Flight Price Prediction Project

Analyzing Airline Data using Machine Learning in python

BY VIRAJ MATHUR

Introduction

1. **Objective:**
    - Our goal is to predict flight prices using machine learning algorithms.
    - Accurate price prediction is crucial for travelers to make informed decisions.
2. **Importance:**
    - Flight prices are dynamic and influenced by various factors.
    - Predicting prices helps users plan cost-effective and efficient travel.
3. **Motivation:**
    - Rising demand for air travel necessitates effective price prediction.
    - Machine learning enables us to analyze complex patterns and make accurate predictions.

Dataset Overview

Source

I gathered a comprehensive dataset from multiple sources like Kaggle that contains information about flights, such as destination, duration, and airlines.

Features

With more than 20 features, including departure time, number of stops, and airline, we have a rich set of data to work with.

Target Variable

Our Target variable is the Flight Price

Data Preprocessing

1

Data Cleaning

Removed duplicates, handled missing values, and standardized the data to prepare it for feature engineering.

2

Feature Selection

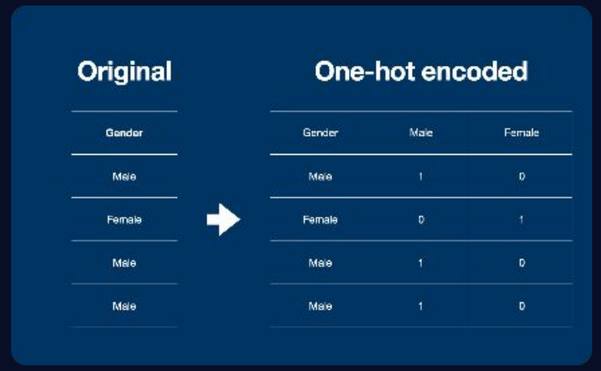
Selected the most relevant features based on their impact on the target variable and their correlation with other features.

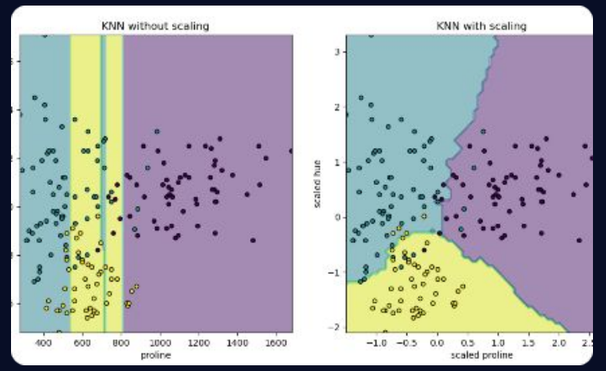
3

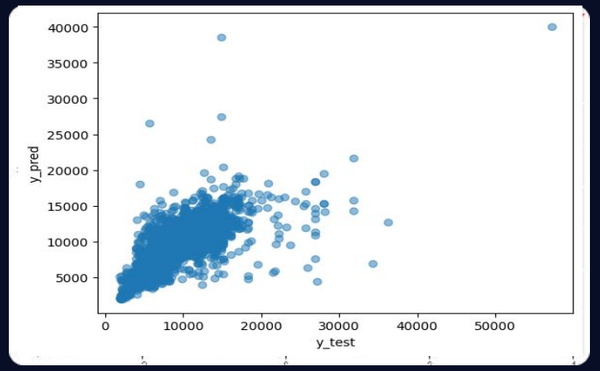
Outlier Detection

Applied robust techniques to identify and handle outliers, ensuring that they do not negatively impact our models.

Feature Encoding and Visualization







One-Hot Encoding

Transformed categorical variables into numerical representations using one-hot encoding to make them compatible with the model.

Feature Scaling

Applied feature scaling techniques to normalize the numerical features and enhance model performance.

Feature Visualization

Created visualizations, such as scatter plots to understand the relationship between features and flight prices.

Machine Learning Algorithms

Linear Regression

We implemented linear regression, a simple yet powerful algorithm, to develop a baseline model for flight price prediction.

Random Forest

Random forest algorithm was employed to handle non-linear relationships and capture complex patterns in the data.

K-Nearest Neighbors (KNN)

Using KNN we classified the majority class of our models among k-nearest neighbors in feature space.

Decision Tree

With Decision Tree we made a Tree-like model making decisions based on features, splitting data to classify or regress with high interpretability.

Model Training and Evaluation

1

Training

Training the model using a portion of the dataset and ensuring that it did not contain redundant and unnecessary data, no missing values and no string values- had uniform data types.

2

Validation

Evaluating the model using cross-validation techniques and assessing their performance metrics, such as RMSE and R^2, to measure accuracy.

3

Testing

Finally, we testing the model on various Machine Learning Model to check which algorithm give the best accuracy for predicting the price of the airline.

Results and Comparison

Model Comparison

We compared the performance of different algorithms to identify the most accurate model for flight price prediction.

Model Performance

LINEAR REGRESSION:

Training score : 0.6241794859094292
r2 score is : 0.6241794859094292
MAE : 1949.458356115105
MSE : 7835152.949901845
RMSE : 2799.1343215183233

RANDOM FOREST:

Training score : 0.9541554467344304
r2 score is : 0.9541554467344304
MAE : 1161.0565396248683
MSE : 4101314.27200073
RMSE : 2025.1701834662513

K-NEAREST NEIGHBOURS:

Training score : 0.6241794859094292
r2 score is : 0.7306851009410837
MAE : 1873.2053912392364
MSE : 8751130.411591165
RMSE : 2958.2309598121587

DECISION TREE:

Training score : 0.9707490055980877
r2 score is : 0.9707490055980877
MAE : 1359.1213652814179
MSE : 5995415.448104538
RMSE : 2448.5537462152097

Conclusion

- In conclusion, our flight price prediction project demonstrates the efficacy of machine learning algorithms in accurately forecasting flight prices.

- We found that Decision Tree and Random Forest Algorithms are the most suitable machine learning model to predict the airline prices.

- Using these model we have the potential to transform the travel industry and empower travelers to make informed decisions.

Future Work

In the future, I plan to further improve the model by incorporating additional features such as weather conditions and airline promotions. I also aim to develop a user-friendly application that allows travelers to obtain real-time flight price predictions.

THANK YOU