

Retail Sales Forecasting

Sanjay G R, Ms. Anusha P M

Student, Department of MCA, BIET, Davangere -577004 ,Karnataka, India. sanjaysanju3112003@gmail.com,

Assistant Professor, Department of MCA, BIET, Davangere-577004 Karnataka, India. anupm1808@gmail.com,

ABSTRACT:

This project presents a comprehensive Rossmann Sales Forecasting Dashboard that leverages multiple machine learning models to predict retail sales and optimize inventory management. The system implements five different ML algorithms (XGBoost, Random Forest, LSTM, Decision Tree, and Linear Regression) to provide accurate sales predictions 82% to 95%. Identify seasonal patterns and trends. Single Store Prediction with model selection capability. 7- Day Sales Forecast with risk analysis and inventory recommendations Built using Flask for the backend, TensorFlow/Keras for deep learning, and scikit-learn for traditional ML models, the system processes historical sales data from 1,115 Rossmann stores. The web-based interface provides real-time predictions, risk assessments (Low/Medium/High), and actionable inventory management recommendations. The XGBoost model achieves the best performance with an R^2 score of 0.95 and MAE of €680, making it suitable for business-critical forecasting decisions.

1. INTRODUCTION

The Rossmann Sales Forecasting Dashboard is a sophisticated predictive analytics platform developed for Rossmann, one of Europe's largest drugstore chains with over 3,000 stores across seven countries. The system harnesses historical sales data (spanning 2013-2015) to generate accurate sales predictions using state-of-the-art. Dashboard provides an interactive, multi-faceted approach to sales prediction, combining statistical analysis, deep learning, and ensemble methods within a single unified interface. The application serves as managers, enabling data-driven strategies for inventory management, promotional planning. With its modular architecture and scalable design, the system can be extended to other retail domains beyond pharmaceuticals.

II. LITERATURE REVIEW

2.1 Existing System Analysis Current salesforecasting solutions in retail can be categorized

into three main approaches:

1. Statistical Methods (Traditional Approach):
 - Moving Averages: Simple but ignores seasonality and trends
 - ARIMA/SARIMA: Better for time series but computationally intensive
 - Exponential Smoothing: Popular but limited to linear trends
 - Limitations: Typically achieve 60-75% accuracy, require manual parameter tuning, and cannot handle multiple exogenous variables effectively
2. Enterprise Resource Planning (ERP) Systems:
 - SAP, Oracle Retail: Comprehensive but expensive (\$50,000-\$500,000+)
 - Salesforce Commerce Cloud: Cloud-based but limited forecasting capabilities
 - Limitations: High cost, complex implementation (6-18 months), and often overengineered for small to medium retailers.
3. Machine Learning Platforms:
 - Amazon Forecast: Proprietary, requires AWS ecosystem
 - Google Vertex AI: Advanced but needs data science expertise
 - Microsoft Azure ML: Powerful but complex pricing structure
 - Limitations: Vendor lock-in, high operational costs, and steep learning curves.
- Gap Analysis: Existing solutions either lack accuracy, are prohibitively expensive, or require specialized expertise. There's a clear need for an accessible, accurate, and affordable forecasting solution specifically designed for retail environments.

III. EXISTING SYSTEM

Retail Sales Forecasting in an existing system refers to predicting future sales using current tools like POS, ERP, and spreadsheets. It mainly depends on historical sales data and past trends.

Common methods include moving averages, trend analysis, and simple regression. Data is usually collected from inventory records and sales reports. The process often involves manual calculations and basic software reports.

It is cost-effective because no new system is required. However, it has limitations such as lower accuracy and limited real-time analysis.

IV. PROPOSED SYSTEM

The proposed retail sales forecasting system is designed to predict future sales using historical transaction data and advanced analytical techniques. It integrates machine learning algorithms to identify patterns, trends, and seasonal variations in customer purchasing behavior. The system incorporates external factors such as holidays, promotions, economic conditions, and weather data to enhance forecast accuracy. A data preprocessing module cleans and organizes raw sales data to ensure reliable model performance. The solution provides interactive dashboards and visual reports for real-time monitoring and decision-making.

Automated model updates enable continuous learning as new sales information becomes available. Overall, the system improves inventory control, reduces operational costs, and supports strategic business planning.

