



## ANÁLISE DE DADOS PARA PREDIÇÃO DE CENÁRIOS DA MATRIZ ELÉTRICA NACIONAL PARA 2050

Helena Jacques Vendramini<sup>1</sup>

Claudio Ernesto Ponce Saldias<sup>2</sup>

**Resumo:** o presente trabalho tem por objetivo principal a predição de dados para o cenário da matriz elétrica brasileira para o ano de 2050. O modelo desenvolvido possui a vantagem de atualização futuras, onde eventuais alterações no cenário econômico ou social, que afetam diretamente ou indiretamente a matriz elétrica nacional, podem ser acrescentadas no modelo para um novo processo de treinamento dos dados. A predição da matriz elétrica é necessária para que a organização estrutural e econômica seja premeditada a fim de garantir o bom funcionamento energético no cenário futuro. Para prever cenários futuros existem várias técnicas que podem ser implementadas, como técnicas estocásticas e técnicas de Inteligência Artificial (IA). Neste trabalho foram utilizadas técnicas estatísticas de auto regressão e também técnicas de inteligência artificial. Os dados utilizados para o treinamento dos modelos foram coletados nos dados históricos no site oficial do Operador Nacional do Sistema Elétrico (ONS) e também foram utilizadas premissas sociais e econômicas para verificar a influência dos mesmos no cenário energético. Para a primeira análise foram utilizadas técnicas de auto regressão, pelo método de *Regressão Linear* e *Arima*, estes apresentaram resultados com um erro demasiadamente grande, justificado pela ausência de estacionalidade nos dados e evidenciado pela falha no teste de *Newcomb-Benford* e de *Dickey-Fuller*, entendendo que fatores externos, como políticas e estratégias do governo e das empresas, influenciam no comportamento dos dados. Como segunda alternativa foi utilizada uma estratégia de IA por meio de uma rede neural recorrente, porém, a performance do modelo continuou com baixa efetividade. Para esse estudo constatou-se a necessidade de regras e ontologias além de dados numéricos, considerando o amplo prazo de previsões, portanto, apenas um tipo de rede neural por si só não é eficiente para essa situação. Finalmente foi implementada uma estrutura que combina redes neurais com um motor de regras na plataforma Jemo, *framework* em Java. Foi realizado o treinamento de dados e obteve resultados com bons indicativos evidenciados pelo baixo erro registrado. De acordo com as predições realizadas pela plataforma em Jemo, concluiu-se que o cenário proposto pelo Plano Nacional de Energia (PNE) que mais representa a matriz elétrica brasileira de 2050 de acordo com os resultados obtidos considera a redução de emissão de gases poluentes e também mudanças climáticas. Comparando com a matriz atual houve o aumento de geração de energia solar e eólica e o decréscimo de energia hidrelétrica e térmica.

**Palavras-chave:** Matriz Elétrica. *Python*. Inteligência artificial. Predição de dados.

---

<sup>1</sup> Graduando em Engenharia Elétrica, 2022: E-mail: helena.261417@ satc.edu.br

<sup>2</sup> Professor do Centro Universitário UniSATC E-mail: claudio.saldias@satc.edu.br



## 1 INTRODUÇÃO

A análise da matriz energética de um país é de grande importância principalmente para convergir com o crescimento da demanda de energia. Enquanto a matriz energética abrange o conjunto de todas fontes de energia disponíveis, a matriz elétrica representa as fontes disponíveis apenas para a geração de energia elétrica (EPE, 2022a). A agência responsável pela fiscalização e regulamentação do setor elétrico no Brasil é a Agência Nacional de Energia Elétrica (ANEEL). Sob a fiscalização da ANEEL, o órgão Operador Nacional do Sistema Elétrico (ONS) coordena e monitora as instalações de geração e transmissão de energia elétrica e, o Sistema Interligado Nacional (SIN) gerencia os sistemas elétricos (ONS, 2022a).

Para o estudo da expansão do setor de energia é preciso analisar e interpretar os dados pertinentes, para que possa fornecer o suporte para estratégias de longo prazo do planejador (EPE, 2022b). Para a análise do sistema elétrico nacional é preciso a implementação de técnicas que consigam processar a grande quantidade de dados do sistema elétrico nacional. Demonstrando a necessidade da implementação da análise de dados, bem como seus obstáculos (STEFANINI, 2020). Estes métodos de estudo permitem a exploração dos dados energéticos com intuito de identificar padrões de interesse e a representação dos dados caracterizados por estes padrões, tornando viável a previsão de dados futuros (LOPES, 2019).

Para a previsão de dados no setor de energia elétrica há alguns estudos periódicos realizados para possibilitar o planejamento de expansão energética. O Indicador de Previsão de Carga (IPC) é uma estimativa de dados disponibilizada no site da ONS que demonstra a relação da previsão de cargas nos últimos 60 meses com o consumo real que foi obtido (ONS, 2022b). A Empresa de Pesquisa Energética (EPE) sob diretrizes gerais do Ministério de Minas e Energia (MME) realiza as previsões estatísticas com maior prazo (EPE, 2020a). Para as previsões a longo prazo há o Plano Decenal de Expansão de Energia (PDE) para previsão de 10 anos e o Plano Nacional de Energia (PNE) para previsão de 30 anos (EPE, 2022b).

Entendendo a necessidade de um prévio estudo do cenário futuro de geração e demanda de energia elétrica, o intuito é auxiliar nas previsões com a análise de dados históricos para o ano de 2050. Explora-se diferentes técnicas algorítmicas e avalia-se a que melhor atende aos critérios, comparando os resultados finais com os



cenários de geração previstos pelo PNE. Eventualmente utiliza-se inteligência artificial para o refinamento das técnicas e o aperfeiçoamento das mesmas com a utilização de raciocínio lógico para análise dos resultados. Entende-se como vantagem da criação do modelo a utilidade a longo prazo com a alteração do cenário nacional, ou seja, para eventuais mudanças sociais ou econômicas pode-se acrescentar como dados de entrada para que haja um novo treinamento do modelo com os dados acrescentados e, conseqüentemente, novas previsões considerando as atualizações do novo cenário encontrado. Outra eventual utilidade que poderia-se obter com o modelo, além de prever o cenário nacional de matriz elétrica, é a previsão de gerações ou cargas regionais, basta atualizar os dados de entrada de acordo com a necessidade de estudo.

Objetiva-se, então, realizar as previsões da matriz elétrica para o ano de 2050, 28 anos à frente (mais de 300 meses), com um modelo e metodologia próprios desse estudo. Há o processo de verificação da precisão do modelo, além da performance medida automaticamente pela programação, também há o processo de validação do modelo pela aproximação dos resultados com os cenários propostos pelo PNE (EPE, 2020b).

## **2 REVISÃO DA LITERATURA**

Nesta seção são abordados tópicos referentes ao processo de análise de dados, contemplando análise lógica e estatística de informações e também o estudo aprofundado de Inteligência Artificial (IA).

### **2.1 MATRIZ ENERGÉTICA**

O programa Performance Organizacional (PO) é um programa do ONS regulamentado pela ANEEL para elaboração de indicadores de desempenho e apontamento de metas, sendo estes dados disponibilizados em área livre do site do ONS. Um dos estudos do programa é o IPC, que apresenta valores mensais acumulados com o objetivo de medir o desempenho das previsões nos últimos 60 meses. Espera-se a melhoria dos modelos de previsão e otimização utilizados no Programa Mensal da Operação (PMO) com este estudo do IPC (ONS, 2022b). Para o estudo de instalações de transmissões futuras existem dois programas que são



publicados duas vezes ao ano, há o Programa de Expansão da Transmissão (PET) com o estudo dos seis anos subsequentes e o Plano de Expansão de Longo Prazo (PELP) com estudo a partir do sétimo ano decorrente da emissão do documento (EPE, 2022c).

A EPE realiza estudos e pesquisas no setor energético, é uma empresa pública federal que presta serviços ao MME. Os agentes reguladores são a ANEEL, Agência Nacional do Petróleo, Gás e Biocombustíveis (ANP), Agência Nacional de Águas (ANA), o ONS e a Câmara de Comercialização de Energia Elétrica (CCEE). O objetivo das atividades é orientar o desenvolvimento do setor energético nacional, assegurando as bases de sustentabilidade e infraestrutura (EPE, 2022c).

O PDE é um documento disponível para consulta no site da EPE com objetivo de indicar perspectivas de futura expansão do setor de energia no intervalo de dez anos. Já o PNE é a pesquisa que avalia as tendências na produção e no uso da energia, direcionando a expansão da energia nas próximas décadas. Atualmente o PNE disponível no site da EPE são as previsões para o ano de 2030 e 2050 (EPE, 2022b). Essas projeções utilizam referências externas para análise, a última predição da matriz energética nacional a longo prazo foi publicada pela EPE e MME em 2020, tendo as seguintes premissas consideradas: análise de demanda e economia, requisitos de geração, autoprodução não injetada na rede, recursos energéticos distribuídos, alternativas e desafios da expansão energética, fontes e tecnologias (EPE; MME, 2018).

