Department of Information Technology
NBA Accredited

A.P. Shah Institute of Technology

G.B.Road,Kasarvadavli, Thane(W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic Year 2021-2022

A Project Report on

Flight Fare Prediction System

Submitted in partial fulfilment of the degree of
Bachelor of Engineering(Sem-6)
in
INFORMATION TECHNOLOGY

By
Vedant Patil(19104065)
Mudra Limbasia(19104058)
Anvit Mirjurkar(19104059)

Under the Guidance of
Prof. Kiran Deshpande

1.Project Conception and Initiation

1.1 Objectives

- To make travelling easy for the passengers.
- To provide all the necessary flight details on one platform to minimize efforts and minimize decision making time.
- To bring multiple Airlines, flights and their rates at the tip of your finger.
- To allow the passengers to book flights according to their convenience of time, rate, location, and stoppage.
- To give multiple best suited flight options to the passengers.

1.2 Literature Review

II. LITERATURE SURVEY

It is very difficult for the customer to purchase a flight ticket at the minimum price. For this several techniques are used to obtain the day at which the price of air ticket will be minimum. Most of these techniques are using sophisticated artificial intelligence(AI) research is known as Machine Learning.

Utilizing AI models, [2] connected PLSR(Partial Least Square Regression) model to acquire the greatest presentation to get the least cost of aircraft ticket buying, having 75.3% precision. Janssen [3] presented a direct quantile blended relapse model to anticipate air ticket costs for cheap tickets numerous prior days takeoff. Ren, Yuan, and Yang [4], contemplated the exhibition of Linear Regression (77.06% precision), Naive Bayes (73.06% exactness, Softmax Regression (76.84% precision) and SVM (80.6% exactness) models in anticipating air ticket costs. Papadakis [5] anticipated that the cost of the ticket drop later on, by accepting the issue as a grouping issue with the assistance of Ripple Down Rule Learner (74.5 % exactness.), Logistic Regression with 69.9% precision and Linear SVM with the (69.4% exactness) Machine Learning models.

Gini and Groves[2] took the Partial Least Square Regression(PLSR) for developing a model of predicting the best purchase time for flight tickets. The data was collected from major travel journey booking websites from 22 February 2011 to 23 June 2011. Additional data were also collected and are used to check the comparisons of the performances of the final model.

Janssen [3] built up an expectation model utilizing the Linear Quantile Blended Regression strategy for SanFrancisco to NewYork course with existing every day airfares given by www.infare.com. The model utilized two highlights including the number of days left until the takeoff date and whether the flight date is at the end of the week or weekday. The model predicts airfare well for the days that are a long way from the takeoff date, anyway for a considerable length of time close the takeoff date, the expectation isn't compelling.

Wohlfarth [15] proposed a ticket buying time enhancement model dependent on an extraordinary pre-preparing step known as macked point processors and information mining systems (arrangement and bunching) and measurable investigation strategy. This system is proposed to change over heterogeneous value arrangement information into added value arrangement direction that can be bolstered to unsupervised grouping calculation. 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.

A study by Dominguez-Menchero [16] recommends the ideal buying time dependent on nonparametric isotonic relapse method for a particular course, carriers, and timeframe. The model gives the most extreme number of days before buying a flight ticket. two sorts of the variable are considered for the expectation. One is the passage and date of procurement.

1.3 Problem Definition

- Booking flights is a tedious task as there might be numerous options on various platforms, which makes it hard and confusing to decide.
- There are abundant platforms which provide varying rates of flights according to time and stoppage of flights, navigating from one website/app to another to make the right decision is very time consuming and puzzling.

1.4 Scope

- The project is a good platform for frequent travellers.
- It is useful for all types of passengers as it is user friendly.
- It is less time consuming as suitable flights are suggested at one single platform.
- It also allows passengers to check predicted flight rates for the near future. As a result they might learn about the best time to book their flights accordingly.

