

DESAFIO CIÊNCIA DE DADOS

GABRIEL RIPPER



Problema

O consumo energético residencial pode ser influenciado por diversos fatores, o que torna difícil a predição de consumo de forma intuitiva, mesmo com uma boa fundamentação de negócio. Por tanto, a proposta visa trazer para o negócio a implementação de modelos de machine learning com o objetivo de enriquecer as análises e fortalecer a tomada de decisão.



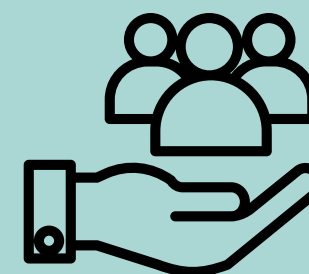
AQUARELA
Advanced Analytics

Objetivo

O objetivo, por tanto, é desenvolver um modelo de machine learning para executar previsões baseando-se em técnicas estatísticas avançadas e poder computacional.



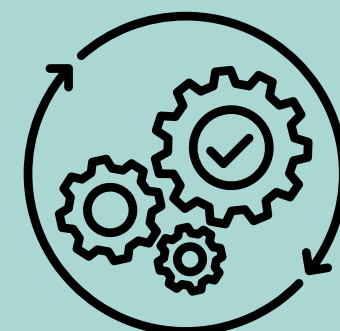
Base de dados



100 clientes



18k de pontos de dados



Dados históricos de consumo e clima



Métricas Finais

RMSE | 2,0

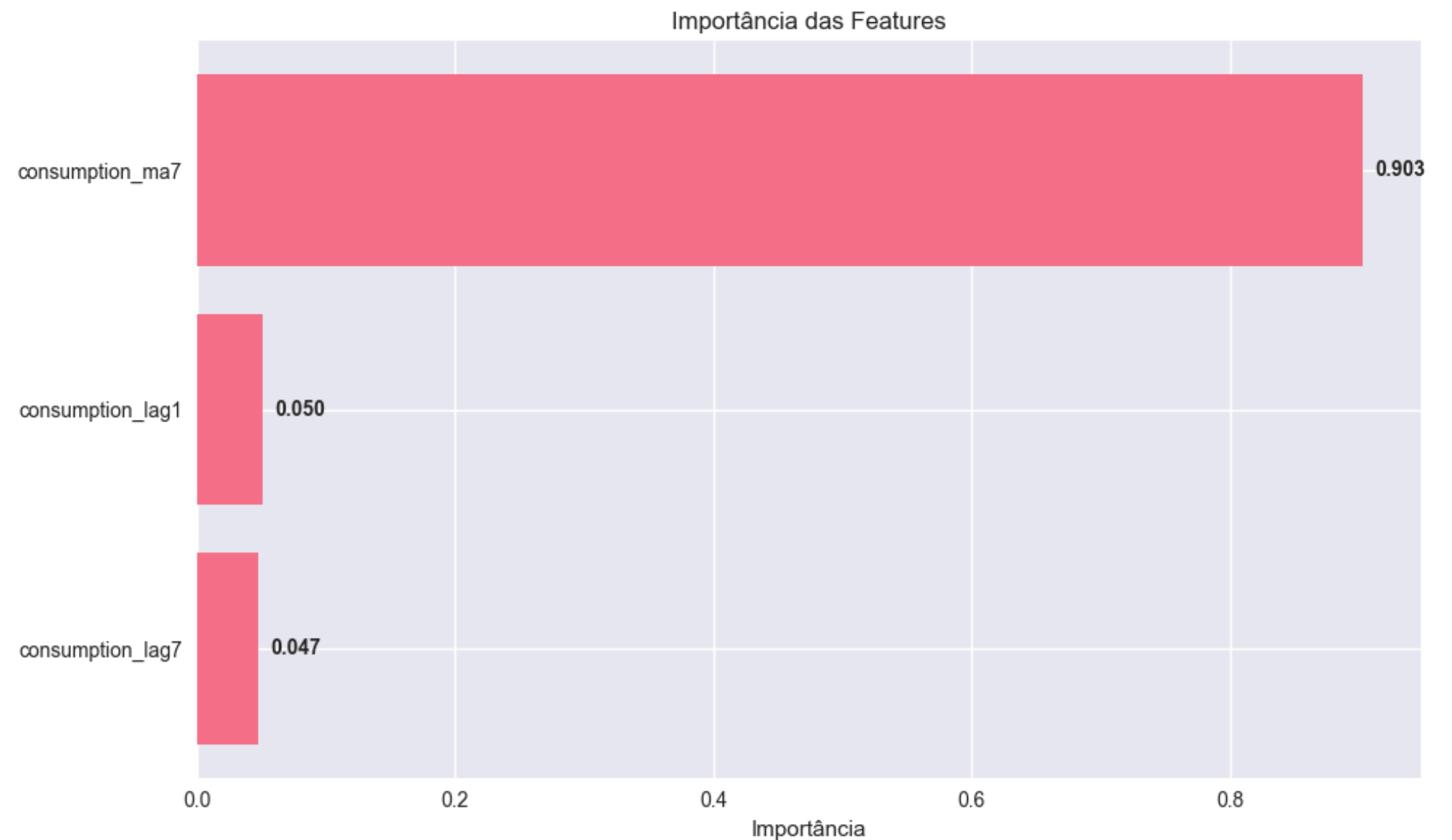
MAPE | 11.43%

R^2 | 0,70



Importância das variáveis

Durante o desenvolvimento do modelo, os dados climáticos apresentaram pouca influência na previsão dos valores de consumo, tal motivo é justificado pela baixa correlação entre essas variáveis e o alvo. Dessa forma, as variáveis que mais influenciaram nos resultados, foram: média móvel dos últimos 7 dias, o consumo do dia -1 e o consumo do dia -7.





Impacto

Os impactos quantitativos quanto ao impacto requerem um estudo mais profundo dos dados históricos. Quanto aos impactos qualitativos, espera-se que o modelo, que possui uma boa taxa de acerto, sirva como um novo ponto de embasamento para a tomada de decisão, não podendo ser utilizado como uma “bola de cristal” mas sim como uma “segunda opinião”, além daquela fundamentada pelo especialista.





Próximos Passos

Os modelos testados possuíam bons comportamentos iniciais, mais é necessário investigar mais profundamente a relação entre as variáveis climáticas e o consumo, uma vez que intuitivamente elas deveriam fazer mais sentido. Uma abordagem interessante poderia ser decompor a série temporal do consumo, já que ela possui um comportamento normal, para então criar modelos individuais para a tendência, a sazonalidade e os resíduos, utilizando variáveis exógenas, como o clima, afim de criar uma previsão mais robusta.



Conclusão

Os modelos desenvolvidos podem ser utilizados como prova de conceito para o cumprimento do objetivo, entretanto, não se pode dispensar a necessidade de um esforço mais elevado para a elaboração de novas variáveis preditivas e modelos mais complexos, como aqueles de rede neural.

OBRIGADO!

GABRIEL RIPPER