

**Ao:** Enacom Group  
**Ao:** Dsc. Leonardo Tavares  
Machine learning manager at Enacom.

**Assunto:** Avaliação Bootcamp Seletivo

**Solicitante:** Marcelo Chalub Llanco

## 1.Principais descobertas na análise exploratória

Ao realizar a análise exploratória dos dados de produção de energia de diversas usinas ao longo dos anos de 2000 à 2018, insights significativos foram coletados, que serão apresentados nas imagens a seguir.

Na figura 01 é visto a produção total de energia na delimitação de período informado anteriormente. É observado a tendência de crescimento, sendo que em poucos anos teve uma regressão, mas não tão significativa quanto os crescimentos contínuos.

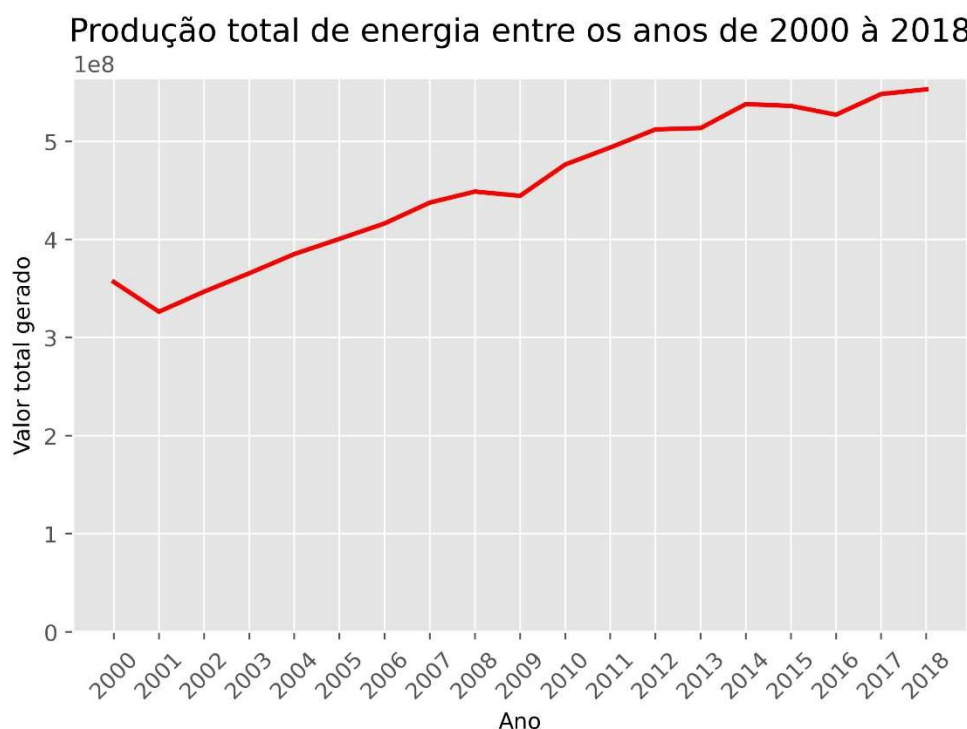


Figura 01 – Produção de energia entre os anos de 2000 à 2018

Em relação à porcentagem de produção, foi realizado um somatório para verificar a maior tendência de utilização energética e os possíveis crescimentos neste período, vistos na Figura 02. Percebeu-se que o bombeamento teve produção mínima comparada aos demais, assim como a fotovoltaica também apresentou essa defasagem.

