a previsão da geração elétrica por fonte para o período de 2023 a 2030.

Tabela 5. Previsão da Geração Elétrica por Fonte 2030 pelo método Regressão do Processo Gaussiano (GWh).

Fontes	2023	2024	2025	2026	2027	2028	2029	2030
Usinas hidrelétricas	373.466	402.534	371.336	417.247	360.569	417.219	366.067	423.762
Não renováveis	119.261	87.668	130.470	72.326	139.133	72.872	130.303	66.284
Renováveis	141.024	165.478	176.821	193.738	198.195	243.960	241.042	283.913
Outras	14.505	14.657	15.802	14.380	16.347	15.380	16.690	15.420
Total	648.256	670.337	694.429	697.691	714.244	749.431	754.102	789.379

Considerando as usinas hidrelétricas, a previsão indica um aumento constante na geração de energia por usinas hidrelétricas ao longo dos anos, podendo ser um indicativo de uma maior dependência de usinas hidrelétricas para atender à demanda de energia elétrica, considerando seu aumento constante.

Para as fontes não renováveis há a previsão de uma queda acentuada ao longo dos anos, podendo ser resultante de políticas de redução de emissões de carbono e uma mudança em direção a fontes mais limpas de energia.

A previsão para fontes de energia renovável, por outro lado, mostra um aumento constante, isso sugere um maior investimento e crescimento na produção de energia renovável, como energia solar e eólica, para atender à crescente demanda.

As outras fontes de geração de energia apresentam variações relativamente pequenas ao longo dos anos.

Assim, a geração total de energia elétrica no Brasil prevista pelo método GPR mostra um aumento constante, atingindo 789.379 GWh em 2030, isso sugere um aumento geral na capacidade de geração de energia no país para atender à demanda crescente.

#### 4.4 Regressão linear

A metodologia de Regressão Linear obteve como resultado os dados apresentados na Tabela 6, onde apresenta a previsão da geração elétrica por fonte para o período de 2023 a 2030.

Tabela 6. Previsão da Geração Elétrica por Fonte 2030 pelo método Regressão Linear (GWh).

Fontes	2023	2024	2025	2026	2027	2028	2029	2030
Usinas hidrelétricas	362.828	376.443	384.748	409.628	340.541	416.418	383.304	384.134
Não renováveis	127.348	98.621	132.763	61.370	186.429	74.148	116.776	76.265
Renováveis	144.526	167.348	189.002	166.595	172.553	197.342	192.248	176.111
Outras	14.153	14.153	14.153	14.153	14.153	14.153	14.153	14.153
Total	648.855	656.565	720.666	651.746	713.676	702.061	706.481	650.663

A geração prevista de energia pelas usinas hidrelétricas mostra uma variação ao longo dos anos, com valores que aumentam gradualmente até atingir o pico em 2026 e, em seguida, diminuem. É notável que a geração em 2027 é significativamente menor, seguida por uma recuperação em 2028. Essa variação pode ser influenciada por fatores climáticos, disponibilidade de recursos hídricos e investimentos no setor.

Notavelmente, em 2027, há uma queda acentuada na geração hidrelétrica, podendo ser indicativo de uma possível redução na disponibilidade de água para a produção de energia, sendo interpretada como um sinal de potencial crise hídrica nesse ano. E, com a diminuição na produção hidrelétrica, resulta em desafios para atender à demanda crescente por eletricidade, levando a adoção de medidas de racionamento e/ou a necessidade de aumentar a participação de outras fontes de energia para compensar a queda na produção hídrica.

A geração de energia a partir de fontes não renováveis, como combustíveis fósseis, mostra uma tendência de aumento gradual, com algumas oscilações. Em 2026, há uma queda acentuada, seguida por uma recuperação em 2027, podendo estar relacionado a mudanças nas políticas energéticas, variações no preço do petróleo e outros fatores.

