Sales Forecasting Using Machine Learning & Deep Learning

Project Overview

This project implements an end-to-end, production-oriented sales forecasting system to predict daily store-level sales up to six weeks ahead. The solution integrates data preprocessing, EDA-driven feature engineering, machine learning, deep learning, and MLOps best practices with a strong focus on reproducibility and evaluation readiness.

Business Objective

- Forecast daily sales with high accuracy
- Capture seasonality, promotions, and store-level effects
- Support inventory, staffing, and revenue planning

Dataset Usage

Only train.csv and store.csv are used in this project. The provided test.csv and sample_submission.csv files are not used, as the project focuses on historical forecasting and evaluation rather than competition-based submission.

Methodology & Architecture

The workflow includes data cleaning, exploratory analysis, feature engineering, time-aware validation, model training, evaluation, experiment tracking using MLflow, and data/model versioning with DVC.

Modeling Approach

- Baseline and tree-based machine learning models
- LSTM-based deep learning model for temporal dependencies
- Time-series aware validation to prevent leakage

Evaluation Metrics

Model performance is evaluated using RMSE, MAE, and MAPE, with comparisons across classical ML and deep learning approaches.

MLOps Integration

- MLflow for experiment tracking and model versioning
- DVC for dataset and pipeline reproducibility

Key Insights

- Promotions significantly impact short-term sales
- Strong weekly and seasonal patterns observed
- Tree-based models outperform linear baselines

Author

Ankita Taneja
Aspiring Data Scientist
GitHub: https://github.com/Ankitatan