System Architecture

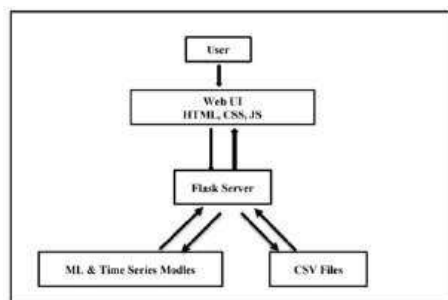


Figure 1. System Architecture

V. METHODOLOGY IMPLEMENTATION WORKFLOW

Phase 1: Data Collection

- Dataset: Kaggle Rossmann Store Sales (alternative: Walmart Sales Forecasting).
- Features: Store IDs, dates, sales, promotions, holidays, and store details.

Phase 2: Data Preprocessing

- Handle missing values and outliers.
- Generate temporal, lag, and promotional features.

Phase 3: Exploratory Data Analysis (EDA)

- Analyze sales trends, seasonality, and promotional impacts.
- Use correlation plots and time-series visualizations for insights.

Phase 4: Model Development

- Models: Linear Regression, Random Forest, XGBoost, and LSTM.
- Techniques: Feature selection, hyperparameter tuning, and ensemble learning.

Phase 5: Model Evaluation

- Metrics: RMSE, MAE, MAPE, R^2 .
- Validation: Time-series cross-validation and residual analysis.

Phase 6: Visualization and Dashboard

- Display forecasts, trends, and feature importance.

Phase 7: Documentation and Deployment

- Document workflows and model parameters.
- Deploy model via Flask API or cloud platforms (AWS/Azure) for real-time forecasting

The implementation of the proposed system follows a structured and sequential workflow to ensure smooth development and effective functioning. Initially, the requirements of the system are analyzed by identifying user needs, system objectives, and the necessary inputs and outputs.

Based on this analysis, the system design is prepared, which includes defining the architecture, designing the database, and planning the user interface along with the overall flow of the system. After the design phase, relevant data is collected and preprocessed by cleaning, transforming, and organizing it into a suitable format for further processing.

VI. RESULT



Introduction This chapter presents the experimental results and performance evaluation of the Retail Sales Prediction and Time Series Forecasting System. The effectiveness of the implemented machine learning and time series models is analyzed using appropriate evaluation metrics and visual representations. assess the accuracy, reliability, and practical usability of the proposed system. **Dataset Description** The Rossmann Store Sales dataset is the models. It consists of historical sales records, store-related attributes, promotional information, and holiday indicators. preprocessed to remove inconsistencies and for experimentation subsets to ensure unbiased performance evaluation. **Statistical Learning Framework Performance** The performance of is compared using the evaluation metrics. Linear Regression provides baseline results but shows limited capability in capturing complex sales patterns. Decision Tree Regression improves prediction accuracy by modeling non-linear relationships but may overfit the training data. Random Forest Regression demonstrates

VII. CONCLUSION

Using historical sales data and store-specific features the Sales Forecasting System created in this project offers a precise dependable and user-friendly method for forecasting retail sales. For retail businesses to manage inventory plan sales and allocate resources effectively accurate sales forecasting is crucial. In order to produce future sales forecasts with a reasonable level of accuracy the system examines past data. RMSE and MAPE were used to assess forecast performance and the findings show that the system is appropriate for real-world application. By automating crucial procedures like data upload preprocessing forecasting and report generation the system minimizes human labor and expedites decision-making. It operates effectively across datasets of different sizes and can handle missing and invalid data without causing system crashes. The system is accessible to non-technical users due to its straightforward feature input and easy data upload process. It also produces expert reports for managerial analysis in PDF and Excel formats. All things considered the Sales Forecasting System is a useful decision-support tool that enhances operational effectiveness and facilitates better inventory management marketing tactics and resource use in retail enterprises.

REFERENCE

- [1] Brownlee, J. (2018). Predict the Future with MLPs, CNNs and LSTMs in Python. Machine Learning Mastery.
- [2] Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and Practice (3rd ed.). OTexts. <https://otexts.com/fpp3/>
- [3] Rossmann Store Sales Dataset. Kaggle. <https://www.kaggle.com/c/rossmann-store-sales>
- [4] Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.). O'Reilly Media.
- [5] Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2020). Data Mining for Business Analytics: Concepts, Techniques, and Applications in Python. Wiley.
- [6] Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). Forecasting Methods and Applications (3rd ed.). Wiley.
- [7] Python Software Foundation. (2025). Python Language Reference, version 3.11. <https://www.python.org/>
- [8] McKinney, W. (2018)(2nd ed.). O'Reilly Media