GLA UNIVERTSITY,

MATHURA

TOPIC: MINI PROJECT SYNOPSIS ON

Flight Price Prediction

Submitted by:

Name: Satyam Tiwari
Id: 191500733

Name: Shaurya Gupta
Id: 191500751

Name: Pranav Pandey
Id: 191500566

Submitted to:

Faculty Name: Mr. Abhishek Tiwari

DECLARATION

We hereby declare that this project work entitled “Flight Price Prediction” has been prepared by our team during 2021-2022 under the guidance of **Mr. Abhishek Tiwari ,Trainer, Training and develpoment GLA UNIVERSITY, MATHURA.** In the partial fulfillment of B.Tech degree prescribed by the college.

I also declare that this project is the outcome of effort of our team that it has not been submitted to any other university, college or any other institute for the award of any degree.

TEAM DETAILS

NAME	UNIVERSITY ROLL NO.
Satyam Tiwari	**191500733**
Shaurya Gupta	**191500751**
Pranav Pandey	**191500566**

Acknowledgment

We would like to take this opportunity to express our profound gratitude and deep regard to Mr. Abhishek Tiwari, for his exemplary guidance, valuable feedback and constant encouragement throughout the duration of the project. His valuable suggestions were of immense help throughout our project work. His perceptive criticism kept us working to make this project in a much better way. Working under him was an extremely knowledgeable experience for us.

The problem statement

Flight ticket prices can be something hard to guess, today we might see a price, check out the price of the same flight tomorrow, and it will be a different story.

To solve this problem, we have been provided with prices of flight tickets for various airlines between various cities, using which we aim to build a model which predicts the prices of the flights using various input features.

INTRODUCTION

Airline companies use complex algorithms to calculate flight prices given various conditions present at that particular time. These methods take financial, marketing, and various social factors into account to predict flight prices.

Nowadays, the number of people using flights has increased significantly. It is difficult for airlines to maintain prices since prices change dynamically due to different conditions. That’s why we will try to use machine learning to solve this problem. This can help airlines by predicting what prices they can maintain. It can also help customers to predict future flight prices and plan their journey accordingly.

ABOUT THE PROJECT:-

Anyone who has booked a flight ticket knows how unexpectedly the prices vary. Airlines use using sophisticated quasi-academic tactics known as "revenue management" or "yield management". The cheapest available ticket for a given date gets more or less expensive over time.

This usually happens as an attempt to maximize revenue based on –

1. Time of purchase patterns (making sure last-minute purchases are expensive)
2. Keeping the flight as full as they want it (raising prices on a flight which is filling up in order to reduce sales and hold back inventory for those expensive last-minute expensive purchases)

So, if we could inform the travellers with the optimal time to buy their flight tickets based on the historic data and also show them various trends in the airline industry we could help them save money on their travels. This would be a practical implementation of a data analysis, statistics and machine learning techniques to solve a daily problem faced by travellers.

The objectives of the project can broadly be laid down by the following questions –

1. Flight Trends –
   Do airfares change frequently? Do they move in small increments or in large jumps? Do tend to go up or down over time?
2. Best Time To Buy -
   What is the best time to buy so that the consumer can save the most by taking the least risk? So should a passenger wait to buy his ticket, or should he buy as early as possible?
3. Verifying Myths-

Does price increase as we get near to departure date? Is Indigo cheaper than Jet Airways? Are morning flights expensive?

The Main Objective of the Project

Automated Script to Collect Historical Data-

For any prediction/classification problem, we need historical data to work with. In this project, past flight prices for each route needs to be collected on a daily basis. Manually collecting data daily is not efficient and thus a python script was run on a remote server which collected prices daily at specific time.

Cleaning & Preparing Data-

After we have the data, we need to clean & prepare the data according to the model's requirements. In any machine learning problem, this is the step that is the most important and the most time consuming. We used various statistical techniques & logics and implemented them using built-in R packages.

Analysing & Building Models-

Data preparation is followed by analysing the data, uncovering hidden trends and then applying various predictive & classification models on the training set. These included Random Forest, Logistic Regression, Gradient Boosting and combination of these models to increase the accuracy. Further statistical models and trend analyzer model have been built to increase the accuracy of the ML algorithms for this task.

Merging Models & Accuracy Calculation-

Having built various models, we have to test the models on our testing set and calculate the savings or loss done on each query put by the user. A statistic of the over Savings, Loss and the mean saving per transaction are the measures used to calculate the Accuracy of the model implemented.

Dataset And features

The dataset used in our project is provided by Professor Gini iv from University of Minnesota. It was originally collected using daily price quotes from a major travel search web site over the period February 22, 2011 to June 23, 2011. The data were used to build a regression model for computing expected future prices and reasoning about the risk of price changes. The data source contains information of seven different routes operated by several flight companies. The features selected to use in our model include: the departure week begin, weekday of the departure, price quote date, weekday of the price quote, number of days between fetch days and the departure, and the number of stops in the itinerary.

