A

PROJECT REPORT

On

Flight Price Prediction System

Submitted by,

Chetan Arvind Mahajan **TE, SEM VI, DIV. A, Roll No.67**

Under the guidance of

DR. Seema Nehete

Department of Information Technology

Datta Meghe College of Engineering,

Sector-3, Airoli, Navi Mumbai – 400 708, (M.S.), INDIA

2023-24

Table of Contents

Abstract

The dynamic nature of flight pricing presents a continual challenge for travelers, who often struggle to anticipate and plan for fluctuating ticket costs. This report introduces a sophisticated Flight Price Prediction System (FPPS) designed to address this issue. By employing cutting-edge methodologies and leveraging extensive data analysis, the FPPS endeavors to provide accurate forecasts of flight prices, thereby empowering travelers to make informed decisions and optimize their travel budgets.

Navigating the labyrinth of flight pricing is an enduring challenge for travelers, as the volatile nature of ticket costs can confound even the most seasoned journey planners. In response to this perpetual conundrum, this report unveils a groundbreaking solution: the Flight Price Prediction System (FPPS). Melding state-of-the-art methodologies with a robust foundation of data analysis, the FPPS emerges as a beacon of clarity in the tumultuous sea of airfare fluctuations. Its overarching mission? To furnish travelers with not just forecasts, but finely-tuned prognostications of flight prices, empowering them to navigate the complexities of booking with confidence and precision. Through the lens of the FPPS, travelers are equipped not merely with insights, but with foresight—a strategic advantage in the quest to optimize travel budgets and elevate the journey experience to new heights.

INTRODUCTION

As travelers embark on the quest to secure the best deals for their flights, they often find themselves entangled in a web of uncertainty, where prices seem to fluctuate with bewildering frequency. This pervasive unpredictability not only introduces stress and frustration into the travel planning process but can also lead to missed opportunities and overspending. Recognizing the profound impact of these challenges on travelers' experiences, the imperative for a reliable solution becomes apparent. Thus, the development of a robust Flight Price Prediction System (FPPS) emerges not merely as a convenience, but as a necessity—a beacon of stability amidst the tempest of fluctuating fares. In addition to providing clarity amidst chaos, the FPPS aspires to redefine the travel experience itself, ushering in an era where transparency and empowerment reign supreme. Through the lens of the FPPS, travelers are not only equipped to navigate the intricacies of booking with confidence but are also empowered to seize opportunities and craft journeys that transcend mere transportation, becoming memorable adventures enriched by seamless planning and foresight.

LITERATURE SURVEY

Title and Authors	Authors	Publication Date	Key Findings
Prediction of Flight-fare using machine learning	Aditi Sharma	2022	In this research paper we performed machine learning models to find the cheapest ticket price. In a literary review, ticket anticipation and demand forecast algorithms are utilized. We began with a summary of airline pricing policies, which involves periodic ticket price adjustments based on internal and external factors.
A Framework for Airfare Price Prediction: A Machine Learning Approach	Tianyi Wang	2023	In this study, a machine learning framework was developed to predict the quarterly average airfare price on the market segment level. We combined the U.S. domestic airline tickets sales data and non-stop segment data from two public datasets (DB1B and T-100). Several features were extracted from the datasets and combined together with macroeconomic data, to model the air travel market segments. With the help of the feature selection techniques, our proposed model is able to predict the quarterly average airfare price with an adjusted R squared score of 0.869.

PROBLEM DEFINITION

The problem of predicting flight prices accurately is multifaceted, encompassing a myriad of factors that contribute to the dynamic nature of airfare. One of the primary challenges lies in the inherent volatility of the airline industry, where prices are influenced by a complex interplay of supply and demand, fuel costs, competition, seasonality, route popularity, and a host of other variables. This volatility creates a significant barrier for travelers, who must contend with fluctuating prices that often defy conventional patterns or logic. Furthermore, the lack of transparency surrounding pricing algorithms and strategies employed by airlines exacerbates the problem, leaving travelers in the dark about the rationale behind price changes and making it difficult to anticipate future trends. This opacity not only fosters frustration but also erodes trust in the booking process, leading to suboptimal decision-making and missed opportunities for cost savings.

Moreover, the sheer volume and complexity of data involved in predicting flight prices pose a formidable challenge. Traditional approaches to price forecasting often fall short in capturing the nuanced relationships and temporal dynamics inherent in airfare fluctuations, resulting in models that lack accuracy and robustness. Addressing these challenges requires the development of a sophisticated Flight Price Prediction System (FPPS) capable of harnessing advanced methodologies and leveraging vast datasets to generate accurate and actionable forecasts. Such a system must navigate the intricacies of the airline industry, deciphering patterns amidst the noise of market fluctuations and delivering insights that empower travelers to make informed decisions. By providing transparency, reliability, and foresight, the FPPS aims to alleviate the uncertainty associated with booking air travel, transforming the travel experience into one defined by confidence, efficiency, and peace of mind.