Porcentagem de produção energética no Brasil durante os anos de 2000 à 2018

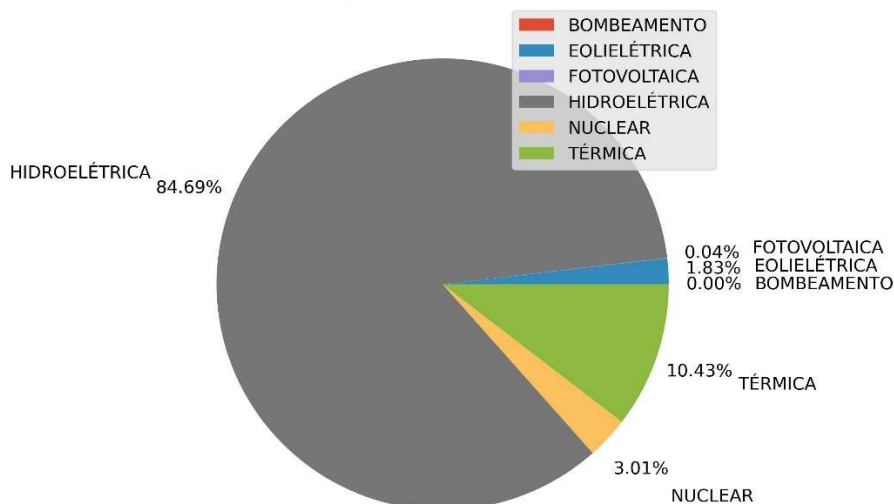


Figura 02 – Produção energética no Brasil durante os anos de 2000 à 2018.

Ao abordar essa porcentagem em valores numéricos evolutivos de acordo com o tempo, é visto na figura 03 que a energia Térmica possuía acréscimo quando a energia Hidroelétrica apresentou recaídas no período de 2011 à 2015, ou seja, diferentemente dos outros tipos de usina, ela se destacou. É percebido também que o tipo de usina Eolielétrica começou sua ascendência em 2013 até 2018.

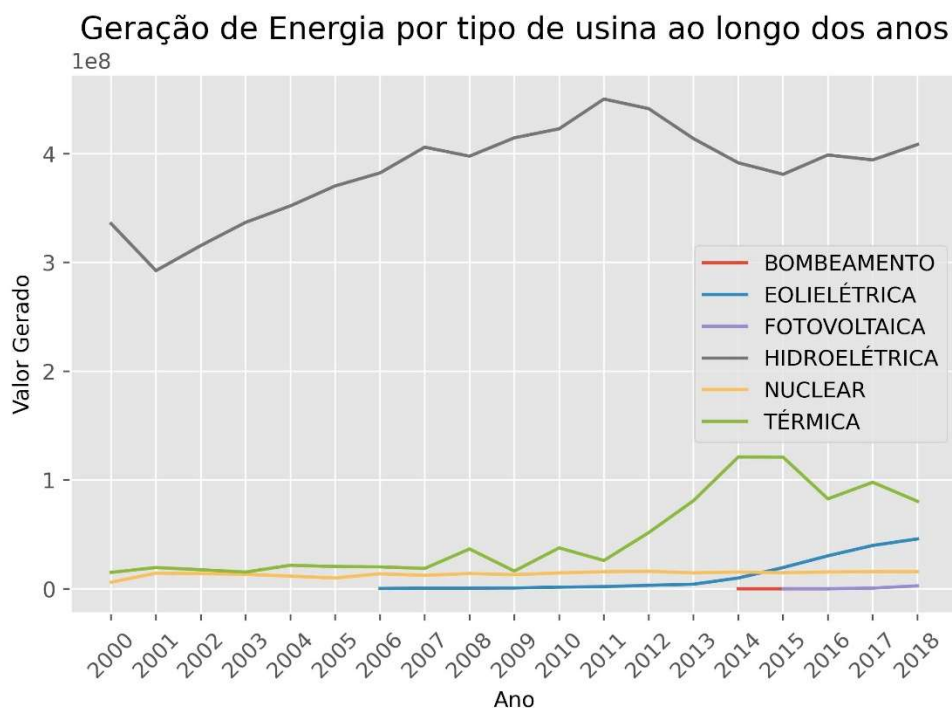


Figura 03 – Geração por tipo de usina (2000 à 2018)

## 2. Modelos utilizados e critérios de seleção

O modelo preditivo utilizado foi um modelo estatístico nomeado ARIMA (Autoregressive Integrated Moving Average) o qual é utilizado para análise e

previsão de séries temporais. Ele é composto por componentes de autorregressão, médias móveis e integração com a realização de diferenciações para a captura de padrões temporais. Este modelo está presente no Python a partir da biblioteca `pmdarima.arima`. Ele é realizado por comparação de Akaike Information Criterion (AIC) para escolher o melhor modelo dentre os realizados testes.