A geração de energia a partir de fontes renováveis, como eólica, solar e biomassa, também apresenta uma tendência de aumento, mas com menos variações em comparação com as outras fontes, isso reflete o crescimento contínuo das energias renováveis no mix energético.

As outras fontes energéticas mostram valores constantes e inalterados ao longo dos anos, o que pode representar fontes de geração constante, como a nuclear.

A geração total de energia elétrica pelo método de Regressão Linear varia ao longo dos anos, refletindo as tendências nas fontes individuais, ressaltando a diminuição em 2027 e a recuperação subsequente em 2028.

#### 4.5 Apontamentos e discussões

As previsões são fundamentais para o planejamento e a gestão do setor de energia elétrica, fornecendo *insights* sobre as contribuições de diferentes fontes de energia ao longo do tempo. Entretanto, ressalta-se que as previsões estão sujeitas a incertezas e podem variar com base em diversos fatores, incluindo condições climáticas, mudanças na demanda e avanços tecnológicos.

É fundamental monitorar e ajustar as previsões conforme necessário para garantir um fornecimento confiável de energia no futuro, assim como, é importante considerar essas métricas ao avaliar a confiabilidade das previsões e ao planejar a geração futura de energia com base nos resultados.

Em suma, as métricas mostram que o método MLP é capaz de fornecer previsões com diferentes níveis de precisão, dependendo da categoria de fonte de energia, no qual as usinas hidrelétricas têm as previsões mais precisas, seguidas outras fontes, as renováveis e as não renováveis.

Para o método GPR é importante observar que a precisão das previsões depende da qualidade dos dados de entrada e da adequação do modelo GPR, onde as incertezas associadas a essas previsões também precisam ser consideradas ao tomar decisões estratégicas no setor de energia.

O método de Regressão Linear parece funcionar bem para prever a geração de energia a partir de várias fontes, no entanto, é importante observar que o desempenho varia entre as diferentes fontes, com as fontes renováveis tendo as previsões mais precisas, assim, a escolha de métricas específicas deve depender dos requisitos de precisão do sistema elétrico e do setor energético.

Por conseguinte, a análise dos dados corresponde a um passo crucial para compreender as tendências futuras na geração de energia elétrica e planejar adequadamente o suprimento de energia para o país, podendo ser valiosa para o planejamento do setor energético, identificando tendências e variações que podem afetar a oferta e a demanda de energia no futuro.

Além do mais, considerar outros fatores, como políticas energéticas, mudanças tecnológicas e impactos ambientais, ao interpretar esses resultados torna-se importante.

O comportamento dos resultados da geração elétrica por fonte para o intervalo anual de 2023 a 2030, obtidos pela aplicação dos métodos MLP, GPR e Regressão Linear, apresentam-se pela Figura 2, onde relaciona o comparativo entre os métodos.

