Flight Price prediction using ml

Ch. Sandeep Reddy[1], Ch. Bharath[2] ,S.Jyothsna[3]
[1] Student, Information Technology, B.V Raju Institute of Technology, Telangana, India
[2] Student, Information Technology, B.V Raju Institute of Technology, Telangana, India
[3] Assistant Professor, Information Technology, B.V Raju Institute of Technology, Telangana, India

ABSTRACT

The pricing of flight tickets are susceptible to large increase or decrease, depending on parameters such as the schedule of the flights, the source, the destination, and the travelling duration of flight, other festivals and events. As a consequence, having a general concept of the prices of the airlines before starting to organise the trip will surely benefit a significant number of individuals in saving money. The gathered flight history will be employed in our approach to guide the building of a prediction system, which will be performed by employing machine learning approaches. Consumers will receive an understanding from this method about the patterns that pricing follow, and it will also give a projected price which they can see before purchasing airline tickets which helps in saving money.

Key words:- *airline price, pricing model, historical data, data visualization, optimal time, data analysis*.

---

1. Introduction

Flight ticket prices are notorious for their unpredictable fluctuations, influenced by an array of factors such as flight timing, destination, and duration. However, with the advent of machine learning algorithms, a predictive model can be created using historical flight data to help consumers determine the optimal time to purchase tickets.
The lack of information available to buyers about future price movements makes it challenging to determine the best timing for airline ticket purchasing. This project aims to uncover underlying trends of flight prices in India using historical data, and suggest the best time to buy a flight ticket.

Through a comparison study among various models, the project aims to implement validations or contradictions towards myths regarding the airline industry, while also determining the amount that can be saved by purchasing tickets at the optimal time
.
Interestingly, flight prices are highly sensitive to various factors, such as the route, month of departure, day of departure, time of departure, and airline carrier. For instance, highly competitive routes like Mumbai-Delhi had a non-decreasing trend where prices increased as days to departure decreased, while other routes like Delhi-Guwahati had a specific time frame where prices were at a minimum.

Additionally, the data uncovered two categories of airline carriers operating in India - the economical group and the luxurious group, with the minimum priced flight often belonging to the former. The data also validated that certain time-periods of the day were expected to have higher prices.
This project has the potential to be extended across various routes to make significant savings on the purchase of flight prices in the Indian Domestic Airline market.

Related work and Methodology:-

Predicting air ticket prices is a complex task due to the dynamic nature of the factors involved in pricing, which leads to fluctuating prices. However, over the past decade, researchers have started using machine learning algorithms and data mining strategies to create better models that accurately reflect observed prices.

Regression models such as Linear Regression (LR), Support Vector Machines (SVMs), and Random Forests (RF) are popularly utilized to predict airfare prices with accuracy.

Since 2017, there has been an inclination towards adopting more sophisticated machine learning models to enhance the accuracy of airfare price prediction.

Flight ticket data is generally unstructured and not instantly suited for analysis, which demands substantial work to acquire and handle the data. To assess their models' effectiveness on multiple datasets, researchers commonly gather data either by scanning it from online or by obtaining personal info from collaborative groups. This makes it hard to duplicate studies and evaluate the performance of various models.

Our suggested system may effectively handle the problem of flight price prediction utilising just publically accessible data sources with minimum characteristics. Unlike prior work, our methodology is not tied to a single market sector, making it suitable to estimate flight costs for any market..

In order to develop the airline ticket pricing model at the market segment segment level, information regarding both the airline traffic and traveller volume for each segment is required. Hence, two public datasets are employed in our suggested system. Data obtained during 2017 are utilised to train and assess the suggested model.

Method:-

Loading the datasets:-

The present study is divided into four distinct phases: (1) identifying the features that affect prices, (2) gathering enough airlines data to train and test the created machine learning models., (3) choosing the machine learning models. to be compared, and (4) conducting evaluations of the machine learning models.

Each processing step is covered in further depth in the following:

Step 1 (Feature Selection) - In this step the most informative aspects of a flight that impact the pricing of the air tickets are chosen. This step is highly significant as it specifies the issue under solution.

The present study is divided into four distinct phases: (1) identifying the features that affect prices, (2) gathering enough airlines data to train and test the created machine learning models., (3) choosing the machine learning models. to be compared, and (4) conducting evaluations of the machine learning models.

For every journey the following characteristics were considered:

1 :- leaving time.

2 :- time difference.

3 :- quantity of free baggage.

4 :- days remaining until departure.

5 :- Total number of stops.

6 :- Festival.

7 :- night travel.

8 : - weekdays.

Stage 2 (Collection of required data) - In this research, our attention is concentrated on the estimation of a single flight price. For the purpose of the testing a collection of airplanes with the same destinations (from mumbai to delhi) over the time between March and August, is gathered. During each trip the eight characteristics 1 to 8 were carefully obtained from online, 2012 trips were recorded entirely.

Stage 3 (Machine learning Model Selection) - various ml models were chosen for the present investigation and subjected to the same dataset of flights. The Machine learning techniques compared in this paper are the following:

o Random Forest Regression.

o Regression Tree.

o Linear Regression (LR) .

Step 4 (Evaluation) - The 1912 flights gathered in phase 2, were utilised in 8-fold cross-validation approach to train the above ML models. The performance metrics used to evaluate the algorithms are accuracy of prediction(% - Mean Squared Error between the intended and projected prices) and the duration in seconds, needed for training each model.