To shed light on how the dataset looks like, Error! Reference source not found. shows a small portion of this dataset. It presents the trend of mean lowest price offered by all the airlines versus the number of days between fetchdays (purchase date) and the departure. To simplify the figure, 2 depart week begin (number of days between 1 Jan 2011 and the departure week's Monday) is set to be 128, 135, 142, and 352, respectively. The weekday of departure date is Monday.

Future Scope of the Project

●More routes can be added and the same analysis can be expanded to major airports and travel routes in India.

● The analysis can be done by increasing the data points and increasing the historical data used. That will train the model better giving better accuracies and more savings.

● More rules can be added in the Rule based learning based on our understanding of the industry, also incorporating the offer periods given by the airlines.

● Developing a more user friendly interface for various routes giving more flexibility to the users

Working Methodology of the Project

In this project, we divided the web-application in two modules first deals with building machine learning model and second module frontend. When access our web-application they will encounter by our frontend module where they will check price of different flight with their specific date and time with their location.

Details about the Hardware and the Software

***System Requirements:* - Windows 7/8/10**

Software Required:

• Technology Implemented:	Machine Learning
• Language :	Python, HTML,CSS
• IDE:	Anaconda, Jupiter,VS Code
• Browser:	Google Chrome

Hardware Requirements: -

• Processor:	Intel i3
• Operating System:	Windows 7/8/10
• RAM:	4+GB

• Hard disk:	64 GB
• Hardware Devices:	Computer System

Listing out testing technology

Language Used:-

- Python
- HTML
- CSS

Environment:-

- Anaconda
- Jupyter
- Spyder
- Vscode

What contribution would the project make and where?

This project will play major role in flight ticket booking it will have several contribution or role someof these are listed below:

1. Save time of user to check different website for ticket

2. Make ticket booking easier

3. Help costumer to compare flight ticket price

Methods: -

Since the price of the flight varies due to the difference in distance, popularity of airport, and other factors, it is hard to build up a model which performs well for all the flights. We decided to train different models for each airline route and the model trained is only applicable to its corresponding route. For the continuous model, the target variable is simply the one-day-average price. For Naïve Bayes and Softmax regression, we classified the prices into five bins using three different classification methods, and the target variable would be 1 to 5, representing each bin. For SVM, the prices are classified using equal interval.

1- Linear regression Models-

Linear regression was performed as the first attempt due to its simplicity. The four features selected in the model include the weekday of the departure date, the weekday of the quote fetch date denoted as the number of days between quote fetch date and the departure date denoted as and the number of stops during the itinerary denoted . Normal equation was used in linear regression models. Both weighted and unweighted linear regression were performed for comparison. The band width values used in weighted linear regression are 0.8, 2 and 10, respectively.

2- Native Bayes Model-

To convert the problem to a classification problem, we applied three different discretization methods: equal probability discretization, equal interval discretization, and K-means cluster. We separated the relative price into five bins and selected the same features as the ones used in linear regression. Multinomial event model of Naïve Bayes with Laplace smoothing is applied to model the trend of the price. Several tests indicate that equal interval discretization method always gives the highest accuracy. Therefore, we decided to apply this method when developing Naïve Bayes model. To parameterize the distribution of relative price, , over 5 possible outcomes, we use 5 parameters specifying the probability of each of the outcomes: Maximizing the log joint likelihood of the training set with respect to the maximum likelihood estimates:

3- SVM-

We divided prices into several bins according to their relative values compared to the overall average air ticket price. An example set of bins would be 60% to 80%, 80% to 100%, 100% to 120% and 120% to 140% (of average price) etc. Using L2-regularization and L2 loss function, we are able to achieve an accuracy of 60.54%, which is not very ideal. However, when using only two bins (higher and lower than average), we are able to achieve an accuracy of 80.6%. SVM regression has also been used as a continuous model, which does not generate satisfying results, thus discarded.

Conclusion-

This study shows that it is feasible to predict the airline ticket price based on historical data. One possible way to increase the accuracy can be combining different models after carefully studying their own performance on each individual bin. Additionally, as the learning curve indicates, adding more features will increase the accuracy of our models. However, limited by the current data source that we have, we are unable to extract more information of a particular flight. In the future, more features, such as the available seat, the departure time of a day, and whether the departure day is a holiday or not, can be added to the model to improve the performance of the predicting model.

Reference-

1. O. Etzioni, R. Tuchinda, C. A. Knoblock, and A. Yates. To buy or not to buy: mining airfare data to minimize ticket purchase price.
2. Manolis Papadakis. Predicting Airfare Prices.
3. Groves and Gini, 2011. A Regression Model For Predicting Optimal Purchase Timing For Airline Tickets.
4. Modeling of United States Airline Fares – Using the Official Airline Guide (OAG) and Airline Origin and Destination Survey (DB1B), Krishna Rama-Murthy, 2006.
5. B. S. Everitt: The Cambridge Dictionary of Statistics, Cambridge University Press, Cambridge (3rd edition, 2006). ISBN 0-521-69027-7.
6. Bishop: Pattern Recognition and Machine Learning, Springer, ISBN 0-387-31073-8.