A partir disto, foi realizado uma análise e treinamento (fit) dos modelos com cada tipo de produção de energia. Como apresentado anteriormente nas Figuras 02 e 03, o tipo BOMBEAMENTO foi desconsiderado, visto que a produção se encerrou.

Além deste critério, outros foram definidos no código, como o período de sazonalidade, número de lags do modelo de acordo com o correlograma da função de autocorrelação (p), número de vezes que a serie será diferenciada (d) e número de lags do modelo de acordo com o correlograma da função de autocorrelação parcial – médias móveis (q).

### 3. Métrica e avaliação do modelo

Como métrica para avaliação do modelo, foi escolhido a comparativa onde, como o método realizado é teste e já possuíamos os dados gerados, realizou-se a avaliação a partir do levantamento desses valores.

Os valores de geração previstos, apresentados na Figura 04, apresentam uma tendência contínua em relação aos dados anteriores da Figura 03. A partir disto, é verificado que em um dos quesitos os valores previstos foram aproximados aqueles observados da série. Em comparação ao valor gerado, apresentados na Figura 05, observou-se a aproximação durante os períodos de alta produção Hidroelétrica, mantendo a mesma sazonalidade. Os outros meios, que possuem uma menor geração energética, também se mantiveram equilibrados na geração prevista e que, na geração da Figura 05, houve uma deformidade em relação a sua forma anterior dos anos 2000 a 2018.

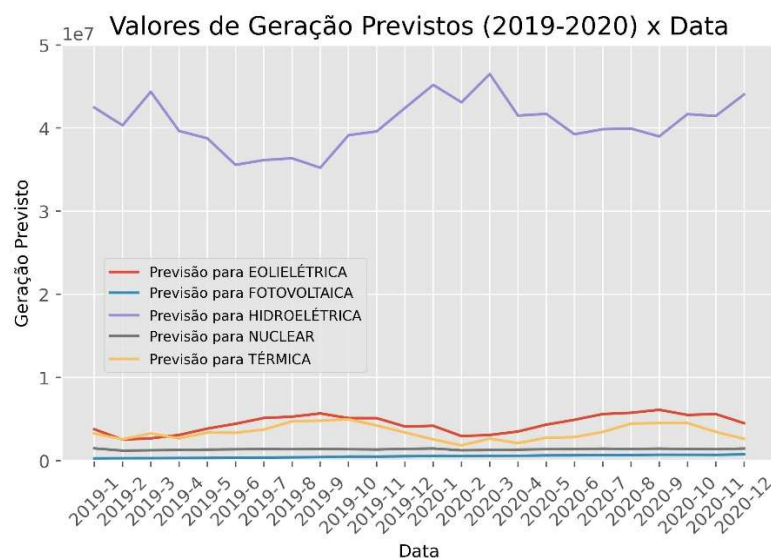


Figura 04 – Geração prevista 2019-2020

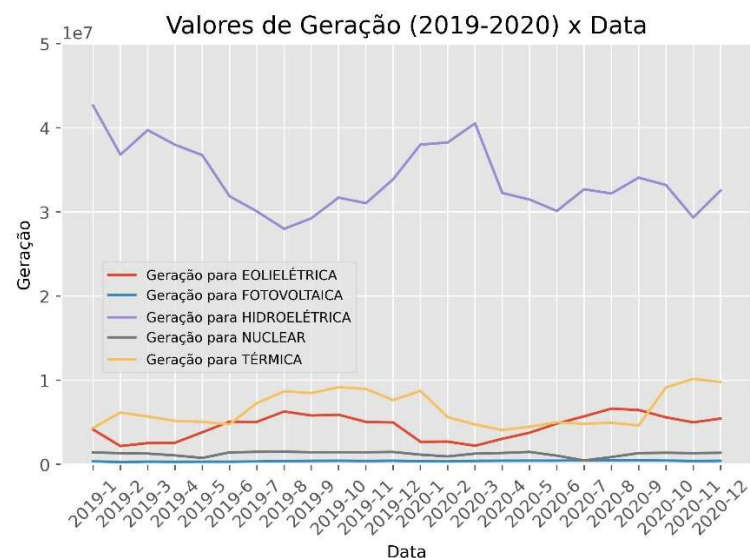


Figura 05 – Geração 2019-2020

Ao ser verificado a porcentagem produzida, presente na Figura 06 e Figura 07, houve grande semelhança nas produções Fotovoltaica, Eolielétrica e

Nuclear. A diferença que ocorreu foi elencada em Térmica e Hidroelétrica do previsto para o gerado. Essa diferença se justifica pelas repetidas crises hídricas no país, realizando-se assim o crescimento não-esperado da energia térmica (CNN Brasil, 2022).