## 2.2 ESTATÍSTICA

A estatística é o estudo de dados que se baseia no recolhimento, organização e análise de informações. Esse procedimento generalizado pode ser descrito por meio de metodologias diversificadas, dependendo-se do objetivo do estudo deve-se escolher o método que melhor explore os dados coletados. Existem três divisões na estatística, descritiva, probabilística e inferencial. A descritiva organiza, demonstra e resume dados, enquanto a probabilística analisa situações sujeitas ao acaso, estudando a probabilidade dos eventos. Já a inferência busca respostas por dados representativos, ou seja, análise com amostras de uma população (SILVESTRE, 2007).



Entre os conceitos da probabilidade destacam-se: o experimento, que é o objeto de estudo; o espaço amostral, onde há a possibilidade de ocorrência; e o evento, que são os resultados ocorridos, podendo ser excludentes ou não. Os resultados não podem ocorrer de forma simultânea e, podem ainda, ser classificados como dependentes ou independentes, onde demonstra a relação entre eventos e a interferência mútua existente ou não (FARBER; LARSON, 2010).

### 2.2.1 Análise de dados

A Análise Exploratória de Dados (AED) busca o aumento da eficácia de estudos estatísticos, principalmente quando se utilizam amostras para representar uma determinada população. Os principais métodos são: coleta, exploração, descrição e interpretação de dados (LOPES, 2019). Os dados podem ser definidos como quantitativos ou qualitativos; os quantitativos são numéricos e podem ser contínuos, com valores reais, ou discretos, com números fixos e inteiros; já os qualitativos representam dados categóricos, podendo ser nominal, sem hierarquia, ou ordinal, com hierarquia (FARBER; LARSON, 2010).

O processo de recolhimento de dados pode ser feito utilizando apenas um percentual da população. Um grupo de dados que representa um alvo de estudo da população é chamado de amostra, podendo ser selecionado por meio de diferentes técnicas. Sua utilização se faz necessária por conta da inviabilidade em alguns casos de analisar em censo, ou seja, pesquisa com toda a população. Em contrapartida, a utilização de amostra para a representação de dados evidencia maior variação nos resultados (FARBER; LARSON, 2010).

Para validar o modelo estatístico pode-se analisar as métricas de erros, sendo essa a diferença entre o que foi previsto e o que realmente aconteceu. Pode-se realizar a previsão de valores categóricos, utilizando métricas específicas para comparação, ou valores numéricos. A análise dos erros define o quanto o modelo estatístico é eficiente, podendo-se analisar utilizando diferentes métodos para diferentes aplicações, destacam-se os métodos *Mean Error* (ME), *Mean Absolute Error* (MAE), *Mean Percentage Error* (MPE), *Mean Absolute Percentage Error* (MAPE) e *Root Mean Squared Error* (RMSE). Outra técnica bastante utilizada na IA é a *Matriz Confusão* (FARBER, 2010).



### 2.2.2 Teste de estacionalidade dos dados

Para analisar a estacionalidade dos dados foram realizados dois testes, teste de *Newcomb-Benford* e teste de *Dickey-Fuller*. O teste de *Newcomb-Benford* analisa um padrão sequencial na disposição de dados. Esse método de análise de dados permite verificar a ocorrência natural dos dados e, para isso, as informações devem representar a magnitude do objeto em estudo e devem seguir algumas condições: a amostra não pode ser pequena ou possuir variações, não pode ser exigido valores mínimos ou máximos, não deve referenciar cadastros ou número de identificação, não podem estar muito agrupados em torno da média, não deve haver influência psicológica como mensagem subliminar nem políticas de decisões (CUNHA, 2013).

Já o teste de *Dickey-Fuller* verifica a presença de raiz unitária em modelos auto regressivos tais como *Arima*, *Sarima*, *Arn*, dentre outros. A falha do teste conclui que a série analisada não é estacionária e, portanto, não é recomendada a aplicação de técnicas auto regressivas sem que os dados originais sejam convertidos, ou seja, transformados em dados estacionários. Sendo “ $\delta$ ” a variável resultante da análise, este teste estabelece duas hipóteses: hipótese zero, para  $\delta=0$  a série é estacionária; hipótese um, para  $\delta<0$  a série não estacionária (DICKKEY; FULLER, 1979).

### 2.2.3 Métodos para predição para séries temporais

A distribuição estatística está relacionada com a probabilidade de ocorrência de um evento. Existe na estatística metodologias que definem essa probabilidade, como o teste *T de Student*, a *Distribuição Binomial*, a *Distribuição de Poisson* e o *QI Quadrado*. Já para a predição de dados as técnicas mais utilizadas são a *Regressão Logística* e *Séries Temporais*, como *Arima*, *Regressão Linear*, *Sarima*, *ARM*. Também existem técnicas de estimativa de estados discretos na forma de equações de estados preditivas, mas são eficientes para curto prazo, para a situação problema deste artigo não seria recomendado (OGATA, 2015).

Para o treinamento de dados pelas séries temporais deve-se respeitar as condições para manter uma boa predição: correlação, que representa a influência mútua entre duas variáveis; tendência, sendo essa o movimento sistemático de aumento ou decréscimo; sazonalidade, que representa situações periódicas



recorrentes; estacionalidade, que é definida como o desenvolvimento aleatório no tempo entorno de uma média constante. Na Eq. (1) mostra-se o "Modelo Genérico Linear Geral" (LEVINE et al., 2005), uma regressão linear múltipla, onde o estudo de uma variável depende de mais de uma variável explicativa.

$$y_i = \beta_0 + \beta_1.X1_i + \dots + \beta_k.Xk_i + \epsilon_i \quad (1)$$

Onde:

$y_i$  = variável dependente, explicada ou resposta;

$\beta_0$  = termo independente de variável;

$\beta_k$  = inclinação de  $y$  em relação à variável  $X_k$ , mantendo constantes as variáveis  $X_1, X_2, \dots, X_{k-1}$

$\epsilon_i$  = erro aleatório em  $y$  para a observação  $i$ ;

$i = 1, 2, \dots, n$ .

### 2.3 INTELIGÊNCIA ARTIFICIAL

O estudo da IA possui 3 tópicos mais conhecidos: aprendizagem de máquina, cujo estudo é voltado para estatística e processos estocásticos; aprendizagem profunda, que envolve redes neurais de várias camadas; e raciocínio, sendo este a representação de conhecimento usando ontologias e regras.

*Machine Learning* (aprendizagem de máquina), é o conceito de técnicas aplicadas a máquinas usando dados para treinar a capacidade de aprendizado. Já para a aprendizagem profunda, há o estudo das redes neurais artificiais, que possuem estrutura semelhante a redes neurais biológicas, contendo os chamados neurônios que realizam conexões entre si e, a função de ativação, que imita as sinapses cerebrais representando a conexão entre os neurônios. A função de ativação mais comum é a *Rectified Linear Unit*, RELU, que é uma função mais utilizada por conta da alteração gradativa de valor, ou seja, uma suavização dos resultados representados.

A camada de entrada na rede é determinada pelos atributos, enquanto a camada de saída depende da classe, ou seja, as entradas dependem dos dados estudados enquanto a saída é representada de acordo com as divisões que



enquadram os dados. Há ainda a camada oculta, que é a estrutura da rede neural artificial englobando as interconexões entre os neurônios entre as camadas externas.

O termo *Backpropagation* é o processo de ajuste de pesos, podendo ser ajustados conforme interações, denominadas *epoch*, passando os dados pela rede para avaliar a métrica de erro e, de acordo com o resultado, é realizada a correção dos pesos visando melhor performance do modelo.

O objetivo da criação de *Machine Learning* pode ser classificado de acordo com tarefas destacadas abaixo:

- Classificação: prevê a classificação de algo categórico ou nominal a partir de dados existentes;
- Regressão: semelhante a classificação, porém sua predição é voltada para previsões numéricas;
- Agrupamentos: objetivo de apenas agrupar e não prever.
- Regras de associação: de acordo com semelhanças realiza a análise do histórico de dados para prever eventos futuros.

O método de treinamento e teste permite a avaliação do aprendizado de máquina, havendo três classificações do modelo de acordo com eficácia de funcionamento. Os melhores resultados ocorrem através de um classificador genérico, que tende a obter maior eficiência no modelo de laboratório bem como na aplicação prática de dados reais. Modelos super-adaptados ocorrem quando a programação se adapta a dados específicos, funcionando bem com dados de treino e pouco desempenho em dados reais. Já os modelos sub-ajustados, não conseguem boas taxas de previsão, a performance é pobre no ambiente de desenvolvimento e de produção.

Os principais tipos de algoritmos que podem ser utilizados para criação do modelo de predição de dados são: árvores de decisão; regras; *Naive Bayes*, classificação; *Redes Bayesianas*; redes neurais artificiais e aprendizado profundo; *Support Vector Machine*; métodos de agrupamento, como *clustering*; aprendizado baseado em instância; regressão logística.

### 2.2.3 Técnicas de previsão de séries com IA

Dentro da técnica de redes neurais artificiais aplicadas a previsão de séries temporais, há uma implementação bastante avançada chamada *deep learning (DP)*,





aprendizado profundo, que é um tipo de rede neural mais complexo contendo mais camadas. Para a predição de dados o tipo de rede neural mais utilizado é a rede neural recorrente (RNN), sendo essa uma rede que possui loops, ou seja, as informações são repassadas entre as etapas. Essa cadeia que interliga as camadas torna o método eficaz para a análise sequencial ou de listas, porém, sua utilização se limita a predições de curto prazo.

