

Product Sales Prediction for E-Commerce – Final Project Report

1. Problem Overview

E-commerce businesses must accurately forecast product sales to optimize inventory, reduce stockouts and overstocking, and improve revenue planning. Manual forecasting methods struggle to scale and adapt to changing demand patterns. This project aims to build a machine learning model that predicts product sales using historical e-commerce data to support data-driven decision-making.

2. Approach

The project followed a structured data science workflow. Exploratory Data Analysis (EDA) was performed to understand sales distributions, feature relationships, and data quality. Sales volumes showed a near-normal distribution with mild right skew, while price demonstrated a negative relationship with demand. Data preprocessing included handling missing values, encoding categorical variables, and feature scaling. Several regression models were evaluated, including a baseline Dummy Regressor, Linear Regression, and Random Forest. Models were compared using cross-validated MAE, RMSE, and R² to ensure robust generalization.

3. Findings

Linear Regression was selected as the final model due to its strong cross-validated performance, low variance, and interpretability. Although Random Forest achieved slightly better test performance, the difference was negligible. Key findings include the importance of pricing and historical sales features in predicting demand, and evidence that simpler models can perform competitively when data relationships are largely linear.

4. Recommendations

1. Use the sales prediction model to inform inventory planning and reduce stockouts.
2. Incorporate pricing insights from the model to support dynamic pricing strategies.
3. Integrate the model into internal dashboards to enable data-driven forecasting decisions.

5. Future Research

Future work could improve performance by incorporating time-series forecasting methods, seasonality and holiday effects, promotional data, and external economic indicators. Deploying the model as an automated forecasting service could further enhance business value.