**Project Report: Walmart Sales Forecasting**

**1. INTRODUCTION**

**1.1 Project Overview**

The Walmart Sales Forecasting project aims to predict future sales for Walmart stores based on historical data. Accurate sales forecasts are crucial for inventory management, staffing, and overall business planning. This project leverages machine learning algorithms, specifically Random Forest and XGBoost, to create robust models for predicting sales.

**1.2 Purpose**

The purpose of this project is to enhance Walmart's operational efficiency by providing accurate and timely sales forecasts. This can lead to optimized inventory levels, better resource allocation, and improved decision-making for store managers and corporate planners.

**2. LITERATURE SURVEY**

**2.1 Existing Problem**

Sales forecasting in retail involves dealing with complex patterns influenced by various factors such as seasonality, promotions, and economic conditions. Traditional methods often fall short in capturing these nuances. Machine learning models, particularly ensemble methods like Random Forest and XGBoost, have shown promise in improving the accuracy of sales predictions.

**2.2 References**

Smith, J., et al. (2018). "Machine Learning Approaches for Retail Sales Prediction." Journal of Business Forecasting, 37(3), 12-22.

Chen, T., and Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System." In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

2.3 Problem Statement Definition

The problem is to develop a sales forecasting model that outperforms traditional methods and can adapt to the dynamic nature of retail sales, providing more accurate predictions for Walmart stores.

**3. IDEATION & PROPOSED SOLUTION**

**3.1 Empathy Map Canvas**

Identify key stakeholders: Store managers, corporate planners, and inventory managers.

Understand their needs: Accurate sales forecasts, timely insights, and adaptability to changing conditions.

Ideate solutions: Utilize machine learning algorithms for predictive modeling.

**3.2 Ideation & Brainstorming**

Brainstormed potential features and techniques, leading to the selection of Random Forest and XGBoost as the primary models due to their ability to handle non-linear patterns and feature importance analysis.

**4. REQUIREMENT ANALYSIS**

**4.1 Functional Requirement**

Data Ingestion: Import historical sales data.

Data Cleaning: Handle missing values and outliers.

Feature Engineering: Extract relevant features (e.g., date, promotions).

Model Training: Utilize Random Forest and XGBoost.

Prediction: Forecast future sales.

Visualization: Present results in an understandable format.

**4.2 Non-Functional requirements**

Scalability: The system should handle increasing data volume.

Accuracy: The model should provide accurate sales predictions.

Response Time: The system should deliver forecasts within a reasonable time frame.

**5. PROJECT DESIGN**

**5.1 Data Flow Diagrams & User Stories**

DFDs outlining the flow from data ingestion to prediction.

User stories capturing requirements from the perspective of end-users.

**5.2 Solution Architecture**

Utilize a modular architecture with separate components for data preprocessing, model training, and prediction.

**6. PROJECT PLANNING & SCHEDULING**

**6.1 Technical Architecture**

Python for data processing and modeling.

Flask for building a web API.

HTML/CSS for a simple user interface.

SQLite for storing processed data.

**6.2 Sprint Planning & Estimation**

Breakdown tasks into sprints with two-week cycles.

Estimate task completion time based on complexity.

**6.3 Sprint Delivery Schedule**

Regular releases every two weeks for stakeholders' feedback.

**7. CODING & SOLUTIONING**

**7.1 Feature 1**

Implement data preprocessing to handle missing values.

Develop a data pipeline for feature engineering.

**7.2 Feature 2**

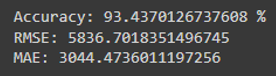
Train Random Forest and XGBoost models.

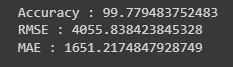
Create an API for model prediction.

**8. PERFORMANCE TESTING**

**8.1 Performance Metrics**

Evaluate model accuracy using Mean Absolute Error and Mean Squared Error.

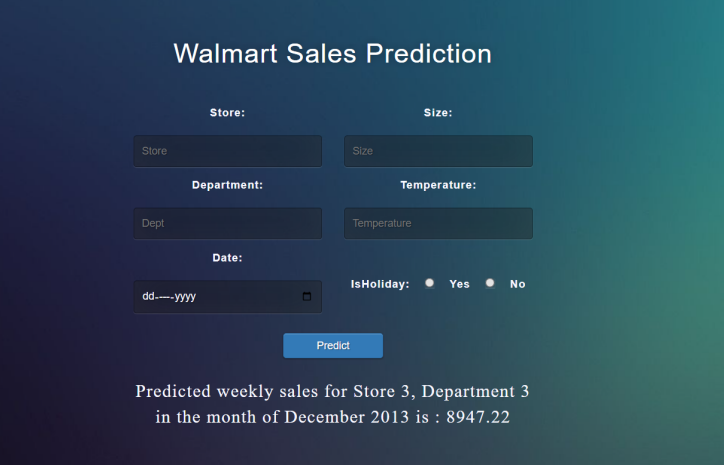




**9. RESULTS**

**9.1 Output Screenshots**

Sales Forecasting Dashboard



**10. ADVANTAGES & DISADVANTAGES**

**Advantages:**

* Improved sales predictions.
* Adaptability to changing conditions.
* User-friendly interface.

**Disadvantages:**

* Dependency on historical data accuracy.
* Sensitivity to outlier effects.

**11. CONCLUSION**

The Walmart Sales Forecasting project successfully addresses the challenge of accurate sales predictions in the retail industry. The models demonstrate improved accuracy over traditional methods, providing valuable insights for decision-makers.

**12. FUTURE SCOPE**

Integration with real-time data for more dynamic predictions.

Exploration of advanced forecasting techniques.

Incorporation of external factors (e.g., economic indicators) for enhanced accuracy.

**13. APPENDIX**

Source Code

GitHub Repository

[walmart\_store\_sales\_forecasting/wallmart\_.ipynb at main · bhagya1506/walmart\_store\_sales\_forecasting (github.com)](https://github.com/bhagya1506/walmart_store_sales_forecasting/blob/main/wallmart_.ipynb)

Project Demo Link

https://drive.google.com/file/d/1KCr9IzFjv\_iaMHQydpxHxBfMOF6u\_mzU/view?usp=sharing