**Optimizing Flight Booking Decisions through Machine Learning Price Predictions**

Abstract: Those who frequently fly for work will be more knowledgeable about the greatest deals and the best times to purchase tickets. Many airline firms adjust their fares in accordance with the seasons or time periods for commercial reasons. When more people go, the cost will go up. Data for the route is gathered using features like Duration, Source, Destination, Arrival, and Departure to estimate the highest airline fares. Features are gathered from a chosen dataset and are included in the cost, where the price of an airline ticket changes over time. We used KNN, decision trees, and random forests to implement flight price prediction for users. For estimating the cost of a flight, Random Forest has the highest accuracy (80%). Moreover, we conducted correlation analyses and metrics.. Nowadays, the majority of ancillary pricing decisions are made manually by analysts using historical data analysis and benchmarking against competitors. Merchandising rules allow ancillary pricing to be further customised to the specifics of the ancillary request after they have been manually calculated and entered in ATPCO (Airline Tariff Publishing Company) or Merchandising systems. We developed a random forest machine algorithm using airline ancillary and itinerary data that can comprehend the complex relationships between numerous attributes, such as passenger type, itinerary, aircraft type, ancillary product, or season, and can automatically determine pricing based on science.

***Key words: Random forest, flight price prediction, KNN, decision trees.***

**Introduction:**

**Data mining is a science which allow business to take data-driven decision from pattern and trends extracted from information.**

**Any individual who has booked a flight ticket previously knows how dynamically costs change. Aircraft uses advanced strategies called Revenue Management to execute a distinctive valuing strategy 1. The least expensive accessible ticket changes over a period the cost of a ticket might be high or low. This valuing method naturally modifies the toll as per the time like morning, afternoon or night. Cost may likewise change with the seasons like winter, summer and celebration seasons. The extreme goal of the carrier is to build its income yet on the opposite side purchaser is searching at the least expensive cost. Purchasers generally endeavor to purchase the ticket in advance to the takeoff day. Since they trust that airfare will be most likely high when the date of buying a ticket is closer to the takeoff date, yet it is not generally true. Purchaser may finish up with the paying more than they ought to for a similar seat.**

**A report says Indias affable aeronautics industry is on a high- development movement. India is the third-biggest avionics showcase in 2020 and the biggest by 2030. Indian air traffic is normal to cross the quantity of 100 million travelers by 2017, whereas there were just 81 million passengers in 2015. Agreeing to Google, the expression Cheap Air Tickets is most sought in India. At the point when the white collar class of India is presented to air travel, buyers searching at modest costs. The rate of flight tickets at the least cost is continuously expanding.**

**A2 - Specify the business problem**

* business problem related to flight price prediction is to improve customer satisfaction by providing personalized pricing and offers to customers. Airlines can use customer data, such as past purchase history, loyalty program status, and demographic information, to develop personalized pricing and offers that can attract and retain customers. Machine learning algorithms can assist in analyzing customer data and developing personalized pricing models that can improve customer satisfaction and loyalty….

**A3 - Business requirement**

* Accurate Flight Price Prediction: The primary business requirement is to develop a reliable and accurate flight price prediction model that can predict prices with a high degree of accuracy. This model should be able to incorporate various factors that affect pricing, including demand, competition, seasonality, and customer behaviour

**A2 - Literature Surve**

**In the research work done by Murenzi et.al,**  they visualize data and build binary classifiers based on machine and deep learning models on real data in predicting loan default probability. The important features from these models are selected and then used in the modeling process to test the stability of classifiers by comparing their performance on separate data. After analysis and visualization of data, they used different models like decision tree, random forest, logistic regression and artificial neural networks to make a real comparison of good predictors in this case.

**In the research work done by** **Neha Sakharkar ,** **Supriya Rajankar .**The value direction is bunched into gathering dependent on comparative estimating conduct. Advancement model gauge the value change designs. A treebased order calculation used to choose the best coordinating group and afterward comparing the advancement model.

**Randomforest-prediction:**

A period, in stowing the random vector is created as the includes in N boxes where N is the number of models in the preparation set of information. In random split, choice includes various autonomous random whole numbers between 1 to K. The dimensionality and nature of theata rely upon its utilization in the development of a tree. After countless trees are created, they select the most famous class.

**PRICE PREDICTION**:

Data preparation is followed by analyzing the data, uncovering the hidden trends and then applying various machine learning models. Also, some features can be calculated from the existing feature. Days to departure can be obtained by calculating the difference between the departure date and the date on which data is taken. This parameter is considered to be within 45 days. Also, the day of departure plays an important role in whether it is holiday or weekday. Intuitively the flights scheduled during weekends have a more price compared to the flights on Wednesday or Thursday. Similarly, time also seems to play an important factor. So the time is been divided into four categories: Morning, afternoon, evening, night.

To develop the model for the flight price prediction, many conventional machine learning algorithms are evaluated. They are as follows: Linear regression, Decision tree[8], Random Forest Algorithm[9], K-Nearest neighbors[7], Multilayer Perceptron[10], Support Vector Machine (SVM) [11]and Gradient Boosting. All these models are implemented in the scikit learn. To evaluate the performance of this model, certain parameters are considered.

**K-Nearest Neighbours:**

In regression techniques, the output obtained is an average value of its k nearest neighbors. It is a non-parametric method like SVM. Using some values, results are evaluated and the best performance value is obtained.

**Decision tree-prediction:**

There are two primary characteristics in the decision tree calculation. One is Information Gain and another is the Gini index. Information Gain is the proportion of Change in entropy. Higher the entropy more the instructive substance, where the entropy is a proportion of vulnerability of arbitrary variable

**Activity 4 : IMPACT**

Better Customer Experience: Personalized pricing and offers can enhance the customer experience, by providing customers with relevant and timely information about flights, pricing, and promotions. This can improve customer satisfaction and loyalty, and create a positive impact on the airline's brand reputation.

Enhanced Competitive Advantage: Personalized pricing and offers can help airlines differentiate themselves from competitors, and create a competitive advantage. By providing customers with tailored pricing and promotions, airlines can create a more personalized customer experience, which can lead to increased customer loyalty and retention.