# Defesa Técnica do Algoritmo

O objetivo do estudo foi desenvolver um modelo preditivo supervisionado capaz de estimar as vendas semanais (Weekly\_Sales) com base em variáveis socioeconômicas e ambientais da base sales.csv.

Foram testados os algoritmos Regressão Linear, Random Forest e Gradient Boosting. O modelo Regressão Linear apresentou o melhor desempenho, com maior Test\_R² e menores Test\_MAE e Test\_RMSE, sendo, portanto, o modelo selecionado para esta aplicação.

## 1. Fundamentação da Escolha

A escolha da Regressão Linear baseia-se em sua adequação ao comportamento dos dados e em seu desempenho estatístico superior.

A análise exploratória indicou relações predominantemente proporcionais e estáveis entre as variáveis independentes e a variável-alvo, o que favorece modelos lineares.

A inclusão de atributos sazonais (Week\_sin, Week\_cos, Month, WeekOfYear) e macroeconômicos (CPI, Unemployment) permitiu linearizar variações cíclicas e contextuais, tornando o modelo linear suficientemente expressivo.

Adicionalmente, o baixo desvio entre as métricas de treino, validação cruzada e teste indica boa generalização e ausência de sobreajuste (overfitting).

## 2. Interpretação das Variáveis e Relevância

A análise dos coeficientes padronizados evidenciou as variáveis de maior impacto no comportamento das vendas:

* Week\_sin, Week\_cos, Month e WeekOfYear: capturam a sazonalidade e os ciclos de consumo.
* Holiday\_Flag: representa variações de demanda em períodos festivos.
* CPI e Unemployment: refletem o poder de compra e as condições macroeconômicas.
* Temperature: apresenta correlação moderada, indicando influência parcial em produtos sazonais.

Variáveis como Fuel\_Price e Quarter tiveram menor contribuição marginal, mantendo, contudo, importância contextual no ajuste do modelo.

## 3. Avaliação de Desempenho

O modelo linear obteve o maior Test\_R² e baixos valores de MAE e RMSE, comprovando capacidade de explicar a variabilidade das vendas e gerar previsões precisas, em relação aos demais estudados.

A comparação com Random Forest e Gradient Boosting evidenciou que a complexidade adicional desses métodos não resultou em ganho preditivo relevante, demonstrando que as relações lineares predominam na base de dados.

A consistência entre a validação cruzada e o desempenho em teste confirma a estabilidade e robustez do modelo.

## 4. Limitações e Oportunidades de Melhoria

Apesar do bom desempenho, a assertividade do modelo pode ser ampliada com a inclusão de novas variáveis explicativas, como:

* Indicadores de promoções e campanhas comerciais;
* Dados regionais e demográficos (localização, renda média, densidade populacional);
* Eventos externos (feriados prolongados, greves, condições climáticas extremas);
* Informações sobre concorrência e variação de preços.

Essas variáveis aumentariam a granularidade do modelo, reduzindo erros residuais e melhorando a sensibilidade às flutuações do mercado.

## 5. Conclusão Técnica

A Regressão Linear demonstrou ser o modelo mais adequado para o problema de previsão de vendas semanais, apresentando melhor desempenho preditivo, baixo erro médio e consistência entre treino e teste.

Sua estrutura simples e estatisticamente sólida permite identificar de forma clara a influência de cada variável sobre o resultado, garantindo transparência, interpretabilidade e confiabilidade na tomada de decisão.

Os resultados obtidos indicam que as relações existentes entre as variáveis independentes e a variável-alvo são predominantemente lineares, o que torna o modelo eficiente e tecnicamente coerente com o comportamento dos dados.

Dessa forma, a Regressão Linear atende plenamente aos critérios de precisão, estabilidade e aplicabilidade prática, configurando-se como a solução mais adequada dentro do escopo do estudo.