

A Thesis/Project/Dissertation Report

on

Flight Price Prediction using AI-ML and Web

*Submitted in partial fulfillment of the
requirement for the award of the degree of*

B.Tech in Computer Science Engineering



(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

**Under The Supervision of
C Ramesh Kumar
Assistant Professor**

Submitted By

Kunal Bhardwaj
18021120010/18SCSE1120011

Kartikey Gupta
18021120003/18SCSE1120003

**SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING /
DEPARTMENT OF COMPUTERAPPLICATION
GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA**

December, 2021



**SCHOOL OF COMPUTING SCIENCE AND
ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA**

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the thesis/project/dissertation,
entitled “**Flight Price Prediction using AI-ML and Web**” in partial fulfillment of the requirements for the award of the B.Tech in Computer Science Engineering –submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of September, 2021 to December and 2021, under the supervision of C Ramesh Kumar Sir (Assistant Professor), Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

Kartikey Gupta 18SCSE1120003

Kunal Bhardwaj 18SCSE1120011

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of Kunal Bhardwaj
(18SCSE1120011) and Kartikey Gupta (18SCSE1120003)-has been held on
_____ and his/her work is recommended for the award of B.Tech. Computer
Science Engineering-

Signature of Examiner(s)

Signature of Supervisor(s)

Signature of Project Coordinator

Signature of Dean

Date: December, 2021

Place: Greater Noida

Acknowledgement

We feel great pleasure in submitting this project paper on Flight fare prediction. We wish to express true sense of gratitude towards our project guide, Dr. C Ramesh sir who at every stage in the study of this project, contributed his valuable guidance and helped to solve every and every problem. Our great obligation would remain due towards Dr. C. Ramesh Sir who was a constant inspiration during our project. He provided with an opportunity to undertake the project at Galgotias University College of Engineering, Greater Noida. I feel highly indebted to them who provided us with all our project requirements, and did much beyond our expectations to bring out the best in us.

Abstract

The cost of an airline ticket is always get changed by a number of factors, such as flight kilometers, flight time, fuel price, etc. So, in that case people are always misjudged about the price of the fare which they will be going to paid at the time of booking. Each model has its unique and authentic rules and algorithms to set the price accordingly. But now things are dependent upon the festivals and many others factors too, which is not been considered earlier.

Most current developments in Artificial Intelligence (AI) and Machine Learning (ML) make it a reality to find such rules and model cost and variation patterns.

People those who will buy flight tickets frequently shall be able to guess the right time to procure a ticket to obtain the good and best deal. All the airplane make difference in ticket cost for their money management. To cure, we will be working on different data sets for getting the optimum results. We have also added the discount prediction page to make things more enjoyable for the user.

Tools which we are using is Wamp server for database, web development technology to make things virtually attractive, Data Models and Machine learning algorithms to train our data sets and make the proper use of it.

When we talk about the scope of the future, we can say that there may be many different things to be added such as flight delays or cancellations or to make things more digital like face scanner etc.

Contents

Title		Page No.
Candidates Declaration		I
Acknowledgement		II
Abstract		III
Contents		IV
List of Figures		V
Chapter 1	Introduction	8-16
	1.1 Introduction	10
	1.2 Formulation of Problem	12
	1.2.1 Tool and Technology Used	13
Chapter 2	Literature Survey/Project Design	17-21
Chapter 3	Methodology of the Project	22-26
Chapter 4	Functionality/Working of Project	27-34
Chapter 4	Results and Discussion	35-37
Chapter 5	Conclusion and Future Scope	38-39
	5.1 Conclusion	38
	5.2 Future Scope	39
	Reference	40

List of Figures

S.No.	Title	Page No.
1	Root Node Model	14
2	Random Forest Classifier	16
3	Use Case model	19
4	Architectural Diagram of system	20
5	Class Diagram of system	21
6	List of columns present in the dataset	24
7	Data Set Columns After data cleaning	25
8	Functionality working of the project chart	26
9	Solution model	27
10	Regression Model figure	30
11	Sample Data set	32
12	Sample Testing data set	33
13	Home Page	34
14	Book Now Page	35
15	Packages Page	36
16	Box Plot for Airline VS Price	36
17	Heat Map	38
18	Importance of attributes	38
19	Box Plot for Destination VS Price	39

CHAPTER-1

Introduction

This Project is basically based upon the flight price prediction using Machine Learning and web technology. We are using web technology for the betterment of users as we are going to provide a full service as a travel agent. In our product we are going to train data and then give all the facilities to the user like gallery page, prediction page, discount prediction page, contact page etc.

