



# Aquarela Energy Forecast

## Previsão de Consumo Energético

Solução de Ciência de Dados para consumidores residenciais



# Contexto do Problema

## Desafio

Prever o consumo diário de energia residencial a partir de:

- Historico de consumo
- Variáveis climáticas
- Perfil do cliente

## Complexidades

- Forte sazonalidade
- Relações não lineares com variáveis climáticas
- Heterogeneidade no padrão de consumo entre clientes
- Possível autocorrelação temporal

# Dados Utilizados



## Consumo (Target)

- Série temporal diária de kWh por cliente
- Histórico multi-período
- Base para modelagem supervisionada



## Clima (Variáveis Exógenas)

- Temperatura (mín., máx., média)
- Umidade e precipitação
- Índice UV
- Indicadores de dias úteis e feriados



## Perfil do Cliente (Features Estruturais)

- Tipo de residência
- Características demográficas
- Histórico agregado de consumo
- Possível segmentação comportamental

Dataset unificado combinando séries temporais e variáveis estáticas para previsão diária de consumo energético.



# Tratamento de Dados e Feature Engineering

01

## Limpeza e Padronização

- Tratamento de valores ausentes por interpolação temporal
- Remoção de outliers via análise estatística
- Alinhamento temporal entre consumo e clima
- Padronização de granularidade diária

02

## Engenharia de Features

Variáveis temporais (dia da semana, mês, sazonalidade)

- Criação de lags ( $t-1, t-7, t-30$ )
- Rolling mean e rolling std
- Interações entre clima e consumo
- Indicadores de feriados

03

## Transformações

- Encoding de variáveis categóricas
- Normalização para modelos sensíveis à escala
- Transformações log para reduzir assimetria
- Split temporal (train/test respeitando ordem cronológica)

# Modelagem e Validação

## Estratégia de Modelagem

- Modelos baseados em árvores (XGBoost, LightGBM) para capturar não linearidades
- LSTM para modelagem sequencial e dependência temporal
- Comparação entre modelos baseados em features vs modelos sequenciais
- Seleção baseada em performance e estabilidade temporal



# Resultados de Performance



RMSE – Linear Regression

Modelo baseline



RMSE – Random Forest

Modelo não linear



MAE Comparison

Desempenho Similar

Modelos lineares apresentaram desempenho equivalente ou superior aos modelos não lineares, sugerindo predominância de efeitos autoregressivos na série.

# Arquitetura do Pipeline



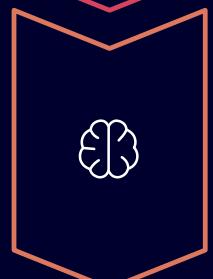
## Ingestão

Extração, de multiplas fontes ETL, versionamento



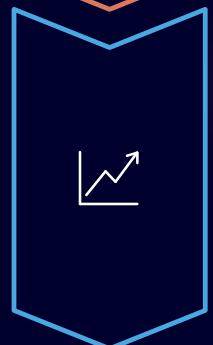
## Transformação

Alinhamento temporal, criação de lags, encoding e geração de dataset



## Treinamento

Sprint temporal (train e test), comparacao de modelos, avaliacao RMSE e MAE



## Inferência

Geração de previsões t+1, possibilidade de re-treinamento periódico, pipeline preparado para novos dados, estrutura compatível com produção (MLOps-ready)





# Dashboard Interativo

Monitoramento da performance do modelo e análise de padrões de consumo

- Comparação entre consumo real e previsto ao longo do tempo
- Distribuição do erro absoluto (MAE) por cliente
- Análise de coeficientes do modelo linear
- Identificação de padrões sazonais e efeitos climáticos

# Principais Insights

## Influência da Temperatura

- Consumo aumenta em períodos de calor e frio extremos
- Evidência de relação positiva entre temperatura e demanda
- Impacto mais evidente em clientes com maior consumo médio

## Sazonalidade Semanal

- Diferença significativa entre dias úteis e finais de semana
- Variação aproximada de 20–25% no consumo médio diário
- Forte componente autoregressivo identificado

## Heterogeneidade de Consumo

- Padrões distintos entre clientes
- Clientes de alto consumo apresentam maior variabilidade
- Indicação de potencial ganho com segmentação futura

- O consumo residencial apresenta forte componente temporal e influência climática, com estrutura predominantemente linear.



# Conclusão e Próximos Passos

## Conclusão

Estruturação de pipeline completo para previsão diária de consumo residencial, com validação temporal adequada e modelos comparativos.

Resultados indicam forte componente autoregressivo e estrutura predominantemente linear do problema.

Projeto foi estruturado com foco em interpretabilidade, robustez temporal e escalabilidade futura

## Próximos Passos

1. Implementar re-treinamento automático periódico
2. Avaliar modelos por segmento de clientes
3. Incorporar novas variáveis externas (ex.: tarifas, eventos)
4. Estruturar API para consumo em tempo real