**FLIGHT TICKET PRICE PREDICTION**

**INTRODUCTION**

As domestic air travel, is getting more and more popular these days in India. With various air ticket booking channels coming up online, travellers are trying to understand how these airline companies make decisions regarding ticket prices over time. Nowadays, airline corporations are using complex strategies and methods to assign airfare prices in a dynamic fashion. These strategies are taking into consideration several financial, marketing, commercial and social factors are closely connected with the ultimate airfare prices.

Due to the high complexity of the pricing models applied by the airlines, it is very difficult for a customer to purchase an air ticket at the lowest price, since the price changes dynamically. For this reason, several techniques ready to provide the proper time to the customer to buy an air ticket by predicting the airfare price, are proposed recently. The majority of those methods are making use of sophisticated prediction models from the computational intelligence research field known as Machine Learning (ML).

In this machine learning in python project there is only one module namely, User. User can login with valid credentials in order to access the web application. A traveller can access this module to get the future price prediction of individual airlines. The prediction will help a traveller to decide a specific airline as per his/her budget. Single entries of current or previous data can be made. This training set is used to train the algorithm for accurate predictions.

**EXISTING SYSTEM**

Flight ticket prices can be something hard to guess, today we might see a price, check out the price of the same flight tomorrow, it will be a different story. We might have often heard travellers saying that flight ticket prices are so unpredictable. So it will be very difficult for customer to choose a flight between various company. For this he/she needs to search a lot in various sites and compare prices to make a decision. This is a lot of time consuming process. At this point the importance of our project arises.

**PREPOSED SYSTEM**

Due to the high complexity of the pricing models applied by the airlines, it is very difficult a customer to purchase an air ticket in the lowest price, since the price changes dynamically. In our project the User can login with valid credentials in order to access the web application. A traveller can access this module to get the future price prediction of individual airlines. The prediction will help a traveller to decide a specific airline as per his/her budget. Single entries of current or previous data can be made. This training set is used to train the algorithm for accurate predictions.

**DATASETS**

We will be using two datasets — Train data and Test data

Training data is combination of both categorical and numerical also we can see some special character also being used because of which we have to do data Transformation on it before applying it to our model.

The dataset consists of data collected from various sources and includes the following features.

* + **Airline**: The name of the airline
  + **Date\_of\_Journey**: The date of the journey
  + **Source**: The source from which the service begins.
  + **Destination**: The destination where the service ends.
  + **Route**: The route taken by the flight to reach the destination.
  + **Dep\_Time**: The time when the journey starts from the source.
  + **Arrival\_Time**: Time of arrival at the destination.
  + **Duration**: Total duration of the flight.
  + **Total\_Stops**: Total stops between the source and destination.
  + **Additional\_Info**: Additional information about the flight.
  + **Price**: The price of the ticket.

In this study, the focus is only on minimizing the airfare charges so a single route is considered without return. This data is collected for one of the busiest routes in India over a period of time. For each flight data with all the features collected manually.

**Cleaning and preparing data**

All the collected data needed a lot of work so after the collection of data, it is needed to be clean and prepare according to the model requirements. All the unnecessary data is removed like duplicates and null values. In all machine learning this technology, this is the most important and time consuming step. Various statistical techniques and logic built-in python are used to clean and prepare the data. For example, the price was character type, not an integer.

**Analysing data**

Data preparation is followed by analysing 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 being divided into four categories: Morning, afternoon, evening, night.

**SYSTEM SPECIFICATIONS**

A **System Requirements Specification (SRS)** (also known as a Software Requirements Specification) is a document or set of documentation that describes the features and behaviour of a system or software application. It includes a variety of elements that attempts to define the intended functionality required by the customer to satisfy their different users.

**HARDWARE REQUIREMENTS**

Computer **hardware specifications** are technical descriptions of the computer's components and capabilities. Processor speed, model and manufacturer. Processor speed is typically indicated in gigahertz (GHz). ...

Processor: intel core i3 or higher

RAM: 4GB

Hard Disk: 500GB

Memory: 5GB min

**SOFTWARE REQUIREMENTS**

Operating System: Windows 10

Front End: Tkinter

Back End: python

Technology: Machine Learning

Web Technologies: HTML

Database: MySQL

IDLE: Pycharm

**MODULES**

**IMPORTING DATASETS**

1. Since data is in form of excel file we have to use pandas read-excel to load the data.
2. After loading it is important to check the complete information of data as it can indication many of hidden information such as

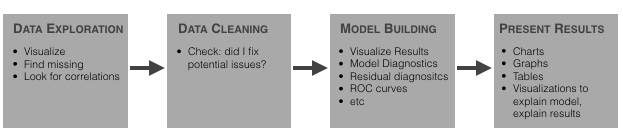
Null values in a column or row.

1. Check whether any null values are there or not. if it is present then the following can be done.
2. Imputing data using imputation method in sklearn.
3. Filling NaN values with mean, median and mode using fillna() method.
4. Describe data🡪 which can give statistical analysis.

**EDA (Exploratory Data Analysis)**

Exploratory Data Analysis, or EDA, is essentially a type of storytelling for statisticians. It allows us to uncover patterns and insights, often with visual methods, within data. EDA is often the first step of the data modelling process. In this phase, data engineers have some questions in hand and try to validate those questions by performing EDA.

EDA may sound exotic if you are new to the world of statistics. However, it’s not actually very difficult to perform an EDA. Nor do you need to know special languages to do it. As this article shows, you can use Python to do an EDA at any point in the Machine Learning (ML) process:



**FEATURE SELECTION**

Finding out the best features which will contribute and have good relation with target variable.

Following are some of the feature selection methods.

1. heatmap
2. feature\_importance\_
3. SelectKBest

**CONCLUSION**

In the proposed paper the overall survey for the dynamic price changes in the flight tickets is presented. this gives the information about the highs and lows in the airfares according to the days, weekend and time of the day that is morning, evening and night. also the machine learning models in the computational intelligence field that are evaluated before on different datasets are studied. their accuracy and performances are evaluated and compared in order to get better result.

For the prediction of the ticket prices perfectly different prediction models are tested for the better prediction accuracy. As the pricing models of the company are developed in order to maximize the revenue management. So to get result with maximum accuracy regression analysis is used. From the studies, the feature that influences the prices of the ticket are to be considered. In future the details about number of available seats can improve the performance of the model.