Porcentagem prevista de produção energética no Brasil durante os anos de 2019 e 2020

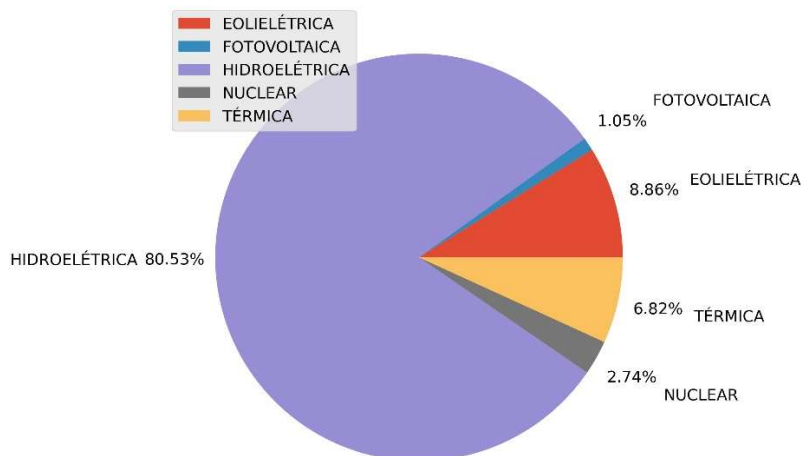


Figura 06 – Porcentagem prevista (2019 – 2020)

Porcentagem de produção energética no Brasil durante os anos de 2019 e 2020

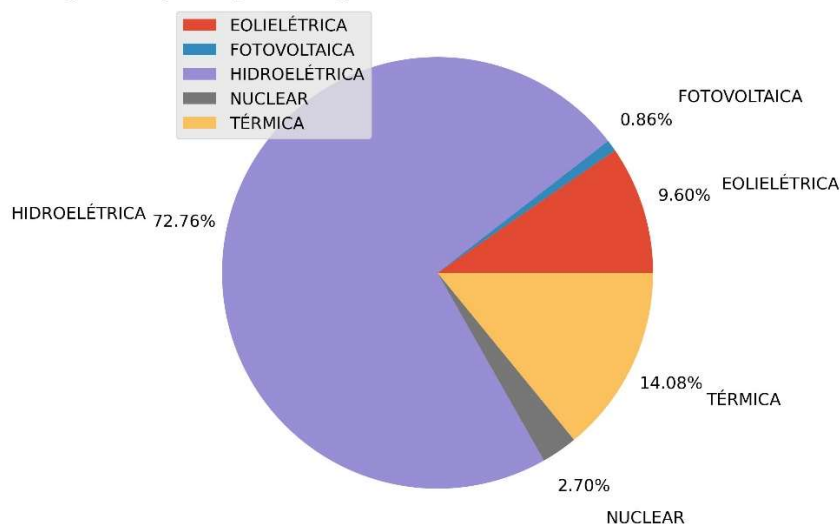


Figura 07 – Porcentagem gerada (2019 – 2020)

Realizado o levantamento por tipo de usina em forma numérica, obteve-se a comparação total do previsto em relação tanto ao valor gerado quanto ao valor previsto, apresentados na Figura 08. É visto que a diferença não se elevou muito em relação aos meios produzidos, se tornando de fato uma métrica de boa avaliação.

Por fim, ao realizar a avaliação mensal de 2019 e 2020, a Figura 09 descreve os principais parâmetros a serem analisados. Ao realizar o comparativo dos valores gerados e previstos, obteve-se uma semelhança até o mês 03 de 2020, onde começou a ocorrer uma diferença no eixo x superior, denotado Variação do Real, devido à Pandemia, que segundo a Empresa de Pesquisa Energética (EPE, 2022) houve uma diminuição do consumo energético e um crescimento na utilização de energia térmica.