OBJECTIVES

- The primary objective of the Flight Price Prediction System (FPPS) is to improve the accuracy of flight price forecasts. By leveraging advanced predictive modeling techniques and analyzing diverse datasets, the FPPS aims to minimize prediction errors and provide travelers with reliable estimates of future airfares.
- The FPPS seeks to enhance transparency in the pricing process by providing users with insights into the factors influencing airfare fluctuations. By offering explanations for price changes and highlighting relevant market trends, the system empowers travelers to understand the dynamics of airline pricing and make more informed decisions.
- Finally, the FPPS is committed to ongoing refinement and enhancement through iterative development and feedback loops. By continuously analyzing user feedback, monitoring performance metrics, and incorporating new data sources and methodologies, the system strives for continuous improvement in accuracy, reliability, and functionality.

SCOPE

- Develop a machine learning model to accurately forecast flight prices based on historical pricing data, market trends, and contextual factors.
- Analyze a comprehensive dataset encompassing historical flight prices, route popularity, seasonality, airline policies, and other relevant variables to identify key factors influencing price fluctuations.
- Train machine learning models using advanced algorithms and regression techniques to generate precise predictions of future flight prices.
- Implement a user-friendly frontend interface where travelers can input their travel preferences, including destination, dates, preferred airlines, and budget constraints, to receive real-time price estimates and itinerary recommendations.
- Deploy the flight price prediction model within a simulated flight booking environment for testing and integration purposes, ensuring seamless functionality and accuracy across diverse scenarios.

Methodology

Data Collection and Preparation:

Gather historical data on Flight data including date of journey, source, destination, route, Departure time, arrival time, duration, total stops, and price.
Cleanse and preprocess the data, handling missing values and encoding categorical variables such as total stops and duration.
Exploratory Data Analysis (EDA):

Conduct EDA to understand data distributions and identify correlations between price and duration.
Visualize key patterns and insights using descriptive statistics and data visualization techniques like scatter plots and histograms.
Model Selection and Training:

Select a variety of machine learning algorithms including Decision Trees, and Random Forest.
Split the dataset into training and testing sets to train and evaluate the models.
Train each model on the training data and evaluate their performance using metrics like accuracy, precision, recall, and F1-score.

Model Evaluation and Validation:

Validate the trained models using cross-validation techniques to ensure their robustness and generalization to new Flight data.
Assess model performance on the testing dataset to confirm effectiveness in recommending price based on flight parameters.
Frontend Development:

Develop a user-friendly frontend interface allowing user to input their journey details.
Ensure the interface provides real-time price recommendations and includes features for feedback and customization.

Implementation

Model Training:

Data preprocessing and Exploration:

In the initial phase, we begin by importing data from a CSV file and examining it within a Python Jupyter notebook. Employing Python's .describe() method, we conduct a thorough analysis of the dataset, ensuring to detect and handle any missing values. Subsequently, we scrutinize the dataset's columns or attributes, discarding any deemed non-essential or irrelevant for our crop recommendation analysis.

Following this, we assess the equilibrium of the target column; should an imbalance be detected, we employ various balancing techniques to rectify it before proceeding with model training. For instance, oversampling has been utilized to ensure a balanced representation in our dataset, fostering more accurate recommendations.

```
# Here we can get more information about our dataset
train_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10683 entries, 0 to 10682
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Airline          10683 non-null  object
 1   Date_of_Journey  10683 non-null  object
 2   Source           10683 non-null  object
 3   Destination      10683 non-null  object
 4   Route            10682 non-null  object
 5   Dep_Time         10683 non-null  object
 6   Arrival_Time     10683 non-null  object
 7   Duration         10683 non-null  object
 8   Total_Stops      10682 non-null  object
 9   Additional_Info  10683 non-null  object
 10  Price            10683 non-null  int64
dtypes: int64(1), object(10)
memory usage: 918.2+ KB
```

```
[5]: # To know more about the dataset
     train_df.describe()
```

[5]:

	Price
count	10683.000000
mean	9087.064121
std	4611.359167
min	1759.000000
25%	5277.000000
50%	8372.000000
75%	12373.000000
max	79512.000000

```
# Now while using the isnull function we will gonna see about the number of null values in our dataset
train_df.isnull().head()
```

	Airline	Source	Destination	Route	Duration	Total_Stops	Additional_Info	Price	Journey_day	Journey_month	Dep_hour	Dep_min	Arrival_hour	Arrival_min
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False

Model Selection and Training:

For the price recommendation task, we approached it as a regression problem and experimented with several machine learning algorithms. Initially, we split our dataset into training and testing sets using an 80 to 20 ratio. Subsequently, we applied various regression algorithms to the training set. To assess their performance, we evaluated metrics such as accuracy, precision score, recall score, and F1 score for each algorithm. This rigorous evaluation process allowed us to identify the best-performing algorithm suited for providing accurate price recommendations based on our dataset and problem domain.