A LSTM (Redes de Memória de Curto Prazo Longo) é um tipo de rede neural recorrente adaptada para melhorar a predição de um prazo relativamente longo, onde a sua memória possui maior alcance conectando uma quantidade maior de informações. A estrutura da LSTM possui a principal diferença na quantidade de camadas, ao invés de uma, possui quatro camadas de rede neural interagindo entre si, tendo a capacidade de remover ou adicionar informações ao estado da célula (KELLEHER, 2019).

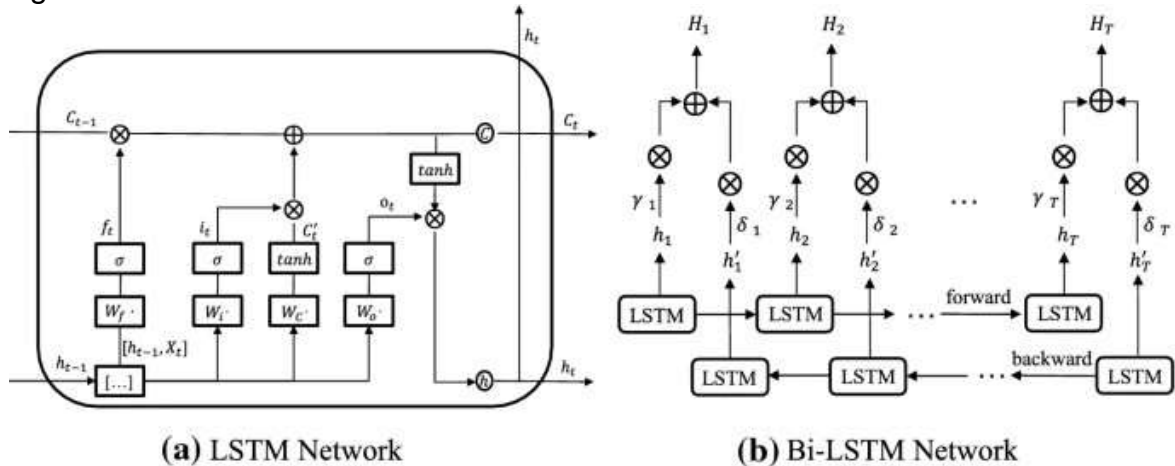
As camadas da rede neural LSTM possuem diferentes células que são blocos de memória. As células coletam as informações e os portões, denominados *gates*, manipulam as memórias. O estado da célula representa a memória a longo prazo, que pode ser modificada pelo *forget gate* e ajustada pelas portas de entrada, onde há o esquecimento de alguns dados e o acréscimo de outros. Considerando que  $C_{t-1}$  representa a entrada e  $h_{t-1}$  representa a saída de célula anterior, segue descrito abaixo o processo estrutural em cadeia da rede neural LSTM:

- *Forget Gate*: possui informações sem utilidade. Entram  $C_{t-1}$  e  $h_{t-1}$  que são multiplicadas por matrizes de peso, seguidas pela adição do *bias* (ajusta as saídas de acordo com os respectivos pesos). O resultante passa por uma função de ativação que providencia saída em zero, informação que será esquecida, ou em um, informação que será utilizada.
- *Input Gate*: adição de informações úteis. Com as entradas  $C_{t-1}$  e  $h_{t-1}$ , a função *sigmoide* filtra os valores a serem lembrados e a função *tanh* dá a saída de -1 à +1, ambas funções são multiplicadas.
- *Output Gate*: extrai informações úteis do estado atual da célula. A função *tanh* gera um vetor que posteriormente é regulado pela função *sigmoid* usando as entradas  $C_{t-1}$  e  $h_{t-1}$ , esse resultado é multiplicado pelo vetor criado pela função *tanh* para servir como entrada para a próxima célula.

A LSTM também pode ser classificada como bidimensional, conforme

ilustrado na Fig. 1, isso acontece quando o processo de interação entre as células da rede neural possui dois sentidos, um sentido linear para frente em simultaneidade com um sentido linear para trás.

Figura 1: Rede neural LSTM unidimensional e bidimensional.



Fonte: CHENG et al (2022)

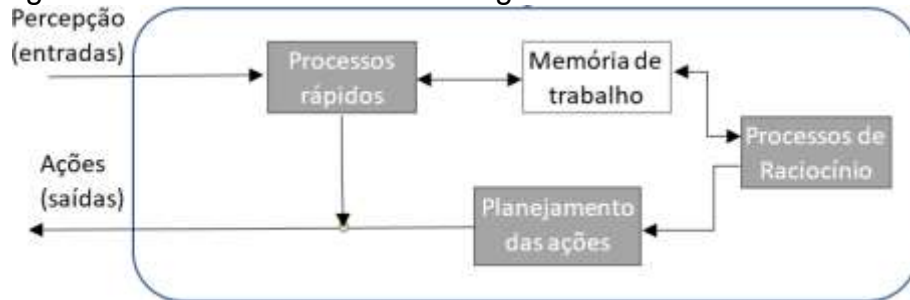
Conforme Fig. 1, percebe-se a combinação dos dois métodos distintos de aprendizagem, que capacitam o modelo para melhor performance de treinamento. A rede Bi-LSTM é especializada para previsão de séries temporais composta por unidades LSTM.

## 2.2.4 Framework Jemo

Foi criado um *framework* (uma biblioteca disponível para uso) em Java (uma linguagem de programação) que possibilita a previsão de dados a longo prazo, em uma tese de doutorado pela Universidade Federal de Santa Catarina (SALDIAS, 2019), cujo estudo é intitulado como "Protótipo de sistema de instrução virtual para treinamento de pilotos de aviões Boeing 737-800". Para esse método houve o acréscimo das premissas teóricas, ou seja, a utilização do raciocínio para analisar os dados numéricos.

O pacote de software, elaborado no estudo da tese, foi o Jemo, uma plataforma disponibilizada em *framework* em Java que implementa agentes seguindo as diretrizes da *Foundation for Intelligent Physical Agents (FIPA)*, que define padrões de software de computador para ambientes multiagentes. Cada agente possui uma estrutura interna customizável mostrada na Fig. 2.

Figura 2: Estrutura interna de um agente Jemo.



Fonte: Dos autores (2022)

O processo de treinamento representado na Fig. 2 é fundamentado por meio do aprendizado pela observação, pela identificação de padrões e pela linguagem natural do raciocínio. Para os processos rápidos há um conjunto de redes neurais orientadas a aprendizado profundo onde podem ser processados dados numéricos ou simbólicos. Para a análise de dados há dois métodos evidenciados na plataforma, análise numérica e análise teórica, ambos métodos compartilham os resultados na “memória de trabalho”, parte do processamento de dados que armazena as informações a curto prazo.

A unidade dos processos de raciocínio é constituída por uma máquina de inferência *bayesiana estocástica*, ou seja, é capaz de realizar raciocínio lógico, processar ontologias, aprender novas regras e verificar a probabilidade de ocorrência. Estes processos são lentos e não acompanham diretamente as atividades da unidade dos processos rápidos. A memória de trabalho implementa um quadro branco com memória de curto prazo, permitindo uma comunicação indireta entre a unidade de processos rápidos e a unidade de processos de raciocínio por meio da leitura e escrita de mensagens no quadro. A unidade de planejamento das ações é um sistema de otimização que agenda uma lista de ações a executar com diferentes prioridades.

### 3 PROCEDIMENTO EXPERIMENTAL

Para a realização dos relatórios finais é preciso conhecimentos estatísticos, além da programação em sistemas que carregam os dados. Inicialmente é realizado o recolhimento dos dados no site do ONS referente ao histórico da matriz energética nacional dos últimos anos. A sintetização dos dados é realizada pelas medidas de estatística descritiva aplicadas na programação na linguagem em Python. A fase final



é o relatório de estudo, contendo os aspectos analisados e conclusões utilizando métodos de raciocínio (SILVESTRE, 2007).

No primeiro momento é implementado o algoritmo utilizando as técnicas de *machine learning* para a programação e métodos estatísticos para a interpretação dos dados obtidos. O primeiro teste é com algoritmos mais simples de Regressão Linear, já no segundo teste utiliza-se a IA por meio das redes neurais artificiais. Como programação complementar utiliza-se a plataforma Jemo, um *framework* em Java desenvolvido em uma tese de doutorado em 2019 (SALDIAS, 2019).

### 3.1 DADOS E PREMISSAS

Inicialmente, para a elaboração do modelo de predição de dados, as premissas foram selecionadas considerando os principais fatores sociais, econômicos e ambientais. A coleta e exploração de dados ocorreu de forma individual, onde cada tipo de informação demanda metodologias próprias de importação. Sendo assim, os processos iniciais da coleta de premissas são descritos separadamente.