The price of an airline ticket is affected by a number of factors, such as flight distance, purchasing time, fuel price, etc. So, in that case people are always misjudged about the price of the fare which they will be going to paid at the time of booking. Each carrier has its own proprietary rules and algorithms to set the price accordingly. But now things are dependent upon the festivals and many others factors too, which is not been considered earlier.

Recent advance in Artificial Intelligence (AI) and Machine Learning (ML) makes it possible to infer such rules and model the price variation.

Someone who purchase flight tickets frequently would be able to predict the right time to procure a ticket to obtain the best deal. Many airlines change ticket prices for their revenue management. To cure, we will be working on different data sets for getting the optimum results. We have also added the discount prediction page to make things more enjoyable for the user.

Tools which we are using is Wamp server for database, web development technology to make things virtually attractive, Data Models and Machine learning algorithms to train our data sets and make the proper use of it.

Talking about the future scope we can say there can be many different things added like flight delay or cancellations or making things more digitals like face scanner etc.

We have trained a random forest classifier model for predicting the price of the flight based on various factors which affect the price of the flight.

Our initial investigation shows that models on the two sides depend on restricted arrangement of highlights, for example, verifiable ticket value information, ticket buy date and flight date.

Parameters on which fares are calculated-

- Airline
- Date of Journey
- Date of Arrival
- Source
- Destination
- Departure Time
- Arrival Time
- Duration
- Total Stops
- Weekday/Weekend

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so.

Machine learning algorithms use historical data as input to predict new output values.

Machine learning is important because it gives enterprises a view of trends in customer behavior and business operational patterns, as well as supports the development of new products. Many of today's leading companies, such as Facebook, Google and Uber, make machine learning a central part of their operations. Machine learning has become a significant competitive differentiator for many companies.

Supervised learning is type of machine learning, data scientists supply algorithms with labeled training data and define the variables they want the algorithm to assess for correlations. Both the input and the output of the algorithm is specified.

Supervised machine learning requires the data scientist to train the algorithm with both labeled inputs and desired outputs. Supervised learning algorithms are good for the following tasks:

1. Binary classification: Dividing data into two categories.
2. Multi-class classification: Choosing between more than two types of answers.
3. Regression modeling: Predicting continuous values.
4. Assembling: Combining the predictions of multiple machine learning models to produce an accurate prediction.

1.2 Formulation of problem

The major problems leading to price prediction is giving the exact amount in website through ML algorithm:

- Differentiating between the weekends and festivals.
- Analysing old data sets to make things on to the point.
- User data storage server.
- Combining the ML and Web development technology.

Incomplete Information: Travelers can just access part of the carrier's inner data. Truth be told, they don't have the admittance to the key information, like the number of the excess tickets and the understanding between various carrier organizations.

Fragmented Information: The data that voyagers can get is divided. For instance, it is undeniably challenging for a normal explorer to find the relationship between flight cost and flight characters, like the quantity of visits, the takeoff time, and so on

Irregular Change: Although explorers can gather verifiable flight value, the cost change isn't smooth. In reality, it is by all accounts profoundly unpredictable. Thus, Travelers can only with significant effort anticipate future flight value as per the chronicled values

1.2.1 Tools and technologies used

We have used and worked on following technologies:

- Wamp servers: To store the query data and details of user.
- SQL: To store user and file metadata received by the User.
- HTML/CSS: For building client-side web application.
- Algorithm: For predicting the price data of the flight.
- ML + User friendly development using web technology: For building web client application. Based on Google's material guidelines and uses latest web architecture pattern for optimal performance and section development.

