FLIGHT PRICE PREDICTION USING MACHINE LEARNING

Student Name: Naga mohan Thotakura

Student Id: 23070010

Supervisor: Darshan Khakkad

GitHub: repo

Introduction and Research Motivation

Airfare pricing is dynamic and unpredictable due to numerous influencing factors such as seasonality, flight duration, airline policies, and market demand. This project aims to build a predictive machine learning model that accurately estimates flight ticket prices. The growing interest in travel tech platforms and cost-conscious consumers has made flight price prediction an important use case for AI-driven solutions.

Research Question

Can machine learning algorithms accurately predict flight prices using publicly available data on flight attributes?

The objective is to develop a regression model that predicts flight prices based on features like departure and arrival times, airline, duration, stops, and class.

Literature Review

Previous research has explored both linear and non-linear models, with Random Forest and Gradient Boosting models showing superior performance due to their ability to capture complex relationships. For instance, studies by Prediction of Flight Fares Using Machine Learning By Manya Tuli and Flight Price Prediction Using Machine Learning by G D Vijaya lakshmi highlight how ensemble methods outperform simpler linear regression and other regression models. Our project builds on these insights and introduces a clean, reproducible pipeline with well-documented preprocessing and hyperparameter tuning steps.

Methodology

- **Dataset:** Kaggle Flight Fare Dataset (Domestic Indian flights). Originally collected from EaseMyTrip.
- **Data Preprocessing:** Included encoding categorical variables (OneHot and Label), converting date/time to numerical features, and removing outliers.
- **Feature Engineering:** Introduced travel duration in minutes, total stops as numeric, and departure/arrival time bins.
- **Models Trained:** Random Forest, and XGBoost
- **Hyperparameter Tuning:** GridSearchCV with 5-fold cross-validation for RandomForest and XGBoost

Exploratory Data Analysis:

- **Bar Plot:** The bar plot represents the relation between the city and the average price.

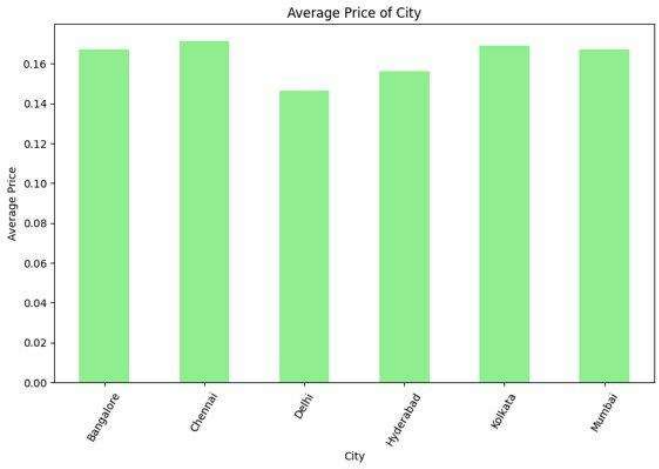
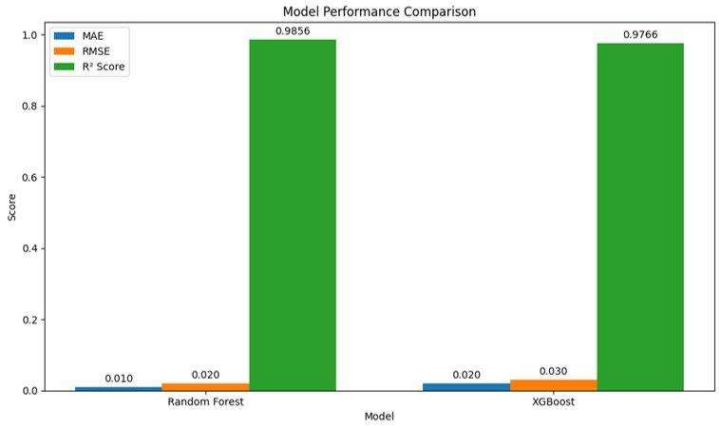


Fig: Bar Plot

Results and Evaluation:



- **Evaluation Metrics:** Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R^2 Score

Model	MAE	RMSE	R^2 Score
Random Forest Regression	0.010	0.020	0.9856
XGBoost Regression	0.020	0.030	0.9766

- **Key Insights:** Airline type and number of stops significantly influence prices. Time-of-day features also showed a moderate correlation.

Discussion and Implications: The model's performance suggests it can be used in applications like fare alert systems, price tracking, or travel advisory bots. Comparing results to literature confirms that ensemble models remain robust for tabular prediction tasks.

Conclusion:

This project demonstrates the effectiveness of machine learning for flight price prediction. Future work could involve deploying the model as a web app using Streamlit or integrating real-time APIs for live fare updates.