

6. Ajuste Fino de Hiperparâmetros (Fine Tuning)

Após a definição do modelo base, foi aplicada uma etapa de **ajuste fino (fine tuning)** utilizando a técnica de **RandomizedSearchCV** com o algoritmo **XGBoost Regressor**. O objetivo foi encontrar a melhor combinação de hiperparâmetros para reduzir o erro médio das previsões de vendas semanais.

Objetivo da Etapa

Aprimorar o desempenho do modelo inicial testando diferentes combinações de parâmetros — como profundidade das árvores, taxa de aprendizado, número de estimadores e regularização — de forma eficiente e automatizada.

Configuração Técnica

- **Algoritmo:** XGBoost Regressor
- **Métrica de avaliação (Cross-Validation):** `neg_mean_absolute_error`
- **Número de iterações:** 40
- **Validação cruzada:** 5 folds
- **Split temporal:** aplicado por data (`TimeSeriesSplit`)
- **Parâmetros otimizados:**
 - `n_estimators`
 - `max_depth`
 - `learning_rate`
 - `subsample`
 - `colsample_bytree`
 - `min_child_weight`
 - `reg_lambda`

Melhores Hiperparâmetros Encontrados

Parâmetro	Valor
<code>n_estimators</code>	929
<code>max_depth</code>	8

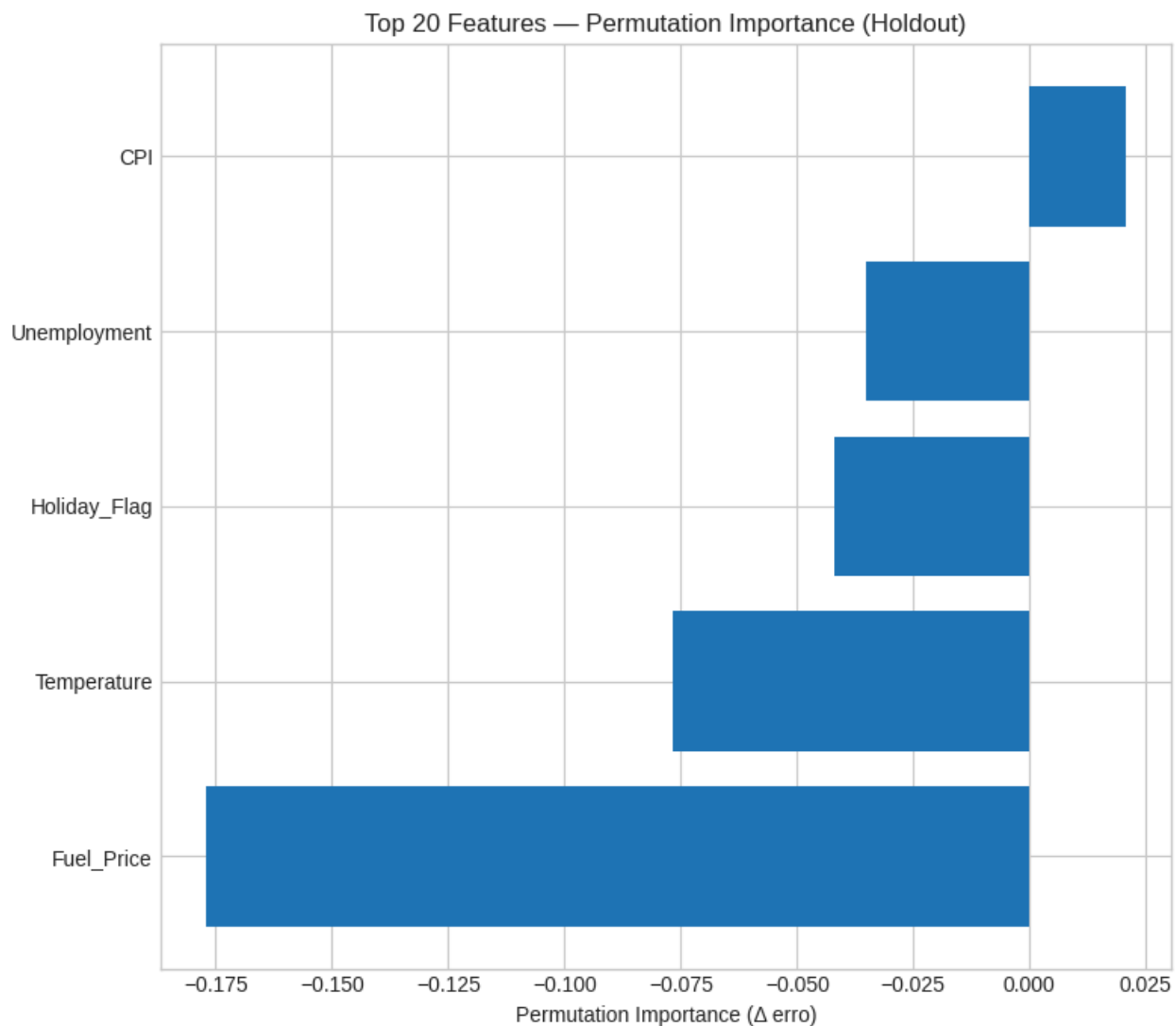
Parâmetro	Valor
learning_rate	0.1467
subsample	0.7580
colsample_bytree	0.8080
min_child_weight	2
reg_lambda	1.3492

