

AI-Based Walmart Sales Forecasting System

This project implements an end-to-end machine learning pipeline to predict weekly sales for Walmart stores using historical sales, store, and feature data. The workflow covers data preprocessing, feature engineering, model training, evaluation, and visualization of results.

Goal

Predict future sales using historical data to improve demand planning.

Use Case

Applicable to retail, supply chain, e-commerce, and FMCG sectors for more accurate sales forecasting.

Features

- **Data Cleaning & Preprocessing:** Handles missing values, encodes categorical and boolean features, and engineers cyclical date features.
 - **Model Training:** Trains and compares Linear Regression, Lasso Regression, and Random Forest models.
 - **Evaluation:** Uses RMSE and R^2 metrics to evaluate model performance.
 - **Prediction:** Generates weekly sales predictions for unseen test data.
 - **Visualization:** Provides clear plots for predicted sales distribution, feature importance, and sales trends over time.
-

Workflow

1. **Data Loading:** Reads Walmart's train, test, features, and stores datasets.
 2. **Merging & Cleaning:** Merges datasets and prepares features for modeling.
 3. **Feature Engineering:** Encodes dates and categorical variables, normalizes data.
 4. **Modeling:** Trains multiple regression models and selects the best based on validation performance.
 5. **Prediction & Visualization:** Predicts future sales and visualizes results for business insights.
-

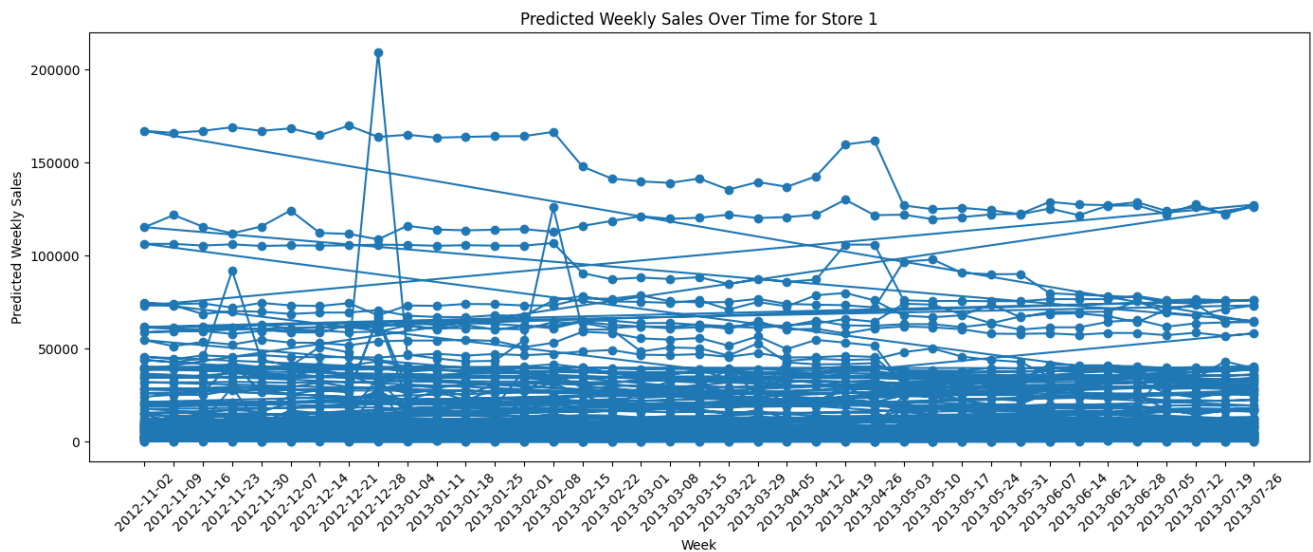
Results

- **Random Forest** achieved the best performance with an **R^2 of ~0.98** on the validation set.
- The notebook includes plots for predicted sales distribution and feature importance.

Model	RMSE	R^2
Linear Regression	21548.43	0.087658
Lasso Regression	21548.43	0.087657

Model	RMSE	R ²
Random Forest	2830.74	0.984256

Predicted Sales by Week



Date	Predicted Weekly Sales
2012-11-02	24,157.93
2012-11-09	25,253.07
2012-11-16	25,117.44
2012-11-23	27,777.37
2012-11-30	27,995.13

Getting Started

1. Clone the repository.
2. Install dependencies from `requirements.txt`.
3. Run the Jupyter notebook or Python scripts to reproduce results.

License

This project is for educational and research purposes.