GITHUB:

- GitHub provides project management features.
- It is used by developers for personal projects or to do experiments with new programming languages.
- For enterprises, it provides the features of SAML single sign-on, Access provisioning, 99.95% uptime, Invoice billing, Advanced auditing, and Unified search and Contribution, etc.
- GitHub provides security features like Security incident response, and Two-factor authentication, etc.

Verdict: GitHub has functionalities for Code review, Project management, Integrations, Team management, social coding, Documentation, and Code hosting. For enterprises, it provides priority support.

Flask - Flask is basically a micro web application framework written in Python. Developers often use Flask for making web applications, HTTP request management, and template rendering. By “micro web application,” we mean that it is not a full-stack framework. It is used for making API which will connect our app to the machine learning model using POST request. Flask is a web application framework written in Python. It is developed by Armin Ronacher, who leads an international group of Python enthusiasts named Pocco. Flask is based on the Werkzeug WSGI toolkit and Jinja2 template engine. Both are Pocco projects.

Data Model for AI predictions:

Random Forest-Random Forest basically uses group of decision trees as group of models. Random amount of data is passed to decision trees and each decision tree predicts values according to the dataset given to it. From the predictions made by the decision trees the average value of the predicted values is considered as the output of the random forest model. Random forest basically uses group of decision trees as group of models. Random amount of data is passed to decision trees and each decision tree predicts values according to the dataset given to it.

It is a machine learning algorithm which we are using to build the project it will take various flight detail from user as input.

A random forest is a machine learning technique that's used to solve regression and classification problems. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems.

A random forest algorithm consists of many decision trees. The 'forest' generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms.

The (random forest) algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome.

A random forest eradicates the limitations of a decision tree algorithm. It reduces the overfitting of datasets and increases precision. It generates predictions without requiring many configurations in packages (like scikit-learn).