A metodologia de estatística utilizada neste processo é a descritiva, onde há o processo de organização e demonstração dos resultados. Atrelando os termos estatísticos ao estudo em questão têm-se a matriz elétrica como experimento, o espaço amostral sendo a matriz elétrica nacional e os eventos como os resultados finais da predição. Os dados utilizados serão variáveis quantitativas do tipo contínuas de toda a população, ou seja, não é analisado apenas um grupo seletivo de informações, todos os dados de geração são computados no modelo. Para a viabilização da predição o treinamento do modelo é feito de forma individual por tipo de geração, ou seja, para prever a geração de energia eólica os dados de outros tipos de geração são utilizados como premissas para essa análise.

Para os dados da geração de energia nuclear foi considerada a geração das duas usinas localizadas na região sudeste do país, Angra 1 atuando desde 1964, Angra 2 desde 2001 e Angra 3 que possui previsão de inauguração próximo à 2028. Durante os 22 anos analisados houve pouca variação, obtendo um desvio médio de aproximadamente 127 MW, não havendo um crescimento ou decréscimo definido de geração de energia nuclear. Devido à baixa variação de dados, não foi realizado o treinamento de dados para a predição, considerando apenas o acréscimo da terceira usina que irá interferir na geração anual de energia nuclear.



### 3.1.1 Dados do Operador Nacional do Sistema Elétrico

Os dados que foram coletados do site ONS contém informações de carga por subsistema em uma determinada data, medida em *MWmed*, *megawatt* médio (ONS, 2022c). A apuração de informação é feita desde o ano 2000 até a última atualização carregada na plataforma na data em que a coleta de dados foi feita, em abril de 2022. Também foi coletado da mesma forma os dados de geração medidos em *MWmed* (ONS, 2022d). Os dados foram baixados em formato “csv”, um tipo de arquivo que pode ser lido pelo Microsoft Excel que separa os valores por vírgula, e separados de acordo com região e mês de referência.

Os dados hidrológicos também foram coletados no mesmo ambiente, da ONS (ONS, 2022e), sendo utilizados como premissas. As informações hídricas se referem ao nível de armazenamento de água nos reservatórios, medidos em metros cúbicos. Os dados de geração e cargas foram agrupados com as informações hidrológicas em formato de uma tabela.

### 3.1.2 Dados de precipitação da chuva

Para a melhor aproximação dos dados reais, é acrescentado o histórico da precipitação de chuva a nível nacional. Os dados históricos anuais foram coletados diretamente do Instituto Nacional de Meteorologia (INMET), registrados por hora em milímetros, “mm” (INMET, 2022). Foi considerada a média de todas as medições realizadas pelas estações a nível nacional. O resultado dos dados coletados por mês da precipitação de chuva, considerando a média em milímetros (“mm”) por mês, foram acrescentados na tabela de premissas, juntamente com os dados de geração, carga e armazenamento hídrico.

### 3.1.3 Dados financeiros e sociais

As informações complementares incluem: histórico do valor econômico do dólar; população das regiões analisadas; valor nominal e em porcentagem do Produto Interno Bruto (PIB); inflação acumulada por ano do Índice Nacional de Preços ao Consumidor Amplo (IPCA); salário mínimo em dezembro do ano em reais, valor da



exportação e importação em dólares do frete na modalidade *Free on Board* (FOB); saldo da balança comercial em dólares do FOB; anos em que ocorreram ou não eleições. Esses indicativos foram acrescentados na tabela juntamente com as demais premissas numéricas.

### 3.1.4 Premissas teóricas

Para complementar a programação em *Machine Learning* na área de raciocínio, foram acrescentados alguns fatores de análise teórica para incrementar o modelo de predição de dados elaborado pelo *framework* em Java, Jemo. Seguem descritas as premissas consideradas do estudo da PNE (EPE, 2018):

- A taxa de crescimento do montante de autoprodução de energia não injetada na rede é de 2,5% ao ano.
- Estimativa de ganho de eficiência energética global de 17%.
- O potencial hidrelétrico inventariado do Brasil é de 176 GW, sendo 108 GW em operação e construção e 68 GW de potencial hidrelétrico inventariado.
- Espera-se uma expansão significativa da fonte solar fotovoltaica, levando em conta apenas a geração centralizada, a fonte solar fotovoltaica atinge aproximadamente entre 27 a 90 GW em termos de capacidade instalada e entre 8 a 26 GW médios em termos de energia em 2050.
- Potencial energético brasileiro em TWh de fontes não renováveis em 2050: petróleo com 105.216,61, gás natural com 34.029,38, carvão mineral com 83.235,91 e urânio com 28.039,93.
- Potencial energético brasileiro em TWh de fontes renováveis em 2050: biomassa com 6.175,53; hidráulica com 860,62; eólica onshore com 348,90; eólica offshore com 15.770,28; solar onshore com 500,090; heliotérmica com 662,91; solar offshore com 61.022,61 e oceânica com 395,42.

Também foram consideradas outras premissas sociais, energéticas, econômicas e naturais:

- A previsão para a população do Brasil em 2050 é de 232.933.276 habitantes (IBGE, 2020).
- De acordo com artigo 27 da lei de número 14.300/2022, o faturamento será por incidência, ou seja, sobre a energia elétrica consumida e sobre a demanda. O faturamento das unidades particulares será escalonado, sendo que a partir de 2029

as tarifas estarão sujeitas às regras tarifárias estabelecidas pela ANEEL: 15% a partir de 2023; 30% a partir de 2024; 45% a partir de 2025; 60% a partir de 2026; 75% a partir de 2027; 90% a partir de 2028 (BRASIL, 2022).

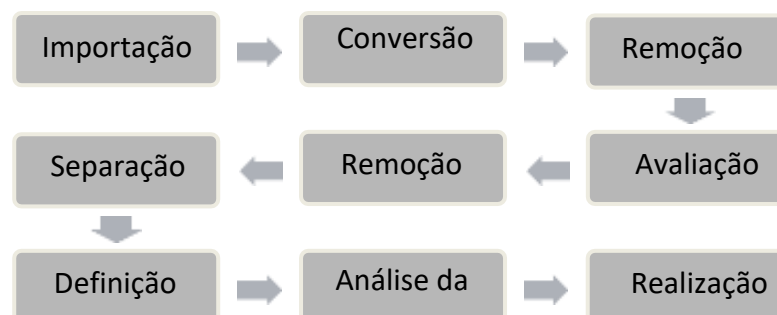
- De acordo com a Agência Nacional de Águas Saneamento Básico, ANA, o cenário de redução da disponibilidade hídrica será de aproximadamente 5%, previsão de acordo com alterações nas UGRH's, Unidades de Gestão de Recursos Hídricos (ANA, 2022).

- Para a geração nuclear foi considerada a potência instalada da Angra 1 de 640 MW e da Angra 2 de 1350 MW. Está previsto o início das atividades da Angra 3 próximo à 2028 com 1405 MW (ELETRONUCLEAR, 2022).

### 3.2 PREVISÃO DE SÉRIES TEMPORAIS

O processo de programação se iniciou com a instalação das bibliotecas e importação e análise dos dados, logo em seguida com o tratamento dos mesmos, transformando dados de data para o tipo específico que fará a leitura do tempo. Retirados os campos com ausência de informação, denominados como *NAN*, avaliada a correlação entre as variáveis e removida as informações irrelevantes, separados dados entre treino, com 70%, e os dados de teste, com os 30% restantes. Por último, é realizada a estrutura do modelo para o treinamento dos dados e a elaboração das previsões.

Figura 3: Processo de programação.



Fonte: Dos autores (2022)

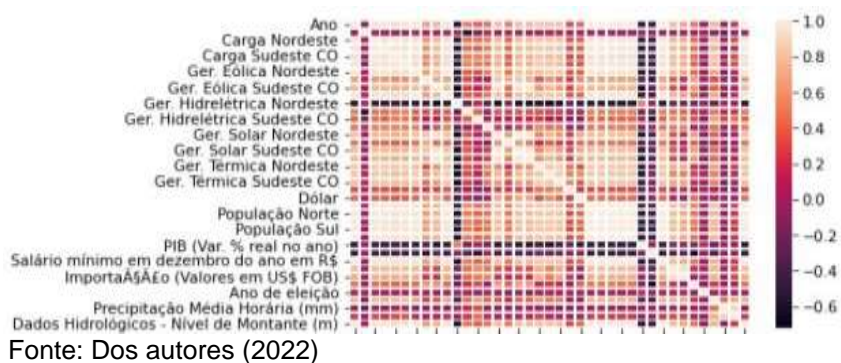
#### 3.2.1 Método de Regressão linear

O processo de programação se inicia com a importação e tratamento de dados, bem como a análise de correlação entre as informações para o treinamento do

modelo. As bibliotecas importadas foram: *pandas*, *numpy*, *seaborn*, *matplotlib.pyplot*, *datetime* e *sklearn.model\_selection* para o tratamento de dados e *sklearn.linear\_model* e *sklearn.metrics* para a criação do modelo de regressão linear. Os dados foram importados diretamente do armazenamento virtual, sendo reconhecidos pelo caminho do diretório referenciado na programação. Depois de importar as bibliotecas e os dados, há o tratamento das informações importadas para que leitura e manipulação das mesmas seja possível na plataforma. Sendo assim, foi ajustado os dados de data, removido os espaços vazios, verificado a correlação entre os elementos e removidas as premissas que tiveram baixa correlação com a geração em estudo. O primeiro processo de avaliação de dados é feito com a regressão linear, carregando o modelo com os dados que obtiveram boa correlação com a variável em estudo.