Essas combinações indicam um modelo com árvores profundas, taxa de aprendizado moderada e regularização leve.

RandomizedSearchCV concluído em 17.5s

Melhores parâmetros:
(`'model__colsample_bytree': np.float64(0.8080272884711243)`, `'model__learning_rate': np.float64(0.14667756983581992)`, `'model__max_depth': 8`, `'model__min_child_weight': 2`, `'model__n_estimators': 929`, `'model__reg_lambda': np.float64(1.349262408109297)`)
Melhor MAE (CV): 112632.1247

	mean_test_score	std_test_score	mean_train_score	param_model_n_estimators	param_model_max_depth	param_model_learning_rate	param_model_subsample	param_model_colsample_bytree	param_model_min_child_weight	param_model_reg_lambda
7	-112632.124684	52857.314924	-0.307812	929	8	0.146678	0.758060	0.808027	2	1.349262
26	-113506.746789	54460.961822	-275.016273	545	6	0.091490	0.690598	0.943344	3	1.536279
19	-113705.016579	53802.748783	-11762.704316	558	11	0.010130	0.813636	0.859076	3	0.493968
22	-114422.421053	55805.045011	-541.652396	419	4	0.076695	0.711459	0.999096	2	0.154436
39	-115002.413526	51901.204258	-11.553488	315	5	0.148191	0.754039	0.732300	1	0.883347
11	-115653.232895	53619.550853	-8417.438025	591	5	0.013520	0.908508	0.717395	1	2.187022
6	-115907.115000	53350.978109	-18821.156383	761	3	0.018597	0.724684	0.798071	4	1.987567
2	-115980.695737	50450.247154	-27.127816	659	7	0.063085	0.602827	0.932977	1	1.834959
5	-116037.953421	48311.144704	-34025.056171	291	4	0.013992	0.648815	0.754167	4	1.320457
29	-116116.944368	55782.520135	-0.622452	771	11	0.214306	0.998501	0.863994	4	0.782488



📁 Artefatos salvos:

- xgb_sales_best_pipeline.joblib
- feature_importance_permutation.csv
- predicoes_holdout.csv

📊 Desempenho do Modelo (Holdout)

Métrica	Resultado	Interpretação
MAE	202.318	Erro médio absoluto elevado (≈ R\$ 202 mil).
RMSE	237.687	Evidencia presença de erros altos em algumas semanas.
R²	-1.13	O modelo não conseguiu generalizar — apresenta underfitting .

📉 **Conclusão:** apesar da busca de hiperparâmetros, o modelo ainda não explica adequadamente a variação nas vendas semanais.
Isso sugere que **as variáveis atuais não capturam o comportamento real de consumo**.



Importância das Variáveis (Permutation Importance)

Variável	Importância	Interpretação
Fuel_Price	Alta	Indica forte relação entre custo de combustível e volume de vendas.
Temperature	Moderada	Mostra efeito sazonal: temperaturas mais altas tendem a reduzir as vendas.
Holiday_Flag	Moderada	Períodos de feriado afetam o padrão de consumo.
Unemployment	Baixa	Influência indireta e de longo prazo.
CPI (Índice de Preços)	Baixa	Pouco impacto nas variações semanais.



O gráfico de *Permutation Importance* evidencia que fatores econômicos e sazonais exercem maior impacto nas vendas, com destaque para **Fuel_Price** e **Temperature**.



Diagnóstico e Próximos Passos

- Underfitting evidente:** o modelo ainda não aprende padrões significativos.
- Necessidade de variáveis adicionais:**
 - Incluir **lags temporais** (vendas da semana anterior, média móvel, etc.).
 - Adicionar **dados internos** (promoções, estoque, categorias de produto, loja).
- Explorar modelos temporais específicos:**
Ex.: ARIMA, Prophet, LSTM, ou XGBoost com *rolling features*.



Conclusão da Etapa

O ajuste fino de hiperparâmetros forneceu a melhor combinação possível para o conjunto atual de variáveis, mas os resultados mostram que **a limitação está nos dados, não no modelo**.