Flight Fare Prediction Using Machine Learning Models

Submission by:

Group 15

Akhila Sannayila – AkhilaSannayila@my.unt.edu

Ram Prasad Jampani – RamPrasadJampani@my.unt.edu

Akhil Ganji – AkhilGanji2@my.unt.edu

1. Motivation:

Airline ticket prices are uncertain as the English weather. In a 24-hour period, you may come across the same seat at different costs. It is also affected by seasons, days and weeks of the week, holidays, and events. Prices may even fall if there are fewer reservations. The reason for this is because hospitality and transit providers, online travel agents, and aggregators all work hard to get users to click the "book" button. Prices rise even more during busy seasons such as Christmas and New Year's.

In this project we will analyse the flight fare prediction using Machine Learning dataset with the help of essential exploratory data analysis techniques, and then make some predictions about the price of the flight based on some features such as source, destination, what type of airline it is, what is the arrival time, what is the departure time, how long the flight is and many more.

2. Task Definition:

Flight fare prediction using machine learning connects numerous technological features to increase the efficiency of forecasting air travel expenses and accuracy, thereby benefiting both passengers and the aviation industry. This method combines digital image processing, medical imaging, machine learning and data preparation, with a heavy emphasis on data security and aviation integration.

a. **Input:**

   Features: These features contain Airline, Duration, Class, Stops, Arrival and Departure timings

b. **Output:**

   Target: We'll predict the Price of the ticket which is target variable.

3. Approach:

- **Data collection:**
  - We will use Flight Price dataset from Kaggle and load the data.
  - Data has no missing values and duplicate values.
- **Data visualization:**
  - We'll plot Price vs Airline, Prive vs Destination, Price vs Source.

- **Preprocessing:**
  - Then we will perform a feature scaling by convert different units and magnitude data into single unit.
  - Then we'll split the data into train data and test data.
- **Feature Engineering:**
  - Divide the features and labels to convert the hours into minutes.
- **Machine Learning model:**
  - The baseline of our project will be a Linear regression model.
  - We'll implement different regression methods and train them to check which model gives the best accuracy and prediction.
  - **Linear regression:** Predicting flight prices using linear regression is an interesting and practical application of machine learning. In this approach, you'll leverage historical flight data to build a model that estimates flight ticket prices based on various factors.
  - **Decision Tree Regressor:** Decision trees are a type of machine learning model that can handle both categorical and numerical data, making them suitable for flight price prediction.
  - **Random Forest Regressor:** Predicting flight prices using a Random Forest Regressor is a powerful and versatile approach. Random Forest is an ensemble learning method that combines multiple decision trees to improve predictive accuracy and generalization.
  - **Support Vector Regressor:** Predicting flight prices using Support Vector Regression (SVR) is another viable approach. SVR is a machine learning technique that can handle both linear and non-linear relationships between features and target values.
  - **K-Nearest Neighbours Regressor:** Predicting flight prices using the K-Nearest Neighbours Regressor (KNeighborsRegressor) is a straightforward and intuitive approach. K-Nearest Neighbours is a simple, instance-based algorithm that estimates the value of a new data point based on the values of its k-nearest neighbours from the training data.

4. Proposed Methods:

To achieve the goal of our project, we plan to do the following. Exploratory Data Analysis, Feature Engineering, Machine learning algorithms like Linear Regression, Support Vector regressor, Random Forest Regressor, K-Nearest Neighbours Regressor and Decision Tree Regressor. We will select a single model with the best accuracy and predictions.

5. Evaluation Metric:

The evaluation measures used in flight price detection are determined by the project's individual goals and needs. To analyse the efficiency of models we would use

- Mean Absolute Error
- Adj_R_Square
- Root Mean Squared Error
- Mean Absolute Percentage Error
- Mean Squared Error
- R2 Score

The criteria used in flight fare prediction are determined by the project's aims and the balance of forecast accuracy, consistency, and fairness. These parameters helps how successfully the model estimates flight prices, allowing travellers to make decisions and airlines to efficiently manage pricing strategies.

6. Data:

This dataset contains 30,0261 data points and 11 features collected from the flight booking website “EaseMyTrip” for flight travels between 6 cities in India. It contains feature like Airline, Flight, Source City, Destination City, Departure Time, Arrival Time, Stops, Destination City, Class, Duration, Days left and Price

7. Experiments:

We plan quantitative studies in addition to quantitative metric evaluations:

a. Quantitative Trials:

- Methodical hyperparameter tuning of machine learning algorithms improves model performance.
- Using k-fold cross-validation to evaluate model generalization.

- Comparing machine learning models to find the best one.

b. Qualitative analysis:

- Method of Analysis: We plan to conduct the following qualitative analyses without numerical data.
- Visual Inspection: Finding patterns to understand the model's limits.

8. Plan

a. Team roles

Name: Akhil Ganji

Student ID: 11701971

Data for Flight Fare Prediction project is organized by me. I find, check, and clean the dataset. My data preprocessing includes feature extraction and augmentation. Data privacy. And will work on the model development and training.

Name: Akhila Sannayila

Student ID: 11702648

I improve models as a Machine Learning Engineer. Support Vector Regressor, Random Forest, Decision Tree, and Linear Regression are my main methods. Baseline model building and hyperparameter optimization are my responsibility. I train and validate models with our data. I'll review the model with the Metrics Analysis Specialist and offer changes.

Name: Ram Prasad Jampani

Student ID: 11702639

My main emphasis is testing machine learning models. I evaluate model performance using Mean Absolute Error, Adj_R_Square, Root Mean Squared Error, Mean Absolute Percentage Error, Mean Squared Error and R2 Score

Finding the best regression model for our project required rigorous comparison. Quantitative measures help me evaluate and enhance our models. I analyze model performance to help our project succeed. And also works on model development and training.

9. Timeline:

- **Oct 21, 2023 –** Project proposal submission.
- **Oct 27, 2023 –** Preprocessing and visualization of data
- **Nov 3, 2023 –** Fitting the model with training data
- **Nov 10, 2023 –** Testing the model using test data and check the performance metrics.
- **Nov 13, 2023 –** selecting the final model for the production deployment.

10. Conclusion

The use of linear regression models to estimate flight prices is a realistic and beneficial activity. It addresses the issues that passengers and airlines encounter, as well as providing insights into an industry that is vital to the global economy. This method improves the experience for all stakeholders involved in air travel by leveraging the power of data and machine learning.