Sales Forecasting Model: Execution Report

Date: September 12, 2025

Script: predictive_analytics.py

Executive Summary

This report documents the execution of the advanced sales forecasting model. The script successfully completed all four stages: data generation, model training with hyperparameter tuning, model evaluation, and future sales prediction. The final XGBoost model demonstrates strong predictive power, explaining 78% of the sales variance with an average error of approximately $19.58. A 7-day sales forecast was successfully generated.

Step 1: Data Generation & Feature Engineering

The process began by creating a sample dataset and engineering advanced features to capture seasonality, trends, and time-based dependencies.

Data Head with Advanced Features:

Date	Temperature	Local_Event	Is_Holiday	Month	Day_of_Week	Week_of_Year	Month_sin	Month_cos	Day_of_Week_sin	Day_of_Week_cos	Sales	Sales_Lag_1
2023-01-02	16	0	0	1	0	1	0.500	0.866	0.000	1.000	160	94.0
2023-01-03	15	0	0	1	1	1	0.500	0.866	0.782	0.623	94	85.0
2023-01-04	20	0	0	1	2	1	0.500	0.866	0.975	-0.223	85	86.0
2023-01-05	31	0	0	1	3	1	0.500	0.866	0.434	-0.901	86	113.0
2023-01-06	15	0	0	1	4	1	0.500	0.866	-0.434	-0.901	113	52.0

Step 2: Model Training & Hyperparameter Tuning

An XGBoost regression model was trained on the dataset. GridSearchCV was employed to systematically test multiple hyperparameter combinations to find the optimal model configuration.

Tuning Process Summary:

- **Method:** 3-fold cross-validation
- **Candidates Tested:** 24 parameter combinations
- **Total Fits:** 72 model training runs

Best Parameters Found: The search identified the following optimal settings for the model:

- learning_rate: **0.1**
- max_depth: **3**
- n_estimators: **100**
- subsample: **0.8**

Step 3: Model Evaluation & Interpretation

The best-performing model was evaluated on a held-out test dataset to assess its real-world accuracy.

Model Performance Metrics:

- **Mean Absolute Error (MAE): $19.58**
    - *On average, the model's sales predictions are off by approximately $19.58.*
- **R-squared (R^2): 0.78**
    - *The model successfully explains 78% of the variability in sales data, indicating a strong fit.*

A feature importance plot was also generated to provide insights into the key drivers of sales.

Step 4: Generating a Sales Forecast

Using the trained model, a forecast was generated for the upcoming 7-day period.

Note on Execution Warning: A FutureWarning was observed during this step. This is an informational message from the pandas library regarding upcoming changes and **does not affect the validity of the current forecast.**

Advanced Sales Forecast for Next Week:

Predicted_Sales	Local_Event	Temperature	Day_of_Week_sin
78	0	15	0.000
84	0	16	0.782
72	0	14	0.975
91	0	18	0.434
162	1	20	-0.434

253	1	22	-0.975
147	0	21	-0.782

End of Report.