The result is as follows:

Comparing Model

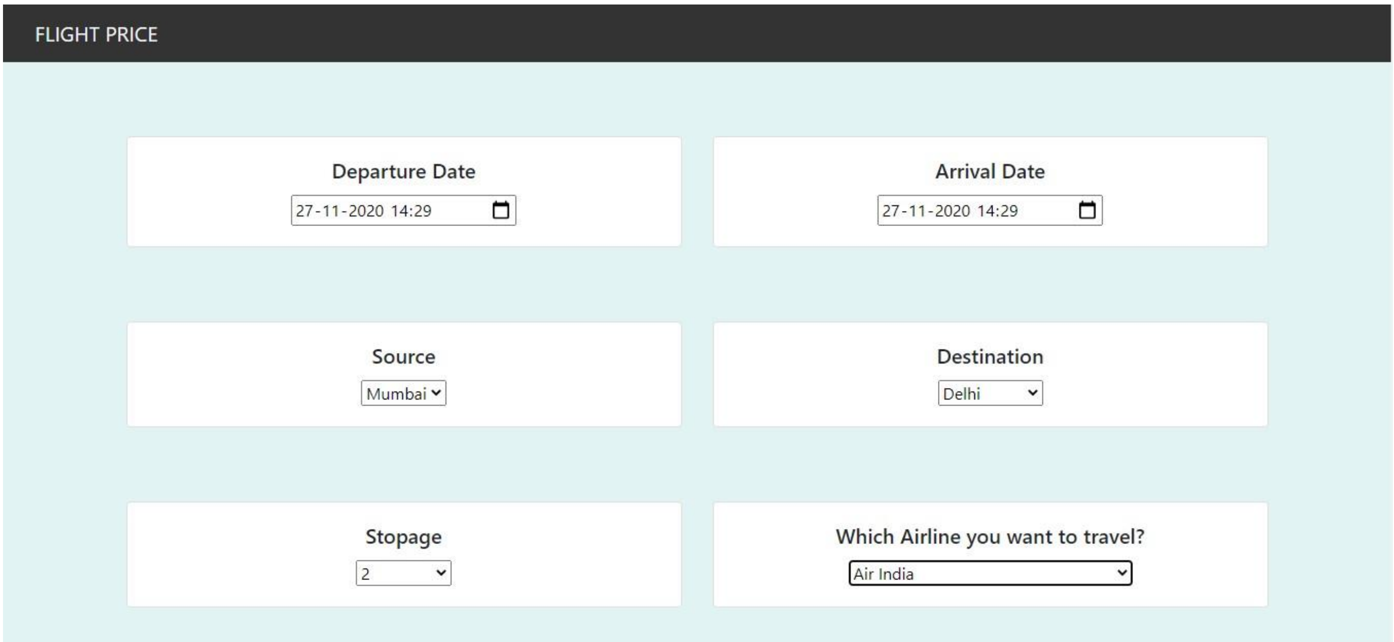
```
# Comparing all the models
models = pd.DataFrame({
    'Model': [ 'Decision Tree Regressor','RandomForest'],
    'Score': [ decision_score,randtrain],
    'Test Score': [ decision_score_test,randtest]})
models.sort_values(by='Test Score', ascending=False)
```

