

Bootcamp de dados - Enacom 2024

Pedro Reis Cabreira

Descobertas da análise exploratória

- **Análises Gráficas:**

- Tendência de Aumento na Produção de Energia: Observou-se um aumento consistente na produção de energia no Brasil ao longo do tempo, indicando investimentos contínuos em usinas de energia.
- Concentração de Investimentos em Poucos Modelos: A análise revelou que os investimentos em energia estão concentrados em alguns modelos, com a hidroelétrica ainda sendo a principal matriz energética ao longo dos anos.
- Crescimento dos Modelos Eólicos e Fotovoltaicos: Apesar da predominância hidroelétrica, há uma tendência crescente de investimento em modelos eólicos e fotovoltaicos, especialmente a partir de 2014, quando os eólicos experimentaram um aumento significativo.
- Variação na Produção de Energia: Entre 2002 e 2013, houve pouca variação anual na produção de energia, indicando uma capacidade de produção quase constante. A partir de 2014, a variação anual aumentou, possivelmente devido ao crescimento de outras fontes de energia além das hidroelétricas.

- **Análises dos Campos:**

- Inconformidades na Coluna 'Ceg': Ao examinar a coluna 'Ceg' (Código Único do Empreendimento de Geração), notou-se uma grande quantidade de valores "-", indicando a falta de um código específico.
- Impacto dos Valores '-': Embora os valores "-" representem cerca de 15% dos dados na coluna 'Ceg', eles contribuem apenas com cerca de 5% da energia total gerada. Isso sugere que essas usinas, classificadas como pequenas ou conjuntos de usinas, provavelmente têm pouca capacidade de geração e, portanto, não possuem um código único registrado na ANEEL.

Descobertas da análise exploratória

- **Análises dos Campos - Estados das Usinas:**

- Distribuição das Usinas por Estado: Os cinco estados com o maior número de usinas elétricas são SP, MG, RS, BA e CE, representando aproximadamente 14%, 11%, 9.5%, 9% e 6.5% do total, respectivamente. Esses cinco estados totalizam 50% de todas as usinas do Brasil.
- Relação entre Estados com Mais Usinas e Produção de Energia: Exceto por MG e SP, os três estados seguintes com o maior número de usinas não estão entre os cinco principais produtores de energia do Brasil. Esses estados são PR, SP, MG, PA e I (Internacional), representando cerca de 17%, 14%, 12%, 8% e 8% da energia total produzida no país, respectivamente.
- Dependência Energética em Poucas Usinas: A análise revela uma grande dependência em poucas usinas de energia, como Itaipu (PR/Paraguai), Belo Monte e Tapajós (PA), destacando uma concentração significativa de produção em determinadas regiões.
- **Análises dos Campos - Tipos de Usina:**
- Predominância das Hidroelétricas: Cerca de 53% das usinas no Brasil são do tipo Hidroelétricas. Essa predominância reflete a dependência do país em relação a esse único modelo de produção de energia, o que pode ser preocupante devido à vulnerabilidade a condições climáticas, como secas, e aos impactos ambientais associados à sua implantação.

Modelos utilizados

- **XGBoost:** É um algoritmo de aprendizado de máquina poderoso e eficiente, usado para problemas de classificação e regressão. Ele se destaca pela alta precisão e velocidade, sendo capaz de lidar com conjuntos de dados grandes e complexos.
- **Prophet:** É uma ferramenta de previsão de séries temporais desenvolvida pelo Facebook, projetada para simplificar o processo de previsão. Ele é ideal para prever séries temporais com sazonalidades claras e tendências, e destaca-se por sua facilidade de uso e capacidade de lidar automaticamente com feriados e incertezas.
- **Modelo escolhido:** Nesse caso seria preferível utilizar o modelo Prophet, pois ele obteve melhores resultados de previsão para três dos cinco tipos de usinas elétricas, e ele se adequa melhor a séries temporais que possuem certa sazonalidade que é o caso das usinas hidroelétricas, fotovoltaicas e nucleares. Porém para obter melhores resultados também deveria se considerar a aplicação do XGBoost nas bases relativas às usinas eólicas e térmicas.

Métricas e avaliação do modelo

- **XGBoost:**

Tipo	MAE	MAPE
Hidroelétricas	2771803.21	8.13
Eolielétricas	727881.65	14.97
Fotovoltaicas	179722.01	44.62
Nuclear	255667.65	28.32
Térmica	2258372.00	36.01

- **Prophet:**

Tipo	MAE	MAPE
Hidroelétricas	2772078.00	7.65
Eolielétricas	998088.97	32.15
Fotovoltaicas	75121.44	18.87
Nuclear	181599.83	22.55
Térmica	3331610.00	63.97

- Para realizar a avaliação da precisão dos modelos, foram calculados o MAE (Erro Médio Absoluto) e o MAPE (Erro Percentual Absoluto Médio), por exemplo, o MAPE calcula a média das diferenças percentuais absolutas entre as previsões do modelo e os valores reais, o que fornece uma medida relativa de precisão, independentemente da escala dos valores da variável alvo, o que significa que quanto menor o MAPE mais preciso é o modelo.

- Um MAPE de 7.65% significa que, em média, as previsões estão incorretas em cerca de 7.65% em relação aos valores reais. Em outras palavras, as previsões tendem a ser, em média, 7.65% maiores ou menores do que os valores reais, o que pode ser considerado uma precisão razoável das previsões em relação aos valores reais.
- Foi possível perceber que dependendo da base de dados e suas características um algoritmo pode se sair melhor do que o outro. O FBProphet por exemplo pode ser útil para previsões em séries temporais com padrões sazonais mais claros, porém nem sempre é eficaz em dados com padrões mais imprevisíveis, como é o caso do gráfico de geração das usinas Eolielétricas e Térmicas.