Duaa Fatima, 21i1667

Syeda Mahum Raza, 21i1662

Project Report: Sales Price Prediction with Machine Learning

Introduction

Predicting sales prices is crucial for businesses in various industries, including real estate and retail, to make informed decisions about pricing strategies and market trends. In this project, we aimed to predict sales prices using machine learning models. We collected data from Kaggle, focusing on sales prices of properties. The objective was to preprocess the data, apply various machine learning models, integrate them with Flask for web deployment, and visualize the predictions using Chart.js and D3.js.

Data Collection and Preprocessing

We collected sales price data from Kaggle, selecting a dataset that aligns with our project requirements. The dataset contained information about property sales, including features such as square footage, number of bedrooms, and location. We preprocessed the data to handle missing values by using interpolation techniques, ensuring the dataset is clean and ready for modeling.

Machine Learning Models

We applied eight machine learning models to predict sales prices:

1. ARIMA (AutoRegressive Integrated Moving Average)

2. SARIMA (Seasonal AutoRegressive Integrated Moving Average)

3. SVR (Support Vector Regression)

4. ETS (Error-Trend-Seasonality)

5. LSTM (Long Short-Term Memory)

6. Prophet

7. ANN (Artificial Neural Network)

8. Hybrid (ANN and Arima)

Each model was trained on the preprocessed dataset to learn the patterns and relationships between the features and the target variable (sales price). We evaluated the performance of each model based on metrics such as Mean Squared Error (MSE) and identified the best-performing models for sales price prediction.

Model Evaluation

Among the models, SVR showed the least promising results, which is expected as SVR is commonly used for classification tasks rather than regression. The rest of the models demonstrated good performance, with low MSE scores, indicating their effectiveness in predicting sales prices accurately.

Web Deployment with Flask

To make the models accessible to users, we integrated them with Flask, a web framework in Python. Users can upload their sales data through a web interface, select the desired model, and obtain predictions. We utilized Chart.js and D3.js for visualizing the predictions, providing users with interactive and informative graphs.

Retraining Models

Although attempted, the retraining functionality of the models was not fully implemented. Retraining allows models to adapt to new data and improve their accuracy over time. Future work involves completing the retraining mechanism to ensure the models remain up-to-date and effective in predicting sales prices.

Conclusion

In conclusion, this project successfully developed and deployed machine learning models for sales price prediction. By preprocessing the data, applying various models, and integrating them with Flask for web deployment, we created a user-friendly platform for sales price forecasting. Future enhancements include completing the retraining functionality and further improving model accuracy.