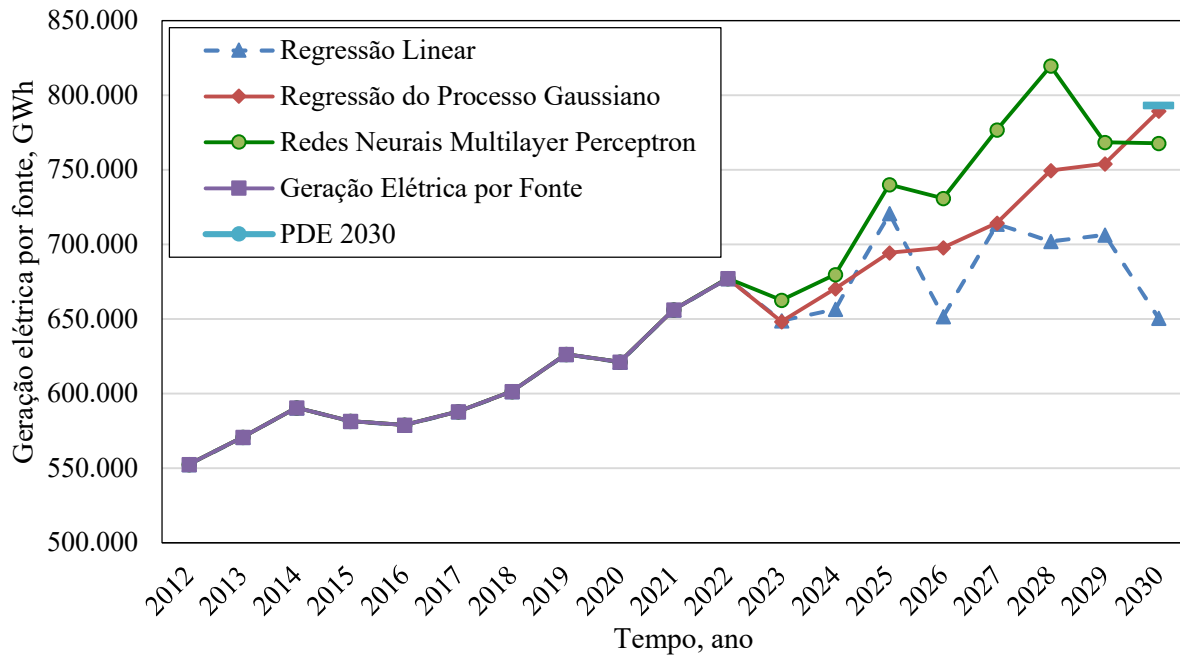


Figura 2. Comparativo entre os métodos para a previsão da expansão da geração elétrica por fonte

O método de Regressão Linear demonstrou uma tendência de crescimento moderado, com alguma variação nos anos finais do período, o método de Regressão do Processo Gaussiano parece ser mais otimista quanto ao crescimento da geração de energia e o Redes Neurais *Multilayer Perceptron* tem uma tendência de crescimento mais variável.

A Regressão Linear e o Processo Gaussiano tendem a produzir projeções mais estáveis, enquanto as Redes Neurais MLP mostra maior variabilidade ao longo do tempo.

Quando comparado ao PDE 2030, mediante a previsão da geração elétrica por fonte, o método Regressão do Processo Gaussiano alcança 789.379 GWh em 2030 e a meta estabelecida pelo Plano Decenal de Expansão de Energia pressupõe que em 2030 se atinge 793.000 GWh, alcançando o RMSE 21.564 GWh, MAE 3.621 GWh e MAPE 0,5%, por conseguinte, evidencia-se uma proximidade nos valores projetados, indicando que o método proposto está alinhado com as metas estabelecidas pelo planejamento energético.

A pequena diferença, de aproximadamente 3.621 GWh, entre a previsão do método de Regressão do Processo Gaussiano e a meta do PDE 2030 sugere uma adequação razoável do modelo na estimativa da geração elétrica futura, tencionando a essa proximidade um indicativo de que o método pode ser considerado confiável para fornecer projeções realistas, ainda que considere as incertezas e variabilidades inerentes ao sistema energético.

Em suma, pela análise comparativa sugere que o método Regressão do Processo Gaussiano apresenta resultados coerentes e alinhados com as metas do PDE 2030, fornecendo uma base sólida para as projeções futuras da geração elétrica por fonte.

Torna-se essencial considerar os resultados das métricas de desempenho, como MAE, MAPE e RMSE, para avaliação da precisão das previsões de cada método, determinando o mais adequado ao propósito desejado, uma vez que as métricas fornecem informações valiosas sobre quão próximas as projeções estão dos valores reais.

Assim, para o método de Regressão Linear, alcançou-se o MAE de 142.337 GWh, MAPE de 17,9% e RMSE de 142.337 GWh, bem como, para as Redes Neurais *Multilayer Perceptron* o MAE corresponde a 25.185 GWh, o MAPE de 3,2% e RMSE de 117.151 GWh.