O processo de análise se inicia com a correlação entre as variáveis. O gráfico de correlação gerado pode ser observado na Fig. 4, onde ilustra de forma mais generalizada os resultados obtidos. Quanto mais forte for a cor respectiva da relação entre variáveis indica maior correlação entre as mesmas (valor próximo a 1), por exemplo, o grau de correlação dos dados de geração de energia eólica com a evolução dos anos possui o valor de 0,97, indicando que a variação do ano analisado vai impactar de forma significativa nos resultados de geração.

Figura 4: Programação - Gráfico de correlação dos dados.



Analisando de modo geral a Fig. 4, nota-se baixa correlação com o mês, com o valor do dólar, com a maioria das premissas econômicas e com a precipitação de chuva. Já as correlações que deram consideravelmente influentes, considerando acima de 50%, são atrelados os valores de cargas, ao ano, mais de 60% com o registro populacional e com o salário mínimo registrado em dezembro do ano

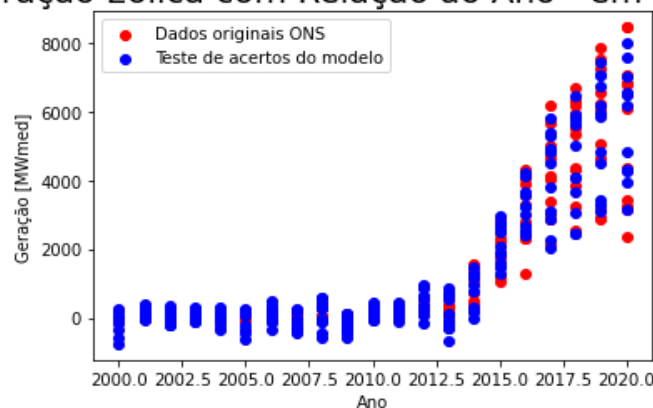


analisado, e 61% de correlação com os dados hidrológicos nos reservatórios.

Ao analisar o método de regressão linear, observa-se pela Fig. 5 que a métrica de erros analisada deu extremamente alta, inviabilizando a previsão por meio desse método. O Erro Quadrático Médio (MSE) calculou o valor aproximado de um milhão (obtendo apenas 20% de treinamento) para previsão de 6 anos à frente, sendo que esse tipo de erro é o mais apropriado para validação dos modelos, pois dá maior peso aos maiores erros. Ressaltando-se que os indicativos em vermelho se referem aos dados utilizados para o treinamento, enquanto os da cor azul indicam as previsões dos mesmos.

Figura 5: Programação - Gráfico dos resultados da Regressão Linear e seus erros.

#### Geração Eólica com Relação ao Ano - em MWmed



Fonte: Dos autores (2022)

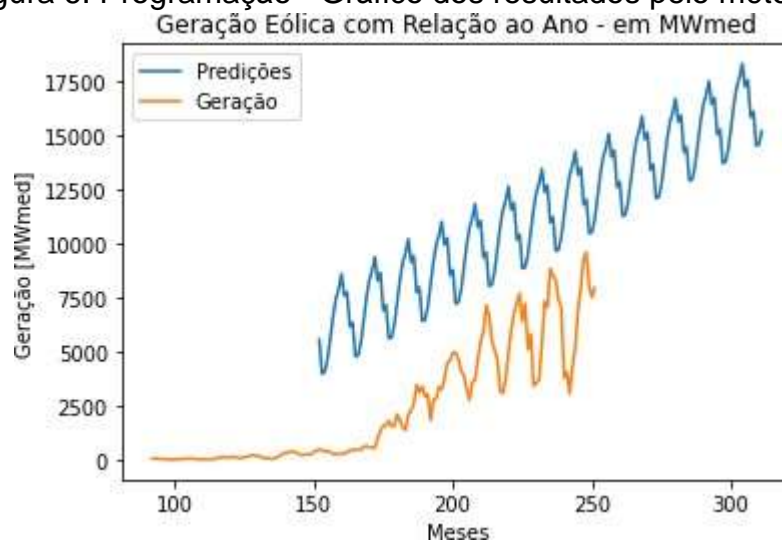
Para que haja um bom treinamento de dados na regressão linear os dados devem ter: correlação, tendência, sazonalidade e estacionalidade. Sendo assim, para que o modelo obtenha boa performance as informações devem possuir alta correlação e um padrão de crescimento conhecido (SILVESTRE, 2007). Portanto, a aleatoriedade dos dados ao longo do tempo deve possuir uma média constante e um padrão definido. Pela análise prévia dos dados coletados já é notada a ausência de um padrão definido, sendo essa análise confirmada com o cálculo do erro posteriormente, contudo, o resultado do modelo estatístico por meio de série temporal é ineficiente para esse tipo de dados.

### 3.2.2 Método de *Arima*

Durante o processo de treinamento e análise dos modelos foi realizado também o método *Arima*, cujos resultados são demonstrados na Fig. 6, outro método

de séries temporais cujos resultados não foram satisfatórios, mesmo como a configuração dos parâmetros gerada automaticamente pela técnica de *auto-arima*, apesar do treinamento de 78% para previsão de 15 anos à frente, o erro para 2050 passou de 100%. Portanto, por meio dessa outra técnica realizada evidenciou novamente a ineficiência da utilização de técnicas auto recorrentes para esse tipo de aplicação proposta, inviabilizada pela ausência de estacionalidade nas séries, onde as características estatísticas são variáveis.

Figura 6: Programação - Gráfico dos resultados pelo método *Arima*.



Fonte: Dos autores (2022)

A ausência de estacionalidade dos dados pode ser corrigida em alguns casos alterando as séries originais, porém, para essa situação em específico, a identificação de padrões recorrentes é inviabilizada pela ausência de uma sequência lógica de dados ao longo do tempo. Portanto, optou-se por manter as séries originais para não supor um padrão que é inexistente, evitando um maior erro nas previsões.

### 3.2.3 Método com Redes Neurais Recorrentes

Com o método de regressão linear não satisfatório inicia-se o processo de treinamento de dados utilizando redes neurais artificiais. É elaborada uma tarefa de “Regra de Associação”, uma metodologia disponibilizada no aprendizado de máquina a qual analisa o histórico para prever futuros eventos. Para o treinamento dos dados em IA o processo de análise de correlação entre variáveis é feito automaticamente pela rede neural. O restante do processo de importação e tratamento de dados é o



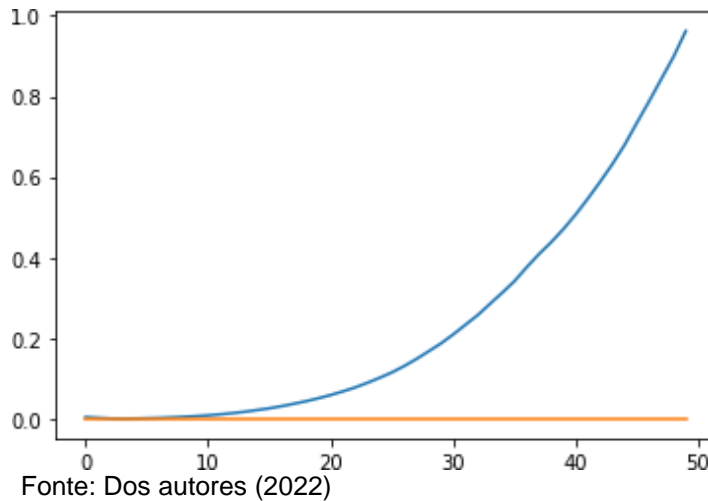
mesmo que foi realizado para o modelo da regressão linear, sendo acrescentado apenas algumas bibliotecas específicas para a programação de rede neural. As bibliotecas importadas foram: *pandas*, *numpy*, *seaborn*, *matplotlib.pyplot*, *datetime* e *sklearn.model\_selection*, *keras.models*, *keras.layers*, *sklearn.metrics*, *keras.callbacks*. Depois do tratamento de dados, a próxima etapa é apenas de parametrização do modelo, onde há a definição das camadas, dos neurônios de entrada e de saída e da função de ativação. Os dados foram separados em 70% de treino e 30% de teste, deixando apenas os dados da geração em estudo definida como variável dependente, “y”, e as premissas como variáveis independentes, “x”.

A função de ativação foi definida como do tipo *relu*, a quantidade de neurônios foi estabelecida como 32 para as duas primeiras camadas e 64 para a terceira camada. Nota-se, ainda, o tipo de rede neural unidimensional utilizada, *sequencial*, caracterizada pela programação em Python, onde a saída de uma camada é conectada à entrada da próxima.