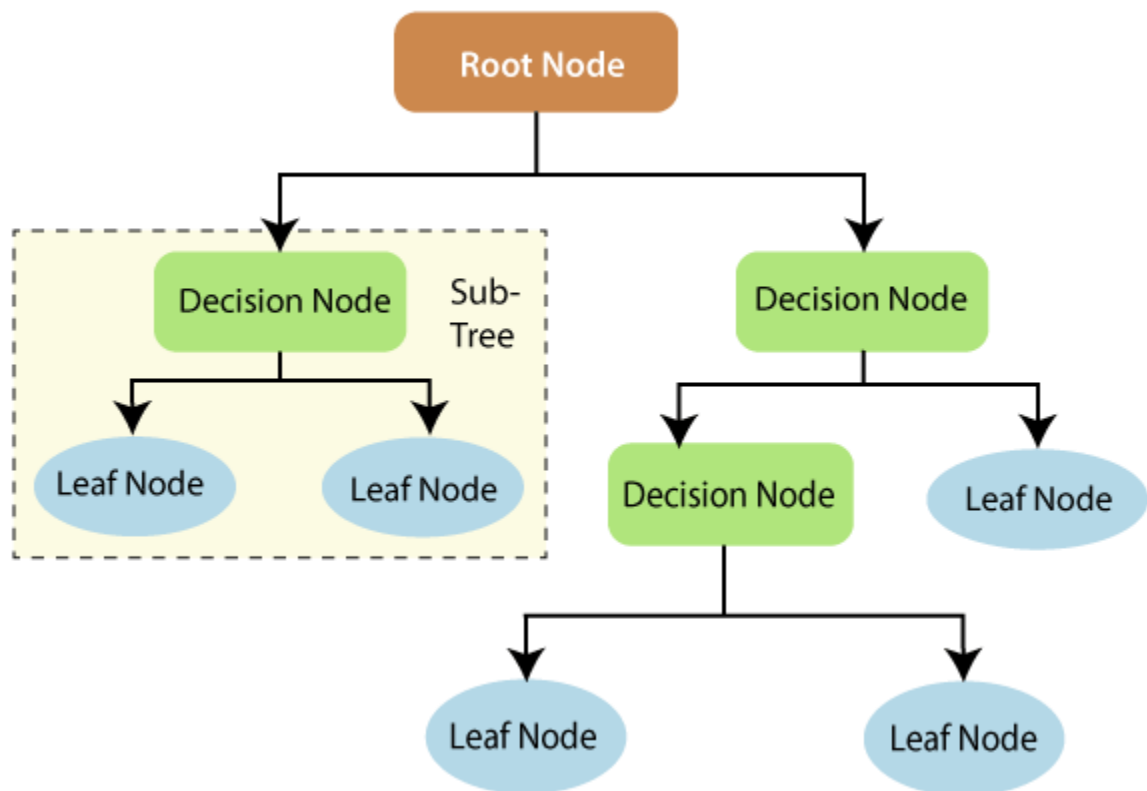
Features of a Random Forest Algorithm

- It's more accurate than the decision tree algorithm.
- It provides an effective way of handling missing data.
- It can produce a reasonable prediction without hyper-parameter tuning.
- It solves the issue of overfitting in decision trees.
- In every random forest tree, a subset of features is selected randomly at the node's splitting point.
- How random forest algorithm works
- Understanding decision trees

Decision trees are the building blocks of a random forest algorithm. A decision tree is a decision support technique that forms a tree-like structure. An overview of decision trees will help us understand how random forest algorithms work.

A decision tree consists of three components: decision nodes, leaf nodes, and a root node. A decision tree algorithm divides a training dataset into branches, which further segregate into other branches. This sequence continues until a leaf node is attained. The leaf node cannot be segregated further.

The nodes in the decision tree represent attributes that are used for predicting the outcome. Decision nodes provide a link to the leaves. The following diagram shows the three types of nodes in a decision tree.



Fig(1) Root Node Model

The information theory can provide more information on how decision trees work. Entropy and information gain are the building blocks of decision trees. An overview of these fundamental concepts will improve our understanding of how decision trees are built.

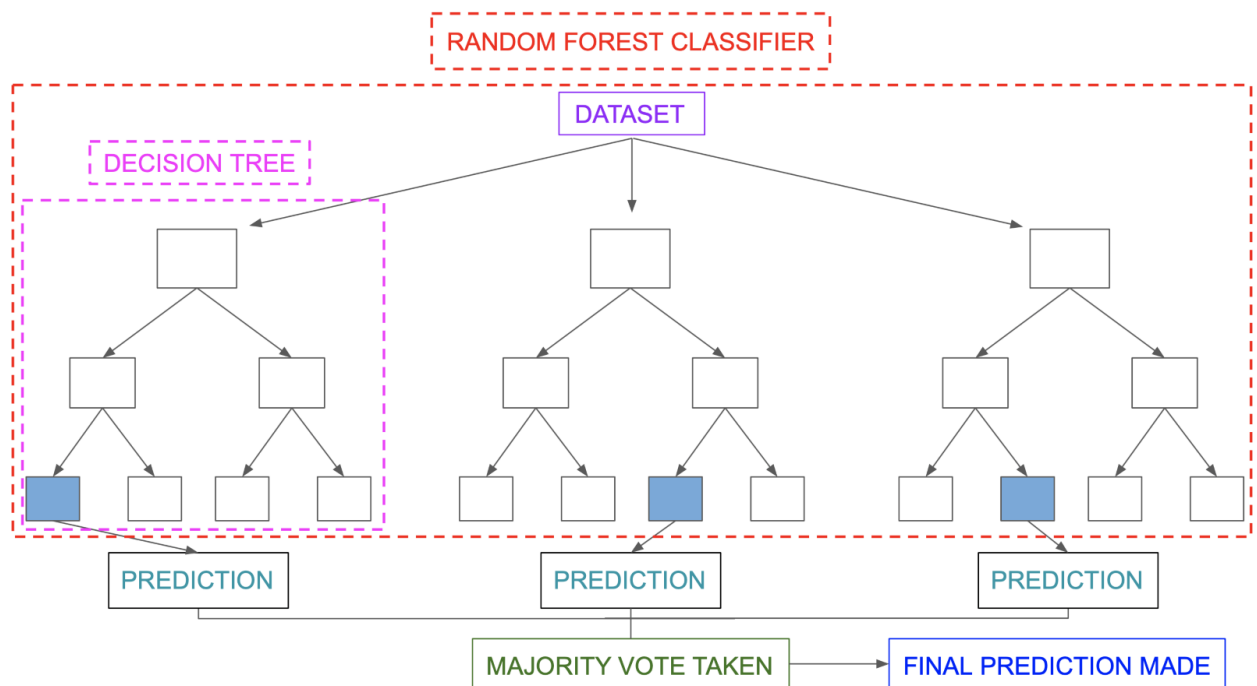
Entropy is a metric for calculating uncertainty. Information gain is a measure of how uncertainty in the target variable is reduced, given a set of independent variables.