## 5 Conclusões

A projeção da previsão do mercado de energia elétrica no Brasil possibilitou a identificação dos padrões de comportamento mercadológico, permitindo antecipar as tendências e mudanças no mercado, contribuindo para um planejamento mais informado e estratégico no setor energético:

- a) Ao delinear o horizonte previsto pelo Plano Decenal de Expansão de Energia 2030, possibilitou a criação de uma visão para avaliação de quão próximas ou distantes as projeções obtidas pelos modelos estão das metas governamentais, sendo fundamental para compreender a viabilidade e a eficácia das estratégias propostas, contribuindo para uma abordagem mais precisa e alinhada com as diretrizes governamentais.
- b) A utilização do software WEKA, em conjunto com os métodos de Redes Neurais *Multilayer Perceptron*, Regressão do Processo Gaussiano e Regressão Linear, representa um avanço significativo na análise de dados e previsão para o mercado energético. Ao longo desta dissertação, foi explorado as capacidades e potencialidades dessas técnicas, evidenciando como elas podem ser valiosas ferramentas na tomada de decisões e versátil na análise de dados e previsão da geração elétrica por fonte no Brasil.
- c) A metodologia proposta de previsão da geração elétrica por fonte, com um mercado brasileiro majoritariamente de participação hídrica, alcançou os resultados esperados, em conformidade com as premissas do Plano Decenal de Expansão de Energia 2030.
- d) O crescimento do mercado energético nacional, mediante aos métodos aplicados retrata que as fontes renováveis tiveram um crescimento considerável, estando próximo ao objetivo do Plano Decenal de Expansão de Energia, em que delinea ser atingido o nível de renovabilidade de 90% abrangendo as fontes hidráulica, biomassa, eólica e solar. Para os métodos de Regressão linear, alcançou a renovabilidade de 86%, assim como, para a Regressão do Processo Gaussiano 90% e para a Redes Neurais *Multilayer Perceptron* a renovabilidade atingiu 88%.
- e) As previsões retratadas pelos métodos visaram a promoção de argumentações que possibilitem o desenvolvimento de ações presumidas para o processo de planejamento energético.

## Agradecimentos

A pesquisa foi parcialmente apoiada pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais, FAPEMIG, pelo financiamento do Projeto: “Simulação e otimização de sistemas híbridos isolados de geração de energia elétrica baseados em energia renovável” (PROCESSO APQ-01932–21) executado na Universidade Federal de Itajubá no âmbito do edital 001/2021—Demanda Universal, registro DPI UNIFEI nº: PVDI208- 2021. Assim como, foi complementada com recursos do projeto nº 403582/2021–0; Chamada CNPq/MCTI/FNDCT Nº 18/2021 – Faixa A – Grupos Emergentes. Projeto financiado Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

## Referências bibliográficas

- Aggarwal, CC (2018) *Neural Networks and Deep Learning: A Textbook*. Springer.
- Barros, EV (2007) ‘A Matriz Energética Mundial e a Competitividade das Nações: bases de uma nova geopolítica’, *Engevista*, v.9, n.1, pp. 47–56.
- Basso, L (2019) ‘Brazilian energy-related climate (in)action and the challenge of deep decarbonization’, *Revista Brasileira de Política Internacional*, v. 62, n. 2, FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/0034-7329201900202>
- Castro, R, e Lyra Filho, C (2005) ‘Um método de suporte a decisões sobre investimento e comercialização de energia elétrica no Brasil’, *Sba: Controle & Automação Sociedade Brasileira de Automatica*, v. 16, n. 4, pp. 478–494.
- Gehrke, P, Goretti, AAT e Ávila, LV (2021) ‘Impactos da matriz energética no desenvolvimento sustentável do Brasil’, *Revista de Administração da UFSM*, v. 14, p. 1032-1049. <http://dx.doi.org/10.5902/1983465964409>
- Gonzalez-Briones, A, Hernandez, G, Corchado, JM, Omatu, S, e Mohamad, MS (2019) ‘Machine learning models for electricity consumption forecasting: A review’, *2nd International Conference on Computer Applications & Information Security (ICCAIS)*, Riyadh, Saudi Arabia, pp. 1–6.