A segunda etapa é de compilação do modelo, onde os dados são treinados de acordo com os parâmetros definidos. Entre os parâmetros destacam-se a limitação de processamento devido o *EarlyStopping*, a função de perdas *categorical\_crossentropy* para verificar a diferença entre o resultado obtido e o esperado, o algoritmo de otimização do modelo *adam*, o *accuracy* como métrica para calcular a taxa de acerto e 50 *epochs* que é a definição da quantidade de vezes que o sistema irá repetir o processo. Como há o ajuste de pesos nas saídas das camadas esse método de rede neural é classificado como recorrente.

De acordo com a programação e o modelo criado, foi gerada a métrica de erro do mesmo, a linha azul representa a evolução do erro, enquanto a linha amarela indica a performance da rede, que aumentaria conforme o erro seria reduzido. De acordo com o gráfico gerado, ilustrado na Fig. 7, é evidente o baixo desempenho do modelo, indicando necessidade de melhoria no processo de treinamento da rede.

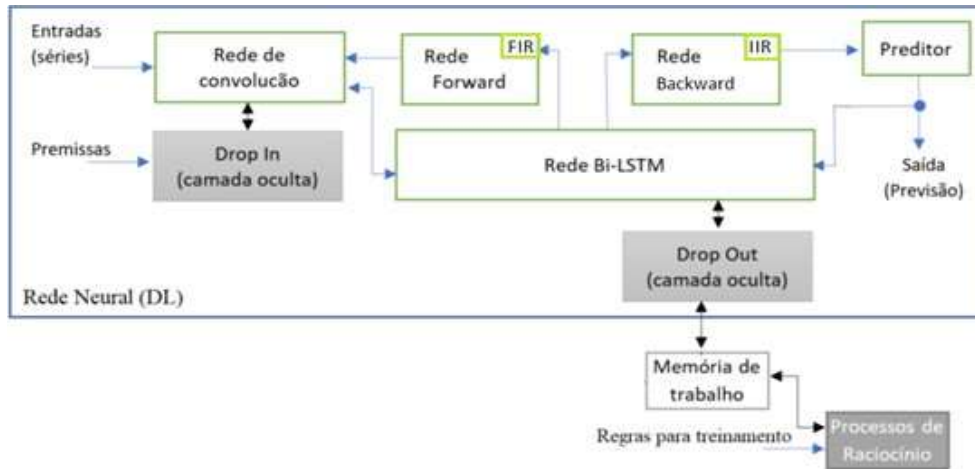
Figura 7: Gráfico de erro de dispersão do modelo.



A evolução constante do erro apurado pelo modelo é justificada pela ausência de informações externas. Com o início da elaboração do modelo é notória a necessidade da utilização da IA para as predições da matriz energética nacional, considerando a ampla relação das cargas de energia com fatores externos. Porém, devido à ligação direta da energia com as imprevisibilidades sociais, apenas os dados de geração e carga de energia elétrica não são o suficiente para realizar as predições. Deve-se, portanto, incluir premissas teóricas além de numéricas, sabendo que apenas os dados históricos não são o suficiente para a boa predição do modelo. Outro importante fator que influencia no treinamento do modelo é a relação de dados históricos e dados futuros, pois há aproximadamente 20 anos de banco de dados que serão utilizados para prever 30 anos à frente. Método complementar usando a plataforma Jemo

Foi implementada uma estrutura usando o *framework* Jemo focada para previsão de séries temporais baseadas em múltiplas variáveis, combinando diferentes redes neurais com um motor de regras para processos de raciocínio, conforme representado na Fig.8. A rede neural interage com a unidade de processos de raciocínio apenas no treinamento. As regras que guiam o treinamento foram estruturadas no formato *Rule Engine* (mecanismo de regras), também conhecido na área de programação como CLP, na unidade de processos de raciocínio.

Figura 8: Estrutura adaptada da plataforma Jemo.



Fonte: Dos autores (2022)

- Rede de convolução: lê, classifica e extrai informação e padrões da entrada que relacionam as variáveis em pares.
- Rede *forward*: filtra incoerências na entrada, é uma unidade de atraso discreto implementando uma rede IIR, *infinite impulse response*.
- Rede *Backward*: filtra incoerências na saída, é um conjunto de neurônios com unidades de atraso discreto implementando uma rede FIR, *finite impulse response*, serve para reforçar padrões e filtrar dados ruidosos, ou irrelevantes, realimentando assim o aprendizado da rede Bi-LSTM.
- Preditor: responsável pela interpolação da série gerada de forma a manter a média móvel estável.
- Rede Bi-LSTM: constrói parte das sequências de dados de previsões em base aos padrões aprendidos. *Drop In/Out*: camada oculta que age apenas durante o processo de treinamento, evita generalidades, insere e analisa as premissas e otimiza o modelo, otimizador para essa aplicação foi escolhido o *adam*.

Segue descrito na Tab. 1 as definições iniciais projetadas.



Tabela 1: Parâmetros do treinamento da rede neural BI-LSTM.

**Definição dos parâmetros**

opt.trPercentage = 0.8; //80% para treinamento, 20% testes  
opt.NumOfHiddenLayers = 2; //camada entrada, camada saída + 2 camadas ocultas  
opt.NumOfUnitsInFirstlayer = 100; //unidades LSTM camada entrada  
opt.NumOfUnitsInSecondlayer = 100; //unidades LSTM 1ª camada oculta  
opt.NumOfUnitsInThirdlayer = 75; //unidades LSTM 2ª camada oculta  
opt.NumOfUnitsInFourthlayer = 75; //unidades LSTM camada saída  
opt.maxEpochs = 200; //máximo de épocas  
opt.NumOfFeedForwardLayers = 2; //realimentações entre camadas  
opt.NumOfNeuronsInFirstlayer = 15; //neurônios camada entrada  
opt.NumOfNeuronsInSecondlayer = 15; //neurônios 1ª camada oculta  
opt.NumOfNeuronsInThirdlayer = 10; //neur. 2ª camada oculta //1 neurônio de saída

Fonte: Dos autores (2022)

O modelo pelo método complementar utilizando a plataforma Jemo obteve melhor resultado, sendo destacado na Tab. 2 a relação dos parâmetros resultantes, nota-se o registro baixo do erro calculado.

Tabela 2: Resultado do treinamento dos parâmetros.

<b>Epoch (Períodos de iterações)</b>	<b>Iteration (Iterações)</b>	<b>Time Elapsed hh:mm:ss (Tempo de execução)</b>	<b>Mini-batch RMSE (Erro absoluto)</b>	<b>Mini-batch - Loss (Erro percentual)</b>	<b>Base Learning Rate (Taxa base de aprendizado)</b>
1	1	00:00:00	0.65	0.2	0.0050
17	50	00:00:03	0.25	3.1e-02	0.0050
34	100	00:00:06	0.22	2.4e-02	0.0050
50	150	00:00:08	0.23	2.6e-02	0.0050
67	200	00:00:11	0.23	2.7e-02	0.0050
84	250	00:00:14	0.17	1.4e-02	0.0050
100	300	00:00:17	0.23	2.7e-02	0.0050
117	350	00:00:19	0.25	3.1e-02	0.0050
134	400	00:00:22	0.19	1.7e-02	0.0010

Fonte: Dos autores (2022)



#### 4 RESULTADOS E DISCUSSÃO

Os métodos estocásticos não obtiveram bons resultados no processo de aprendizagem de máquina, erro justificado pela ausência de estacionalidade dos dados que foram utilizados no processo de treinamento e o horizonte distante das previsões, tentar prever 30 anos a partir de 20 anos de dados conhecidos. O indicativo de dependência dos dados com fatores externos foi evidenciado pela falha obtida ao realizar o teste de *Newcomb-Benford*, resultando como “não aceitável” para todos os conjuntos de dados de geração. Outro indicativo foi pelo teste de *Dickey-Fuller*, reafirmando a ausência de estacionalidade dos dados, ou seja, as características estatísticas das séries são variáveis. Ambos testes foram realizados pela plataforma online Octave. Sendo assim, o principal fator que inviabilizou a técnica de série temporal é a evidência de uma forte correlação dos dados históricos coletados com fatores externos, gerando instabilidade.

O indicativo de bom treinamento do modelo pode ser observado na Tab. 3 no índice de treino, notando-se as gerações que possuíram o melhor treinamento dos dados. Para a análise dos resultados, foi considerado o indicativo acima de 90% como bom, acima de 80% como médio e abaixo de 70% como baixo nível de treinamento. No geral, as previsões ficaram com a performance de treinamento acima de 70%. Segue representado na Tab. 3 os resultados dos dados de geração com os testes de estacionalidade e de performance do modelo.