The information gain concept involves using independent variables (features) to gain information about a target variable (class). The entropy of the target variable (Y) and the conditional entropy of Y (given X) are used to estimate the information gain. In this case, the conditional entropy is subtracted from the entropy of Y.

Information gain is used in the training of decision trees. It helps in reducing uncertainty in these trees. A high information gain means that a high degree of uncertainty (information entropy) has been removed. Entropy and information gain are important in splitting branches, which is an important activity in the construction of decision trees.

Classification in random forests employs an ensemble methodology to attain the outcome. The training data is fed to train various decision trees. This dataset consists of observations and features that will be selected randomly during the splitting of nodes.

A rain forest system relies on various decision trees. Every decision tree consists of decision nodes, leaf nodes, and a root node. The leaf node of each tree is the final output produced by that specific decision tree. The selection of the final output follows the majority-voting system. In this case, the output chosen by the majority of the decision trees becomes the final output of the rain forest system. The diagram below shows a simple random forest classifier.



Fig(2) Random Forest Classifier

Advantages of random forest are: -

- It can perform both regression and classification tasks.
- A random forest produces good predictions that can be understood easily.
- It can handle large datasets efficiently.
- The random forest algorithm provides a higher level of accuracy in predicting outcomes over the decision tree algorithm.

CHAPTER-2

Literature Survey/Project Design

There are many different people of different domain who have worked over this idea of predicting flight price to make things sorted. It is hard for the client to buy an air ticket at the most reduced cost. For these few procedures are explored to determine time and date to grab air tickets with minimum fare rate.

The majority of these systems are utilizing the modern computerized system known as Machine Learning. To determine ideal purchase time for flight ticket Gini and Groves exploited Partial Least Square Regression (PLSR) for building up a model. The information was gathered from major travel adventure booking sites from 22 February 2011 to 23 June 2011. Extra information was additionally gathered and are utilized to check the correlations of the exhibitions of the last model.

Janssen implemented a desire model using the Linear Quantile Blended Regression methodology for San Francisco–New York course where each day airfares. Two features such as number of days for departure and whether departure is on weekend or weekday are considered to develop the model. The model guesses airfare well in advance from the departure date. But the model isn't convincing in a situation for an extensive time allotment, it closes the departure date.

Wohlfarth proposed a ticket purchasing time improvement model subject to a significant pre-processing known as masked point processors, data mining frameworks (course of action and grouping) and quantifiable examination system. This framework is proposed to change various added value arrangements into included added value arrangement heading which can support to solo gathering estimation. This value heading is packed into get-together reliant on near evaluating conduct. Headway model measure the value change plans. A tree-based analysis used to pick the best planning gathering and a short time later looking at the progression model.