1.5 Technology stack

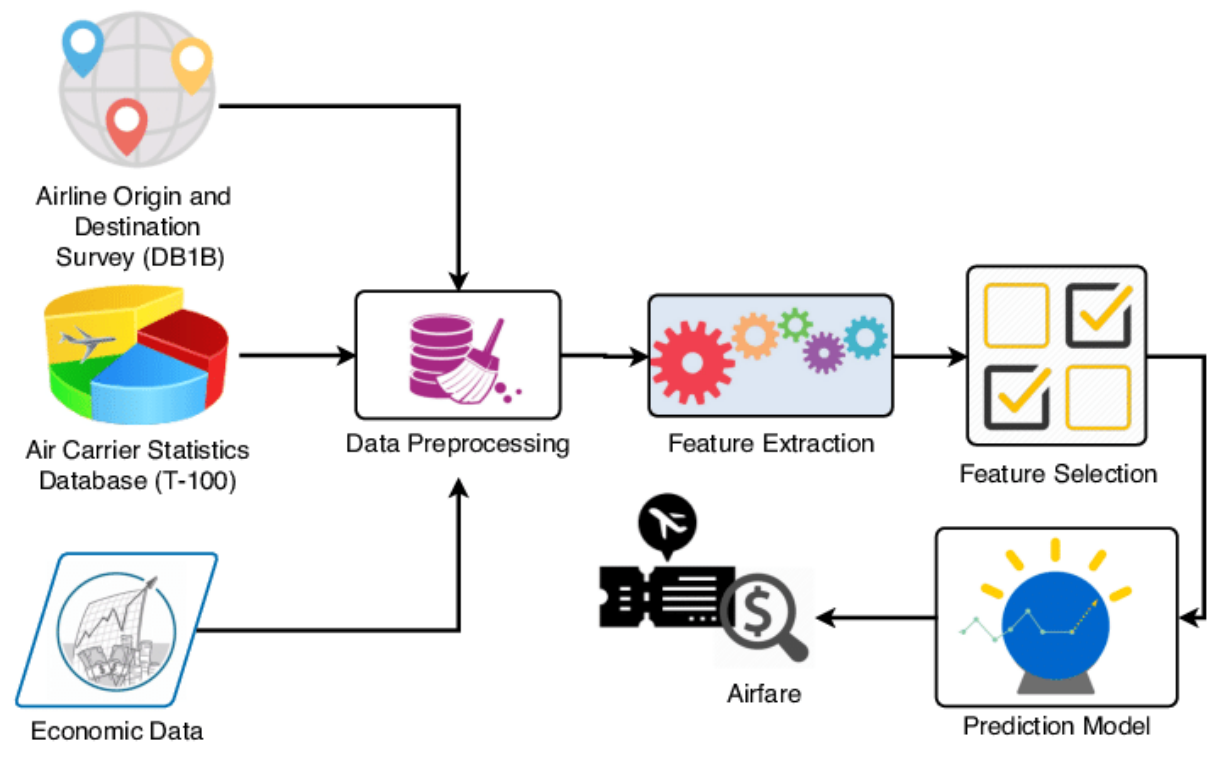
- PYTHON
- HTML, CSS
- FLASK
- JUPYTER NOTEBOOK
- GITHUB & GIT
- CODE EDITOR

2. Project Design

2.1 Proposed System

- The project allows the travellers to check flight rates, according to their travel date, destination and stoppage.
- It also allows passengers to check predicted flight rates for the near future. As a result they might learn about the best time to book their flights accordingly.
- It brings various Airline companies, and flights at one single platform, which makes it easier for the passengers to choose pocket friendly options.

2.2 Design(Flow Of Modules)



2.2 Design(Flow Of Modules)

- The first module consists of Airline origin and destination survey, air carrier statistics database and economic data. It is the research process where we have collected the required data for our project.
- Next step is data pre-processing. Where training data is pre-processed to transform the raw data into useful and efficient format.
- After pre-processing feature extraction and feature selection is done.
- The last step is to create a prediction model using AI algorithms. And finally the air fare prediction system is built.

3. Implementation

3.1 Features and Functionality

- **Flight rate prediction:** Rates of Airlines and flights are predicted as per the passengers requirements. Passengers can check flight rates some time prior so that they can book whenever the rates are feasible.
- **Flight Stoppage:** Passengers can choose to travel by taking halts or direct to their destination which may also affect their flight rates, and hence giving them options according to their convenience.

3.2 Algorithms Used

- **Random forests** or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance.

3.2 Algorithms Used

- Adding one further step of randomization yields **extremely randomized trees**, or **ExtraTrees**. While similar to ordinary random forests in that they are an ensemble of individual trees, there are two main differences: first, each tree is trained using the whole learning sample (rather than a bootstrap sample), and second, the top-down splitting in the tree learner is randomized. Instead of computing the locally optimal cut-point for each feature under consideration (based on, e.g., information gain or the Gini impurity), a random cut-point is selected.

3.2 Algorithms Used

- This value is selected from a uniform distribution within the feature's empirical range (in the tree's training set). Then, of all the randomly generated splits, the split that yields the highest score is chosen to split the node. Similar to ordinary random forests, the number of randomly selected features to be considered at each node can be specified.
- We have used **Extra tree regressor** in our project, for prediction.

5. Result

5.1 Project Outcomes

- User can choose their date of departure, and source and destination.
- User can choose stoppage of flight, and their preferred airlines to get the best match of flights available.
- User is given best suited flight options according to their travel requirements.

6. Conclusion and Future Scope

6.1 Conclusion

- Thus in conclusion, our project would help to save money of inexperienced people by providing them the information related to the trends that the flight rates follow.
- It will also give them a predicted value of the price, by which they can decide whether to book ticket immediately or at some later point of time when the cost of ticket is comparatively lower.

6.2 Future scope

- Currently there are many fields where machine learning algorithms are used to predict outcomes of certain events, such as a “Stock price prediction system”. Which gives an approximate estimate value of the stock.
- Therefore, there is a requirement for such services in the aviation industry too, which can help customers in booking tickets.
- More accurate datasets can be used to get more accurate predictions.

References

- Machine Learning Basics: Decision Tree Regression - Medium
- Flight Fare Prediction Using Machine Learning - Analytics ...
- https://medium.com/analytics-vidhya/regression-flight-price-prediction-6771fc4d1fb3

Thank You