**Tabela 3: Conformidade com os testes de *Newcomb-Benford* e de *Dickey-Fuller*.**

<b>Tipo de Geração</b>	<b>Teste Newcomb-Benford</b>	<b>Dickey-Fuller (<math>\delta</math>)</b>	<b>Treino (%)</b>	<b>Tipo de Geração</b>	<b>Teste Newcomb-Benford</b>	<b>Dickey-Fuller (<math>\delta</math>)</b>	<b>Treino (%)</b>
Eólica Nordeste	0.0238374	-0,12	91,47	Solar Nordeste	0.0571227	-0,14	81,24
Eólica Norte	0.0421212	-0,16	74,67	Solar Norte	0.0872034	-0,42	70,22
Eólica Sudeste CO	0.048207	-0,22	89,26	Solar Sudeste CO	0.0782219	-0,06	88,74
Eólica Sul	0.068728	-0,09	81,02	Solar Sul	0.0835605	-0,11	71,44
Hidrelétrica Nordeste	0.0824065	-0,09	88,71	Térmica Nordeste	0.0424905	-0,41	70,14



Hidrelétrica Norte	0.0645676	-0,32	91,03	Térmica Norte	0.0666485	-1	88,24
Hidrelétrica Sudeste CO	0.147864	-0,32	90,94	Térmica Sudeste CO	0.0525074	-0,77	90,13
Hidrelétrica Sul	0.0657977	-0,07	90,94	Térmica Sul	0.0918825	-0,38	91,02

Fonte: Dos autores (2022)

Demonstra-se abaixo na Tab. 4 o valor em *MW<sub>med</sub>* anual de geração de energia elétrica pelo modelo criado em Jemo, comparando com o cenário da matriz elétrica do PNE (EPE, 2020b) que mais se encaixou com os resultados encontrados, mostrando a diferença das previsões do modelo com o cenário da PNE.

Tabela 4: Resultados da predição de geração em *MW<sub>med</sub>* pelo modelo Jemo.

	2030	Diferença (%)	2040	Diferença (%)	2050	Diferença (%)
Hidrelétrica	62.970,54	1,24	60.867,72	4,60	59.416,09	13,77
Eólica	13.967,83	-44,40	20.885,28	-62,26	26.581,05	-69,12
Solar	11.824,52	263,38	26.501,37	103,67	34.200,75	36,27
Térmica	11.943,2	-4,05	9.803,6	-11,80	6.715,3	-19,80
Nuclear	3.112,08	3,70	3.112,08	5,03	3.112,08	6,65

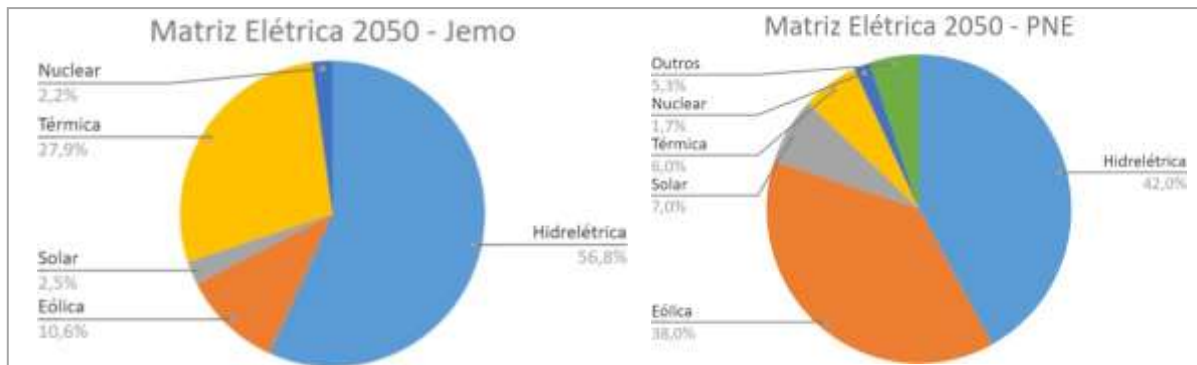
Fonte: Dos autores (2022)

O cenário geral comparado na Tab. 4 considera apenas usinas hidrelétricas sem áreas sem interferência em unidades de conservação ou em terras indígenas e quilombolas, cujo cenário principal do PNE presume “Efeitos das Mudanças Climáticas (Redução de disponibilidade hídrica) sem emissões” (EPE, 2020b, p.11). Vale ressaltar que os três cenários mais próximos das previsões realizadas consideram a redução de emissão de gases poluentes.

Com relação a porcentagem dos tipos de gerações na matriz elétrica o cenário que mais se aproximou considera todo potencial hidrelétrico inventariado disponível, cujo cenário principal do PNE presume “Matriz Elétrica com expansão a partir de tecnologias não emissoras de gás de efeito estufa” (EPE, 2020b, p.12). A comparação da matriz elétrica prevista para o ano de 2050 com o cenário da PNE segue representada na Fig. 9.



Figura 9: Comparação da matriz elétrica para 2050.



Fonte: Dos autores (2022)

Apesar de comparar os resultados com o cenário mais próximo sugerido pelo PNE, a diferença total dos resultados varia em torno de 50%. A alta variação dos resultados encontrados se deve pela divergência de dados base, demonstrados nas gerações registradas do ano de 2015, e também pelas diferentes premissas utilizadas para o treinamento do modelo. A principal diferença na proporção de gerações para o ano de 2050 comparado ao cenário do PNE é na energia eólica e solar, sendo que a previsão realizada pela plataforma Jemo identificou menos geração eólica e mais geração solar que as previstas pelo PNE.

## 5 CONCLUSÃO

Ao analisar os dados históricos de geração de energia elétrica notou-se a ausência de estacionalidade das séries, sendo verificado com os testes de *Newcomb-Benford* e de *Dickey-Fuller*, evidenciando a ausência de um crescimento padrão ao longo dos anos, bem como a carência de tendência, de correlação, de variância e de sazonalidade constantes. Entende-se, portanto, que o processo de estudo da futura matriz elétrica nacional não considera apenas os dados históricos de carga e geração, mas sim, diversas outras premissas que influenciam nas decisões. De acordo com o PNE 2050, o potencial de recursos energéticos na economia nacional é impactado por estruturas legais e regulatórias e por políticas públicas e de inovações tecnológicas. Sendo assim, o estudo do PNE contribui para a análise inicial de uma provável realidade, não com o objetivo de prever o futuro, mas sim, de auxiliar nas decisões nas áreas políticas, sociais, econômicas e ambientais (EPE, 2020a).

A ausência de estacionalidade nos dados inviabilizou os métodos



estatísticos de séries temporais implementados na regressão linear e nas técnicas auto recorrentes. A rede neural LSTM por si só também não conseguiu atingir bons resultados, sua estrutura unidimensional e a combinação de dados e parâmetros não foram adequadas para esse tipo de previsão, considerando que redes neurais recorrentes possuem boa performance apenas a curto prazo. Deste modo, observou-se que as redes neurais por si só não são o suficiente para desenvolver um bom modelo de previsão de dados. Ressalta-se, ainda, que para o processo de treinamento dos modelos há dados que possuem pouco histórico de geração, o que torna as previsões mais instáveis devido o prazo de previsão ser mais extenso que o próprio registro dos dados usados para o aprendizado. Outro fator influente é a diferença no banco de dados, pois os dados históricos coletados do ONS são diferentes dos dados utilizados no PNE.

O método de aprendizado de dados por meio da plataforma Jemo mostrou-se promissora para o estudo das previsões, com alta performance e semelhança próxima aos resultados publicados pelo PNE em 2020 (EPE, 2020b). A IA utilizada obteve um bom empenho, entendendo-se que redes neurais recorrentes possuem boa performance para processar eventos correlacionados, realizando o ajuste de pesos e mantendo a correlação entre dados. Além do ajuste de pesos, também houve a melhoria do modelo pela utilização de LSTM, um tipo de rede neural recorrente que consegue manter uma memória.

Contudo, de acordo com as previsões realizadas na estrutura implementada na plataforma Jemo, o cenário da matriz elétrica para o ano de 2050 se enquadrará em cenários que sugerem a redução de emissão de gases poluentes e influência por mudanças climáticas. Entendendo que a emissão de gases poluentes dificilmente será erradicada, sendo assim, interpreta-se o resultado como uma possível diminuição nos impactos ambientais. A redução da emissão de gases poluentes é ainda justificada pela diminuição de energias térmicas. Os resultados do cenário sem emissão de gases poluentes vão de encontro com as metas estipuladas pela Conferência das Nações Unidas de Mudanças Climáticas, COP26, objetivando a neutralização de gases de efeito estufa até 2050 (COP26. 2021). Outras considerações é a diminuição de energia hidrelétrica, ainda no mesmo cenário de conservação natural, e também o aumento da energia solar, entendendo que esses aumentos surgem para a substituição dos métodos de geração de energia elétrica que possuem maior impacto direto ao meio ambiente.



Ademais, conclui-se que o estudo aprofundado do setor de energia é imprescindível para as ações internas e externas do país, entendendo que a estratégia de desenvolvimento progressivo está diretamente atrelada com as decisões embasadas em possíveis cenários de desenvolvimento. Sugere-se como pesquisas futuras a utilização de premissas políticas para que os resultados sejam mais assertivos, considerando a ampla participação da política na matriz energética nacional brasileira. Deve ser considerado o processo burocrático do mercado de energia, analisando as políticas de decisões das empresas de geração e do governo, o mercado atual e também o cenário econômico.

## REFERÊNCIAS

ANA. **Cenário de Redução da Disponibilidade Hídrica (Classes de convergência por nível de alteração na UGRH - Redução Superior a 5%)**. [S. l.], 14 set. 2022. Disponível em: <https://dados.gov.br/dataset/cenario-de-reducao-da-disponibilidade-hidrica-classes-de-convergencia-por-nivel-de-alteracao-na>. Acesso em: 19 set. 2022.