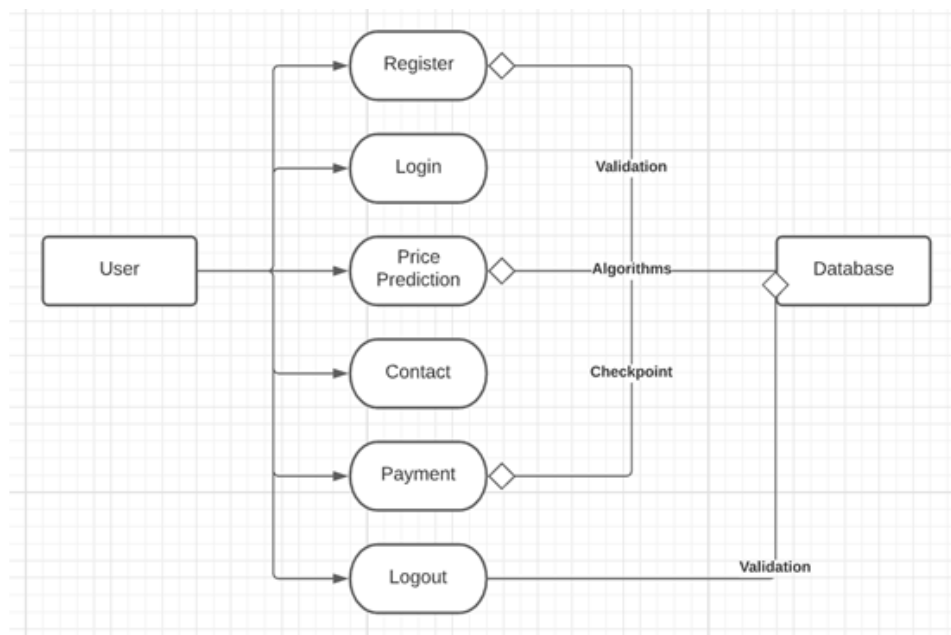
An investigation by Dominguez-Mencher suggests the perfect purchase timing reliant on a nonparametric isotonic backslide technique for a specific course, carriers, and time frame. The model provides the most acceptable number of days before buying the flight ticket. The model considers two types of a variable such as the entry and is date of obtainment.

Project Design Approach

Use Case Diagram

A use case diagram is used to represent the dynamic behavior of a system. It encapsulates the system's functionality by incorporating use cases, actors, and their relationships. It models the tasks, services, and functions required by a system/subsystem of an application. It depicts the high-level functionality of a system and also tells how the user handles a system.

In our Use Case Diagram all the entities are been used and explained also the results of each is also been depicted from it.

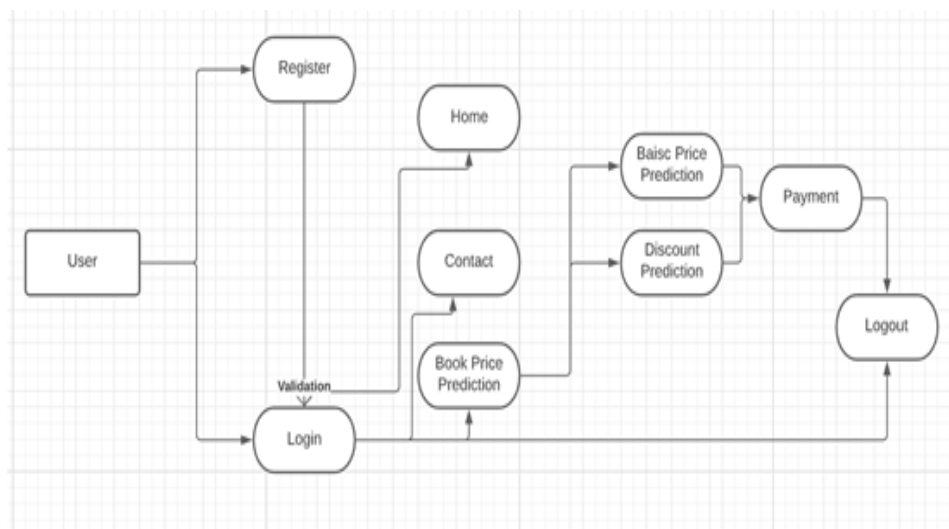


Fig(3) Use Case model

Architectural Diagram of system

Architecture as 'the practice of the architect' refers to planning, designing and constructing form, space and ambience. It extends from urban design and master planning to building design, the design of individual spaces and even fixtures and fittings.