.

Architecture:-

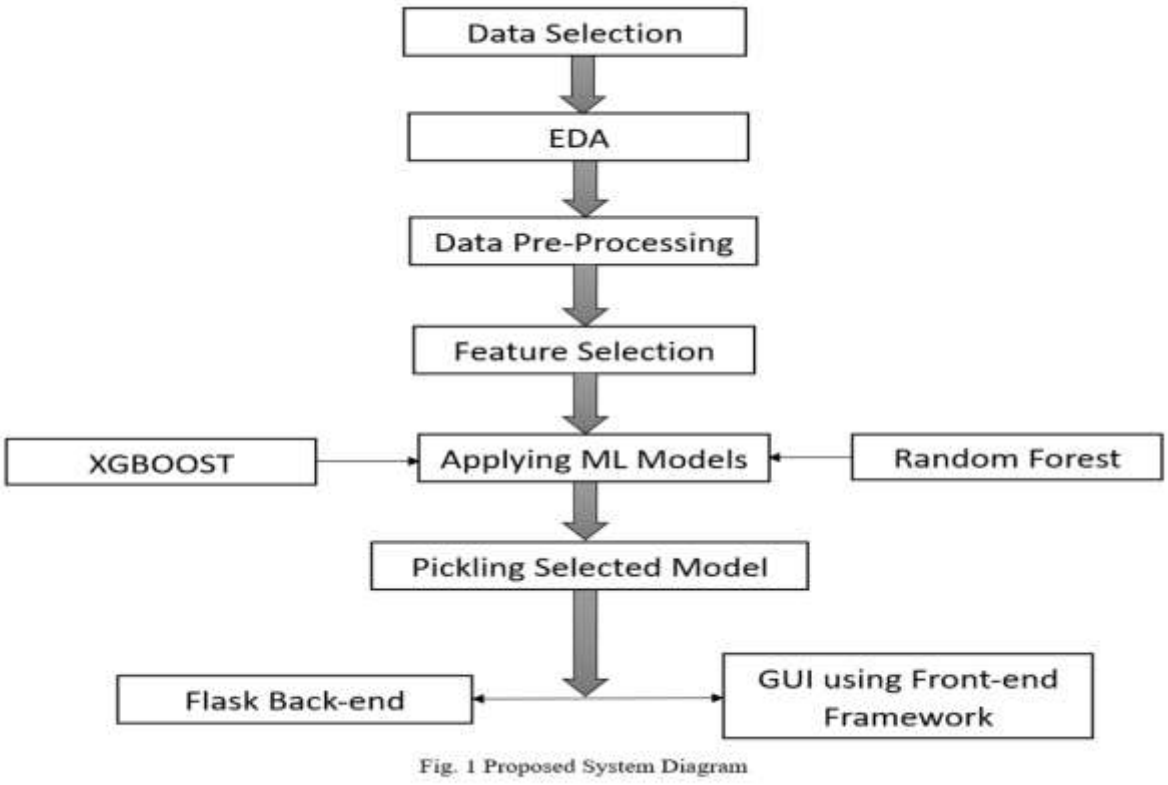
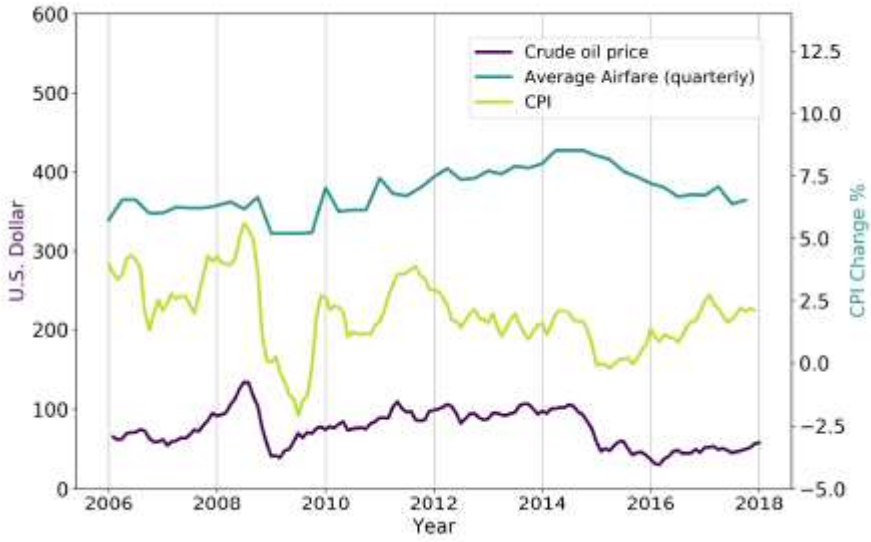


Fig. 1 Proposed System Diagram

Fig 2. Architecture of CNN



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. The random forest algorithm was utilized, resulting in an accuracy of 99%.

REFERENCES:-

[1] https://www.kaggle.com/datasets/nikhilmittal/flight-fareprediction-mh?select=Test_set.xlsx

[2] https://www.tutorialspoint.com/google_colab/index. htm

[3] https://www.kaggle.com/datasets/nikhilmittal/flight-fare-prediction-mh