- Goodfellow, I, Bengio, Y e Courville, A (2016) *Deep Learning*. MIT Press.
- Hyndman, RJ e Athanasopoulos, G (2018) *Forecasting: Principles and Practice*. OTexts.
- Hyndman, RJ, Koehler, AB (2006) ‘Another look at measures of forecast accuracy’, *International Journal Of Forecasting*, v. 22, n. 4, p. 679-688, Elsevier BV. <http://dx.doi.org/10.1016/j.ijforecast.2006.03.001>
- IRENA (2022) ‘Panorama das Transições Energéticas Mundiais 2022’, Agência Internacional para as Energias Renováveis, Abu Dhabi.
- Korzeniewicz, MBDV (2021) ‘Análise da matriz energética brasileira e a participação das energias renováveis a partir das políticas ambientais energéticas’, Dissertação de Mestrado, Programa de Pós-Graduação em Economia do Desenvolvimento, PUCRS, Porto Alegre, p. 55.
- Makridakis, SG, Wheelwright, SC e Hyndman, RJ (1998) *Forecasting: methods and applications*. Wiley.
- Montgomery, DC, Peck, EA e Vining, GG (2021) *Introduction to Linear Regression Analysis*. Wiley.
- Moore, DS, McCabe, GP e Craig, BA (2014) *Introduction to the Practice of Statistics*. W. H. Freeman.
- Moret, MA (2014) ‘Análise da expansão do setor elétrico brasileiro considerando os impactos socioeconômicos e ambientais’, Dissertação de Mestrado, Universidade Estadual de Campinas - UNICAMP, São Paulo.
- Noguchi, A e Nobre, FS (2023) ‘Oil and Gas Companies - Are They Shifting to Renewables? A Study of Policy Mixes for Energy Transition in Brazil’, *Bar - Brazilian Administration Review*, v. 20, n. 1. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/1807-7692bar2023220087>
- Oliveira, ACL, Tótola, LA, Lorentz, JF, Silva, AA; Assis, LR, Santos, VJ e Calijuri, ML (2022) ‘Spatial analysis of energy indicators and proposition of alternative generation sources for the Brazilian territory’, *Journal Of Cleaner Production*, v. 356, p. 131894, Elsevier BV. <http://dx.doi.org/10.1016/j.jclepro.2022.131894>
- Oliveira, FJA (2020) *O planejamento da operação energética no sistema: conceitos, modelagem matemática, previsão de geração e carga*. São Paulo: Artliber.
- Pereira, TCG (2014) *Planejamento Energético e as políticas públicas: Aspectos conceituais e metodológicos*. Curitiba: COPEL.
- Rasmussen, CE e Williams, CKI (2006) *Gaussian Processes for Machine Learning*. The MIT Press.
- Shumway, RH e Stoffer, DS (2017) *Time Series Analysis and Its Applications: with R examples*. Springer.
- Tan, PN, Steinbach, M e Kumar, V (2006) *Introduction to data mining*. Pearson Addison Wesley.
- Ventura Filho, A (2014) ‘A Política Energética do Brasil’, *Cadernos Adenauer XV*, v. 3, pp. 121–143.
- Witten, IH, Frank, E, Hall, MA e Pal, CJ (2016) *Data mining: practical machine learning tools and techniques*. Morgan Kaufmann.
- Zarur, MAF (2005) ‘Modelo para Elaboração de Cenários do Setor Energético, utilizando técnicas de data mining’, Dissertação de Mestrado, Ciência em Engenharia Civil, COPPE/UFRJ, Rio de Janeiro, p. 107.