Our figure gives the clean meaning of the working of our project and also the working of it and all the modules are also covered in the below figure.



Fig(4) Architectural Diagram of system

Class Diagram of system

The purpose of class diagram is to model the static view of an application. Class diagrams are the only diagrams which can be directly mapped with object-oriented languages and thus widely used at the time of construction.

UML diagrams like activity diagram, sequence diagram can only give the sequence flow of the application, however class diagram is a bit different. It is the most popular UML diagram in the coder community.

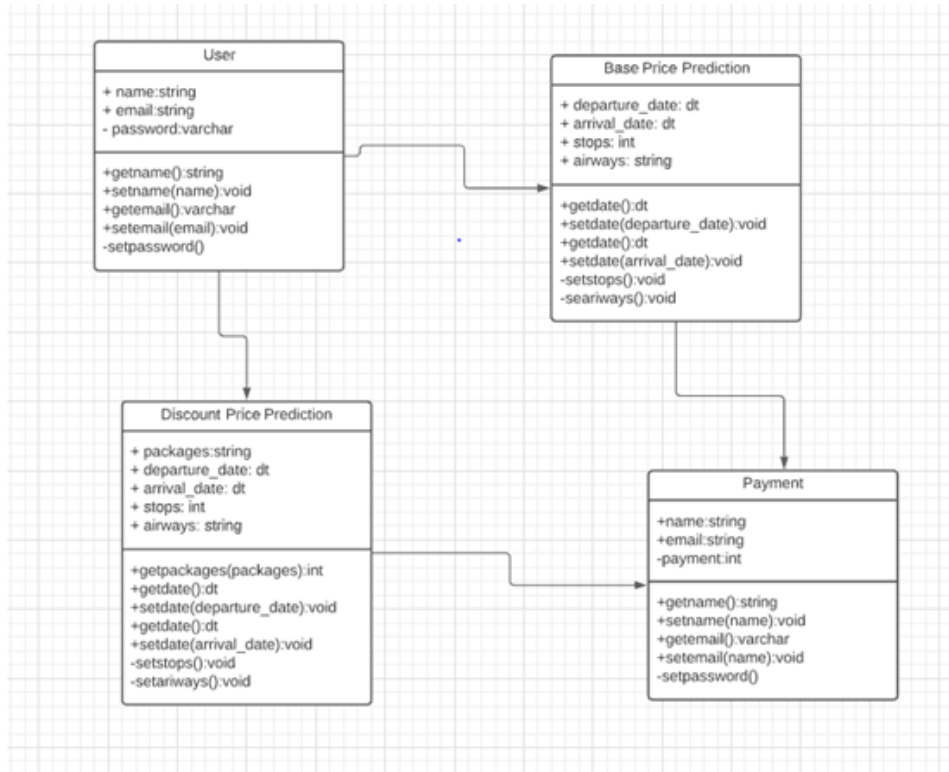


Fig (5) Class Diagram of system

Chapter-3

Methodology

Proposed System

To develop a mobile application for “Airfare Prediction” based on previous airline ticket sales dataset for improving sales in Indian Domestic Airline. Our main motive is to provide the client with a prediction system from which it can take a right decision of increasing or decreasing the Airfare so that the flight doesn’t go empty or no money is lost due to sudden increase in crude oil.

- a. To perform data analytics on customer’s ticket booking data for a brief amount of time.
- b. To refine the data i.e., Removing duplicate records, ambiguity etc.
- c. To perform Feature engineering in order to extract important feature from dataset for prediction.
- d. To Brainstorm the Features i.e., to decide how to use those features
- e. To create features i.e., to derive new features from those useful features.

The contribution of the proposed system includes the following activities:

1) Airfare prices prediction in India for domestic airline

The dataset contains 45 different columns from which we are extracting those columns (features) which will be used to train the model and predict the given goal.

2) Investigation and analysis of the features that affects the airfare.

