

# Project Demonstration & Documentation

---

## Optimizing Flight Booking Decisions Through Machine Learning Price Prediction

### INTRODUCTION

#### Project Description

People who work frequently travel through flight will have better knowledge on best discount and right time to buy the ticket. For the business purpose many airline companies change prices according to the seasons or time duration. They will increase the price when people travel more. Estimating the highest prices of the airlines data for the route is collected with features such as Duration, Source, Destination, Arrival and Departure. Features are taken from chosen dataset and in the price wherein the airline price ticket costs vary overtime. We have implemented flight price prediction for users by using KNN, decision tree and random forest algorithms. Random Forest shows the best accuracy of 80% for predicting the flight price. Also, we have done correlation tests and metrics for the statistical analysis.

#### Technical Architecture:

#### Project Flow:

User interacts with the UI to enter the input.

Entered input is analysed by the model which is integrated.

Once model analyses the input the prediction is showcased on the UI

To accomplish this, we have to complete all the activities listed below,

## Define Problem / Problem Understanding

Specify the business problem

Business requirements

Literature Survey

Social or Business Impact.

## Data Collection & Preparation

Collect the dataset

Data Preparation

Exploratory Data Analysis

Descriptive statistical

Visual Analysis

## Model Building

Training the model in multiple algorithms

Testing the model

Performance Testing & Hyperparameter Tuning

Testing model with multiple evaluation metrics

Comparing model accuracy before & after applying hyperparameter

## **2.LITERATURE SURVEY**

## Existing problem

We have implemented flight price prediction for users by using KNN, decision tree and random forest algorithms. Random Forest shows the best accuracy of 80% for predicting the flight price. Also, we have done correlation tests and metrics for the statistical analysis.

## 2.2 PROPOSED SOLUTION

As the data is increasing daily due to digitization in the banking sector, people want to apply for loans through the internet. Machine Learning (ML), as a typical method for information investigation, has gotten more consideration increasingly. Individuals of various businesses are utilising ML calculations to take care of the issues dependent on their industry information. Banks are facing a significant problem in the approval of the loan. Daily there are so many applications that are challenging to manage by the bank employees, and also the chances of some mistakes are high. Most banks earn profit from the loan, but it is risky to choose deserving customers from the number of applications. There are various algorithms that have been used with varying levels of success. Logistic regression, decision tree, random forest, and neural networks have all been used and have been able to accurately predict loan defaults. Commonly used features in these studies include credit score, income, and employment history, sometimes also other features like age, occupation, and education level.

## 3.THEORITICAL ANALYSIS

### 3.1 block diagram

#### Descriptive Statistical Analysis

Descriptive analysis is to study the basic features of data with the statistical process. Here pandas has a worthy function called describe. With this describe function we can understand the unique, top and frequent values of categorical features. And we can find mean, std, min, max and percentile values of continuous features.

#### Toggle navigation

0

Dashboard

Internship

Support

Training Calendar

Guided Project

Project Workspace

## Visual Analysis

Visual analysis is the process of using visual representations, such as charts, plots, and graphs, to explore and understand data. It is a way to quickly identify patterns, trends, and outliers in the data, which can help to gain insights and make informed decisions.

Plotting countplots for categorical data

We now plot distribution plots to check the distribution in numerical data (Distribution of 'Price' Column)

The `seaborn.displot()` function is used to plot the displot. The displot represents the univariate distribution of data variable as an argument and returns the plot with the density distribution. Here, I used `distribution(displot)` on 'Price' column.

It estimates the probability of distribution of continuous variable across various data.

Training The Model In Multiple Algorithms

Using Ensemble Techniques

`RandomForestRegressor`, `GradientBoostingRegressor`, `AdaBoostRegressor`

A function named `RandomForest`, `GradientBoosting`, `AdaBoost` is created and train and test data are passed as the parameters. Inside the function, `RandomForest`, `GradientBoosting`, `AdaBoost` algorithm is initialized and training data is passed to the model with `.fit()` function. Test data is predicted with `.predict()` function and saved in new variable. For evaluating the model, `r2_score`, `mean_absolute_error`, and `mean_squared_error` report is done.

## Testing The Model

In ANN we first have to save the model to the test the inputs

This code defines a function named “predict\_exit” which takes in a sample\_value as an input. The function then converts the input sample\_value from a list to a numpy array. It reshapes the sample\_value array as it contains only one record. Then, it applies feature scaling to the reshaped sample\_value array using a scaler object ‘sc’ that should have been previously defined and fitted. Finally, the function returns the prediction of the classifier on the scaled sample\_value.

## 4.RESULT

### Testing Model With Multiple Evaluation Metrics

#### Hypertuning the model

RandomSearch CV is a technique used to validate the model with different parameter combinations, by creating a random of parameters and trying all the combinations to compare which combination gave the best results. We apply random search on our model.

From sklearn, `cross_val_score` is used to evaluate the score of the model. On the parameters, we have given `rf` (model name), `x`, `y`, `cv` (as 3 folds). Our model is performing well.

### Evaluate The Results

From sklearn, `cross_val_score` is used to evaluate the score of the model. On the parameters, we have given `rfr` (model name), `x`, `y`, `cv` (as 3 folds). Our model is performing well. So, we are saving the model by `pickle.dump()`.

## 5.ADVANTAGES AND DIS ADVANTAGES

### 5.1 benefits

One of the primary benefits of using machine learning algorithms to predict flight prices is the accuracy of the predictions. The algorithms analyse vast amounts of data, including historical data on flight prices, weather patterns, customer booking behaviour, and other variables.

### 5.2 disadvantages

Using machine learning algorithms on the dataset, one can predict the dynamic fare of flights, thereby obtaining the predicted flight fare values to obtain a flight ticket at the lowest cost. The accuracy of the model is determined by the R-squared values obtained from the algorithm.

## 6.APPLICATIONS

This model helps its users by advising them whether to buy tickets or wait for a suitable time to get the optimal deal. It uses data mining techniques like Rule Learning, Reinforcement Learning, time-series methods, and their combinations to achieve greater accuracy in predicting the fare of flights.07-Feb-2022

Below are some most trending real-world applications of Machine Learning:

Image Recognition: ...

Speech Recognition. ...

Traffic prediction: ...

Product recommendations: ...

Self-driving cars: ...

Email Spam and Malware Filtering: ...

Virtual Personal Assistant: ...

Online Fraud Detection:

## 7.CONCLUSION

### 7.1

Conclusion:- By

Using machine learning algorithms on the dataset, one can predict the dynamic fare of flights, thereby obtaining the predicted flight fare values to obtain a flight ticket at the lowest cost. The accuracy of the model is determined by the R-squared values obtained from the algorithm.



## 8.FUTURE SCOPE

We have implemented flight price prediction for users by using KNN, decision tree and random forest algorithms. Random Forest shows the best accuracy of 80% for predicting the flight price. Also, we have done correlation tests and metrics for the statistical analysis.