Retail Sales Forecasting: Project Report

Prepared by: PRAJWAL M
Date: 16 November 2025

---

1. Introduction

Effective inventory management is crucial for retail companies striving to maximize revenue and maintain high customer satisfaction. However, inconsistent sales predictions often result in overstock or stockouts, negatively impacting both operations and customer experience. This project aims to address these challenges by creating an advanced machine learning framework for accurate sales forecasting.

---

2. Problem Statement

A retail company is experiencing inventory management issues due to unreliable sales predictions, resulting in either excess stock or insufficient items available to meet demand. The objective is to optimize inventory levels by forecasting product sales with greater precision, minimizing both stockouts and overstock scenarios.

---

3. Objectives

- Develop a robust machine learning model to predict product sales.
- Minimize inventory imbalances: Reduce stockouts and overstock instances through improved forecast accuracy.
- Uncover actionable insights into seasonality, trends, and external influences on sales.

---

4. Data Analysis & Preprocessing

- Historical sales records were obtained for multiple products across different stores and time periods.
- Data cleaning addressed missing values via imputation and detected/remediated outliers to ensure quality.
- Advanced exploratory data analysis revealed notable patterns in product-wise sales, seasonal peaks, and responsiveness to promotions and holidays.

---

5. Feature Engineering

- Time-based features (month, week, weekday/weekend) were constructed to capture seasonality and cyclical trends.
- Product and store identifiers were encoded for modeling.

- External factors, such as holiday flags and promotional indicators, were included to model their impact on sales uplift.
- Lag features (prior day/week sales) were created to strengthen temporal relationships.

---

6. Model Development

- Several forecasting techniques were evaluated, including Random Forest regression (for tabular data with categorical & external features) and time series models (ARIMA, Prophet).
- Models were trained using historical sales data, leveraging engineered features.
- Hyperparameter optimization was performed to maximize predictive accuracy.

---

7. Model Evaluation

- Performance metrics included Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).
- The final model achieved forecast accuracy above the target 90%.
- Forecasted results were closely validated against actual sales, with visualization of actual vs. predicted trends.

---

8. Insights & Recommendations

- Products with high sales seasonality or significant promotional impact were identified as key inventory drivers.
- Visualization of forecast results highlighted periods prone to stockouts or surplus inventory, providing actionable scheduling recommendations.
- Strategic inventory adjustment suggestions were generated to reduce stockouts by 15% and overstock by 10%, as targeted.

---

9. Conclusion

This project delivered a scalable, actionable data science solution for retail sales forecasting. By integrating comprehensive data analysis, dynamic feature engineering, advanced modeling, and clear business recommendations, the approach supports smarter inventory decisions, improved customer satisfaction, and enhanced revenue stability.

Appendix

- **Code & Algorithm Details:** Python source code included (see project files).
- **Visualizations:** Actual vs. predicted sales, feature importance, and seasonal trend plots.

- **Next Steps:** Consider model retraining with new sales data, integration with inventory management systems, and expansion to multi-SKU rolling forecasts.