Comparação entre Valor de Geração e Valor de Geração Prevista nos anos de 2019 e 2020

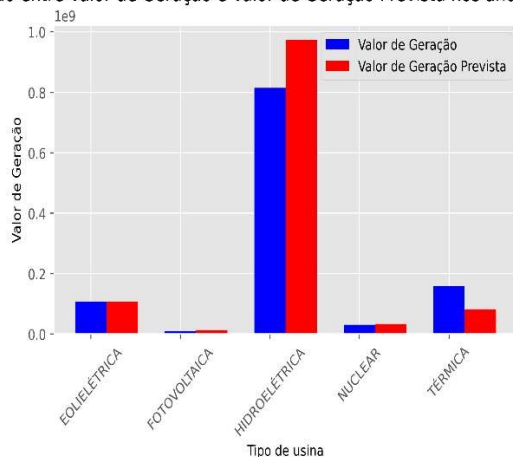


Figura 08 – Tipo de Usina x Valor de geração e geração prevista

Comparação do Somatório Ano-Mês Previsto/Gerado

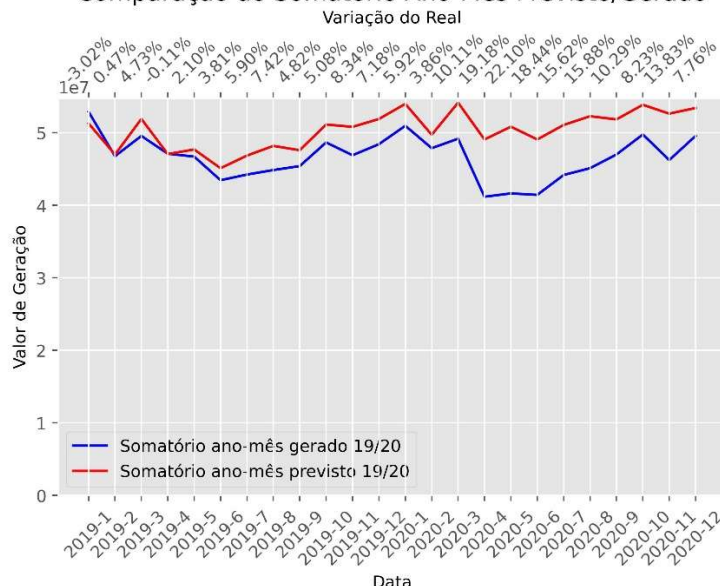


Figura 09 – Comparação do somatório mensal

#### 4. Conclusões e recomendações

De acordo com a análise de dados feita para todo o apunhalado disponibilizado, verificou-se como um modelo efetivo para a previsão, sendo a diferença em períodos normais de valores máximos 8,34% de variação do valor real em 2019. Em 2020, devido a fatores não sazonais, como a COVID, crise hídrica e demais problemas vinculados a estes, teve-se um erro não esperado, tendo que tomar medidas especiais para este fator, como o aumento da energia térmica. Vê-se que, no mês 12 de 2020, o gráfico já iria retornar valores estimados próximos aos valores reais.

Para recomendações futuras, é visto a necessidade de manter o teste para verificar possíveis divergências e, caso ocorram, explorar o motivo. Como a análise foi realizada para todos os tipos de usina em funcionamento, é importante explorar as que obtiverem maior divergência e verificar as razões diretas e indiretas na diferenciação do resultado esperado.

#### Referências

CNN Brasil. (2022). **Produção de energia por termelétricas cresce 77% em 2021, aponta estudo.** Disponível em: <https://www.cnnbrasil.com.br/nacional/producao-de-energia-por-termeltricas-cresce-77-em-2021-aponta-estudo>. Acesso em 11 de março de 2024.

Empresa de Pesquisa Energética (EPE, 2022). **Balanco Covid-19.** Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-500/Balanco Covid-19%20rev.pdf>. Acesso em 11 de março de 2024.