Sales Forecasting Project

Final Iteration Report

Ayman Mushtaq Ahmad, Amisha Tiwari, Ronhit Neema

November 24, 2024

Contents

1 Introduction

The Sales Forecasting Project aims to provide accurate forecasts for sales data using advanced machine learning techniques. This report details the progress and outcomes achieved in the final iteration, building upon the work completed in previous iterations.

2 Project Updates

2.1 Previous Iteration Summary

The previous iteration focused on data preprocessing, exploratory data analysis (EDA), and visualization. Each team member contributed to ensuring a streamlined and accurate dataset for model training.

2.2 Current Iteration Focus

The current iteration focused on:

- Implementing and comparing multiple machine learning models.
- Training models using optimized hyperparameters.
- Evaluating models based on R^2 scores and documenting insights.
- Visualizing results for better interpretability.

3 Task Breakdown

3.1 Ayman Mushtaq Ahmad

- Task: Model Development and Training
- Description: Implemented and trained machine learning models, including Linear Regression, Lasso, Ridge and Decision Trees. Applied hyper parameter tuning to improve performance.

3.2 Ronhit Neema

- Task: Model Evaluation and comparison
- Description: Evaluated models using R^2 scores, generating comparative analyses. Contributed to identifying the best-performing model and documenting its results.

3.3 Amisha Tiwari

- Task: Model Implementation, Visualization and Reporting
- Description: Worked alongside Ayman on machine learning models focusing on Random Forest and Extra Trees. Created insightful visualizations to represent model performances and forecast results. Consolidated findings into a comprehensive report.

4 Skills and Tools Used

- **Programming Languages:** Python
- **Libraries:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, sklearn, catboost, xgboost, lightgbm
- **Environment:** Jupyter Notebooks
- **Version Control:** GitHub (`https://github.com/ahmadaym/IDMP_SalesForecasting_Project`)

5 Progress Review

5.1 Achievements

- Successfully implemented and evaluated multiple models for sales forecasting.
- Achieved the best performance using Random Forest with an R^2 score of 0.89.
- Documented key insights into model performance and improvement strategies.

5.2 Challenges

- Ensuring the computational efficiency of complex models.
- Handling data preprocessing challenges, including outlier treatment and feature scaling.

6 Conclusion and Future Work

The final iteration successfully demonstrated the project objectives highlighted in our Project scope. We plan to complete the project over the coming weeks focusing on Project presentation and our technical report as well.

A Appendix

Detailed code and additional documentation are available in our GitHub repository: `https://github.com/ahmadaym/IDMP_SalesForecasting_Project`.