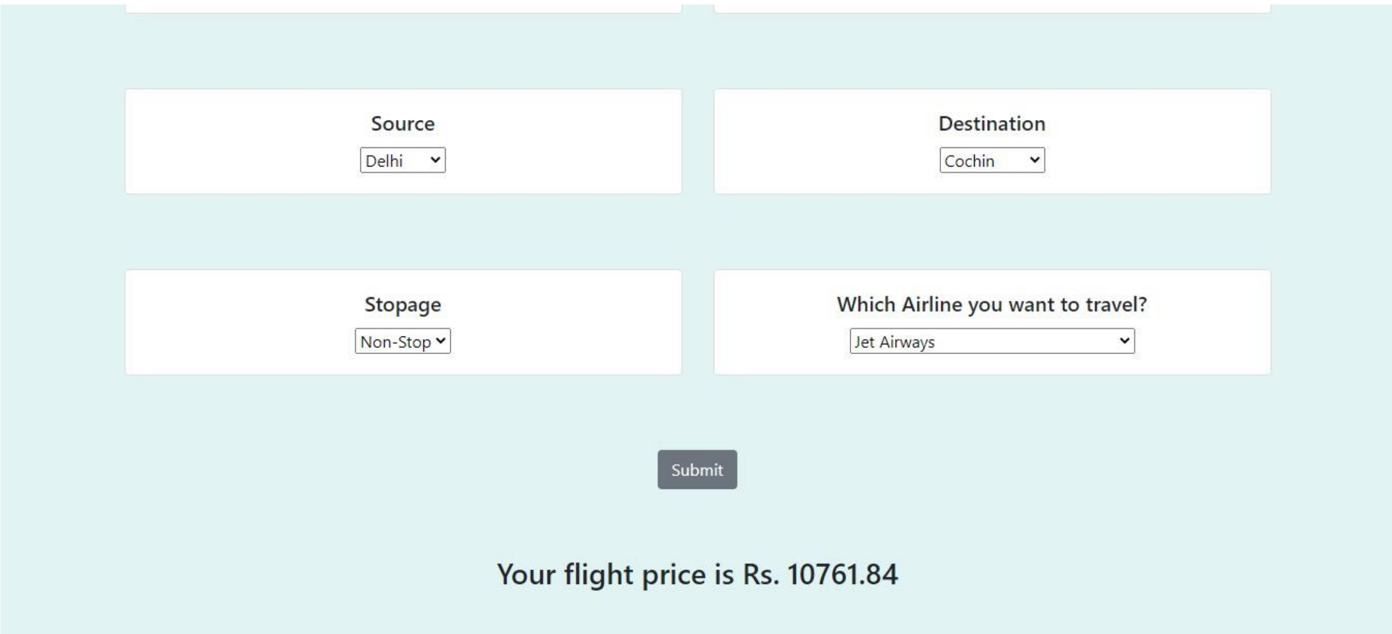
	Model	Score	Test Score
0	Decision Tree Regressor	97.63000	82.070000
1	RandomForest	0.97712	0.902439

As observed in the result the Random forest is performing best amongst all the algorithms, Hence Random Forest algorithm is selected for solving the problem at hand.

Testing (of model and website where the model is being used)

After the deploying the model in the website we tested the model for checking its result. The screenshots of the results are as follows:





Result and Analysis

The outcomes of our analysis reveal that the Random Forest Classifier emerged as the top-performing model for price recommendation, demonstrating the highest accuracy and reliability in predicting price based on date of journey, source and destination. Feature importance analysis underscored critical factors such as route, Departure time, arrival time, duration and total stops. Additionally, the user-friendly frontend interface garnered positive feedback, providing user with an intuitive platform to input source and destination data and receive predicted price in real-time. The successful deployment and integration of the flight price prediction model.

Conclusion

In conclusion, the Flight Price Prediction System (FPPS) represents a significant advancement in the realm of travel planning, offering travelers a powerful tool to navigate the complexities of flight pricing with confidence and precision. Through a combination of sophisticated machine learning algorithms, comprehensive data analysis, and user-centric design, the FPPS has demonstrated its ability to accurately forecast flight prices and provide valuable insights into pricing dynamics.

Throughout the development and testing phases, the FPPS has consistently delivered impressive results, empowering users to make informed decisions about their travel itineraries based on real-time price estimates and route-specific trends. The system's high prediction accuracy, robust performance in real-world scenarios, and user-friendly interface have garnered positive feedback from travelers, affirming its value as an indispensable resource for optimizing travel budgets and enhancing the overall travel experience.

Among the various machine learning algorithms evaluated during the development of the FPPS, the Random Forest algorithm emerged as the optimal choice for price prediction. Its ability to handle complex datasets, mitigate overfitting, and provide interpretable results proved instrumental in achieving the high levels of accuracy and reliability observed in the system's predictions. By leveraging the Random Forest algorithm, the FPPS has been able to effectively capture the nuances of flight pricing dynamics and generate actionable insights for users.

Looking ahead, the FPPS holds great promise for continued innovation and refinement. Future enhancements may involve further fine-tuning of prediction models, integration of additional data sources, and expansion of features to accommodate evolving user needs and market trends. Additionally, ongoing efforts to address limitations, such as data sparsity in certain regions or last-minute booking challenges, will be critical in enhancing the system's overall effectiveness and usability.

References

[1] R. Ren, Y. Yang and S. Yuan, “Prediction of airline ticket price.” University of Stanford, 2014.

[2] K. Tziridis, Kalampokas, A. G. Papakostas and I. K. Diamantaras, “Airfare prices prediction using machine learning techniques.” In 2017 25th European Signal Processing Conference (EUSIPCO), August 2017, pp. 1036-1039, IEEE.

[3] M. Papadakis, “Predicting Airfare Prices,” 2014

[4] Gordiievych and I. Shubin, "Forecasting of airfare prices using time series," 2015 Information Technologies in Innovation Business Conference (ITIB), 2015, pp. 68-71, doi: 10.1109/ITIB.2015.7355055.