BRASIL. **Lei nº 14.300, de 06 de janeiro de 2022**. Institui o marco legal da microgeração e minigeração distribuída, o Sistema de Compensação de Energia Elétrica (SCEE) e o Programa de Energia Renovável Social (PERS); altera as Leis nºs 10.848, de 15 de março de 2004, e 9.427, de 26 de dezembro de 1996; e dá outras providências, Brasília, DF, 06 jan. 2022. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2019-2022/2022/lei/L14300.htm](http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2022/lei/L14300.htm). Acesso em: 19 set. 2022.

CHENG, Q., CHEN, Y., XIAO, Y. et al. **A dual-stage attention-based Bi-LSTM network for multivariate time series prediction**. J Supercomput 78, 16214–16235 (2022). Disponível em: <https://doi.org/10.1007/s11227-022-04506-3>. Acesso em: 13 out. 2022.

COP26. 2021, Scottish Event Campus, Glasgow, Reino Unido. COP26: THE GLASGOW CLIMATE PACT. Disponível em: <https://ukcop26.wpenginepowered.com/wp-content/uploads/2021/11/COP26-Presidency-Outcomes-The-Climate-Pact.pdf>. Acesso em: 19 nov. 2022.

CUNHA, F. C. R. da. **Aplicações da lei Newcomb-Benford à auditoria de obras públicas**. Orientador: Maurício Soares Bugarin. 2013. 486f. Dissertação (Mestrado em Regulação e Gestão de Negócios) - Universidade de Brasília, Brasília, 2013. Disponível em: <https://repositorio.unb.br/handle/10482/16379>. Acesso em: 13 out. 2022.

DICKEY, D. & FULLER, W. (1979), **Distribution of the estimators for autoregressive time series with a unit root**, Journal of the American Statistical Association 74(366), 427-431. Acesso em: 13 out. 2022.



ELETRONUCLEAR. **Angra 3**, 2022. Disponível em: <https://www.eletronuclear.gov.br/Nossas-Atividades/Paginas/Angra-3.aspx>. Acesso em: 19 nov. 2022.

EPE; MME. ANEEL. BRASIL. 07/2018. **Premissas e Custos da Oferta de Energia Elétrica**: No horizonte 2050, Rio de Janeiro, nov. 2018. Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-227/topico-456/NT%20PR%20007-2018%20Premissas%20e%20Custos%20Oferta%20de%20Energia%20EI%C3%A9trica.pdf>. Acesso em: 18 abr. 2022.

EPE. ANEEL. BRASIL. Plano Nacional de Energia 2050: Ministério de Minas e Energia. Empresa de Pesquisa Energética. **Premissas e Custos da Oferta de Energia Elétrica no Horizonte 2050**, Brasília, p. 1-127, nov. 2018. NT PR 07/18. Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-227/topico-456/NT%20PR%20009-2018%20Premissas%20e%20Custos%20Oferta%20de%20Combustiveis.pdf>. Acesso em: 03 maio 2022.

EPE. ANEEL. BRASIL. Plano Nacional de Energia 2050: Ministério de Minas e Energia. Empresa de Pesquisa Energética. **Plano Nacional de Energia**, Brasília, p. 1-230, 16 dez. 2020a. Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-227/topico-563/Relatorio%20Final%20do%20PNE%202050.pdf>. Acesso em: 03 maio 2022.

EPE. ANEEL. BRASIL. Plano Nacional de Energia 2050: Ministério de Minas e Energia. Empresa de Pesquisa Energética. **Anexo - Plano Nacional de Energia**, Brasília, p. 1-78, 16 dez. 2020b. Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-227/topico-563/PNE%202050%20-%20Anexo.pdf>. Acesso em: 03 maio 2022.

EPE. ANEEL. BRASIL. **Matriz energética e elétrica**. [S. l.], 2022a. Disponível em: <https://www.epe.gov.br/pt/abcdenergia/matriz-energetica-e-eletrica>. Acesso em: 01 jun. 2022.

EPE. ANEEL. BRASIL. Plano Nacional de Energia 2050: Ministério de Minas e Energia. Empresa de Pesquisa Energética. **Plano Nacional de Energia 2050**. [S. l.], 2022b. Disponível em: <https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/Plano-Nacional-de-Energia-2050>. Acesso em: 01 jun. 2022.

EPE. ANEEL. BRASIL. **Quem somos**. [S. l.], 2022c. Disponível em: <https://www.epe.gov.br/pt/a-epe/quem-somos>. Acesso em: 18 abr. 2022.

IBGE. **Projeções da Populações**: Projeções da População do Brasil e Unidades da Federação por sexo e idade simples: 2010-2060. [S. l.], 09 dez. 2020. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9109-projecao-da-populacao.html?=&t=resultados>. Acesso em: 19 set. 2022.

INMET. **Dados históricos anuais**. [S. l.], 2022. Disponível em:



<https://portal.inmet.gov.br/dadoshistoricos>. Acesso em: 3 set. 2022.

KELLEHER, John D. **Deep Learning**: (The MIT Press Essential Knowledge series). [S. l.]: The MIT Press, 2019. 295 p.

LEVINE, D.M; BERENSON, M.L.; STEPHAN, D. **Estatística: Teoria e Aplicações**. Rio de Janeiro: LTC, 2005. Acesso em: 27 out. 2022.

OGATA, Katsuhiko. **Engenharia de Controle Moderno**. [s. l.]: Pearson, 2015. Disponível em: c. Acesso em: 13 out. 2022.

ONS. ANEEL. BRASIL. **O que é ONS**. [S. l.], 2022a. Disponível em: c. Acesso em: 28 mar. 2022.

ONS. ANEEL. BRASIL. **Indicador de previsão de carga - IPC**. In: Indicador de previsão de cargas - IPC. [S. l.], 2022b. Disponível em: [http://www.ons.org.br/Paginas/sobre-ons/indicadores-ons/indicador\\_ipc.aspx](http://www.ons.org.br/Paginas/sobre-ons/indicadores-ons/indicador_ipc.aspx). Acesso em: 06 abr. 2022.

ONS. ANEEL. BRASIL. **Carga de energia**. [S. l.], 2022c. Disponível em: [http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/carga\\_energia.aspx](http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/carga_energia.aspx). Acesso em: 01 jun. 2022.

ONS. ANEEL. BRASIL. **Geração de energia**. [S. l.], 2022d. Disponível em: [http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/geracao\\_energia.aspx](http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/geracao_energia.aspx). Acesso em: 01 jun. 2022.

ONS. ANEEL. BRASIL. **Dados hidrológicos/níveis**. [S. l.], 2022e. Disponível em: [http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/dados\\_hidrologicos\\_niveis.aspx](http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/dados_hidrologicos_niveis.aspx). Acesso em: 04 out. 2022.

FARBER, Betsy; LARSON, Ron. **Estatística Aplicada**. 4. ed. [S. l.]: Pearson, 2010. 658 p. ISBN 9788576053729. Disponível em: <https://plataforma.bvirtual.com.br/Leitor/Publicacao/1242/pdf/0?code=GeB46XkE3T0xlacUJy44kqvLPTQJtJWR/c2ik0THSzBsZKscR/DasnUVeLvoJ7u1Sb8FNivBY8RQygFWUjPa3Q==>. Acesso em: 12 abr. 2022.

LOPES, Gesiel Rios; ALMEIDA, Alessandro Wilk Silva; DELBEM, Alexandre C. B.; TOLEDO, Cláudio Fabiano Motta. **Introdução à Análise Exploratória de Dados com Python**, [s. l.], v. 8, 2019. Disponível em: [https://www.researchgate.net/profile/Gesiel-Lopes/publication/336778766\\_Introducao\\_a\\_Analise\\_Exploratoria\\_de\\_Dados\\_com\\_Python/links/5db225d2a6fdccc99d9426f2/Introducao-a-Analise-Exploratoria-de-Dados-com-Python.pdf](https://www.researchgate.net/profile/Gesiel-Lopes/publication/336778766_Introducao_a_Analise_Exploratoria_de_Dados_com_Python/links/5db225d2a6fdccc99d9426f2/Introducao-a-Analise-Exploratoria-de-Dados-com-Python.pdf). Acesso em: 16 mar. 2022.

SALDIAS, Claudio Ernesto Ponce. **Protótipo de sistema de instrução virtual para treinamento de pilotos de aviões Boeing 737-800**. 2019. Tese (Doutorado) - Pós Graduação de Engenharia Mecânica, Universidade Federal de Santa Catarina, Florianópolis, 2019. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/215638/PEMC2025-T.pdf?sequence=-1&isAllowed=y>. Acesso em: 14 out. 2022.



STEFANINI, Rafael. **Quais os impactos da análise de dados na indústria 4.0**, [s. l.], 17 ago. 2020. Disponível em: <https://stefanini.com/pt-br/trends/artigos/impactos-da-analise-de-dados-na-industria>. Acesso em: 28 mar. 2022.

SILVESTRE, António Lupis. **Análise de Dados e Estatística Descritiva**. [s. l.]: Escolar Editora, 2007. ISBN 9789725922088. Acesso em: 16 mar. 2022.