The proposed system will use Data analytics for completely exploring the data, identifying relationship among those columns, finding some patterns of work, cleaning and refining the dataset to reduce complexity of data. Then the system will extract useful features from the bulky dataset and also will derive new features for making system processing easier.

3) Performance analysis of the ML models.

Random forest regression, Support vector machine, K-nearest neighbors Regression etc.

Machine Learning algorithms will be used to train the model. Since the airfare is continuous value therefore Regression techniques will be used. At the end the system will generate report for “AIRFARE PREDICTION”.

The proposed system is composed of four phases:

1. Data input
2. Feature extraction
3. Machine learning model selection
4. Prediction

Phase 1: Data Input

The input data file in .csv file will be provided to system and that input file contains all customer ticket booking information. The training data contains 45 columns from which important features are extracted. The information is limited to domestic airline.

```
Data columns (total 11 columns):
# Column Non-Null Count Dtype
---
0 Airline 10683 non-null object
1 Date_of_Journey 10683 non-null object
2 Source 10683 non-null object
3 Destination 10683 non-null object
4 Route 10682 non-null object
5 Dep_Time 10683 non-null object
6 Arrival_Time 10683 non-null object
7 Duration 10683 non-null object
8 Total_Stops 10682 non-null object
9 Additional_Info 10683 non-null object
10 Price 10683 non-null int64
```

Fig 6 List of columns present in the dataset

Phase 2: Data Cleaning

According to the research 80% of the work is done in cleaning data and retrieving useful information from it. As the data is collected from live public domain i.e. Airline industry it contains many Null values, redundant entries, merged values, referential features and many unnecessary columns. Data cleaning steps are as follows :

1. Removing null values
2. Formatting of date columns
3. Removing outliers
4. Conversion of object, string and other data types into numeric form (Encoding).

Fig 5 List of columns present in the dataset

```
Data columns (total 30 columns):
```

#	Column	Non-Null	Count	Dtype
0	Total_Stops	10682	non-null	int64
1	Price	10682	non-null	int64
2	Journey_day	10682	non-null	int64
3	Journey_month	10682	non-null	int64
4	Dep_hour	10682	non-null	int64
5	Dep_min	10682	non-null	int64
6	Arrival_hour	10682	non-null	int64
7	Arrival_min	10682	non-null	int64
8	Duration_hours	10682	non-null	int64
9	Duration_mins	10682	non-null	int64
10	Airline_Air India	10682	non-null	uint8
11	Airline_GoAir	10682	non-null	uint8
12	Airline_IndiGo	10682	non-null	uint8
13	Airline_Jet Airways	10682	non-null	uint8
14	Airline_Jet Airways Business	10682	non-null	uint8
15	Airline_Multiple carriers	10682	non-null	uint8
16	Airline_Multiple carriers Premium economy	10682	non-null	uint8
17	Airline_SpiceJet	10682	non-null	uint8
18	Airline_Trujet	10682	non-null	uint8
19	Airline_Vistara	10682	non-null	uint8
20	Airline_Vistara Premium economy	10682	non-null	uint8
21	Source_Chennai	10682	non-null	uint8
22	Source_Delhi	10682	non-null	uint8
23	Source_Kolkata	10682	non-null	uint8
24	Source_Mumbai	10682	non-null	uint8
25	Destination_Cochin	10682	non-null	uint8
26	Destination_Delhi	10682	non-null	uint8
27	Destination_Hyderabad	10682	non-null	uint8
28	Destination_Kolkata	10682	non-null	uint8
29	Destination_New Delhi	10682	non-null	uint8

Fig 7. Data Set Columns After data cleaning

Phase 3: Feature Extraction

During this phase most of the informative features from the airline dataset that determines the prices of the air tickets are extracted. Features that can be considered are as follows:

Feature 1: Booking date and time

Feature 2: Departure date and time

Feature 3: Numbers of days till flight departure

Feature 4: Category of passenger (Adult/Child)

Feature 4: Cabin (Economy/Business)

Feature 5: Source Location

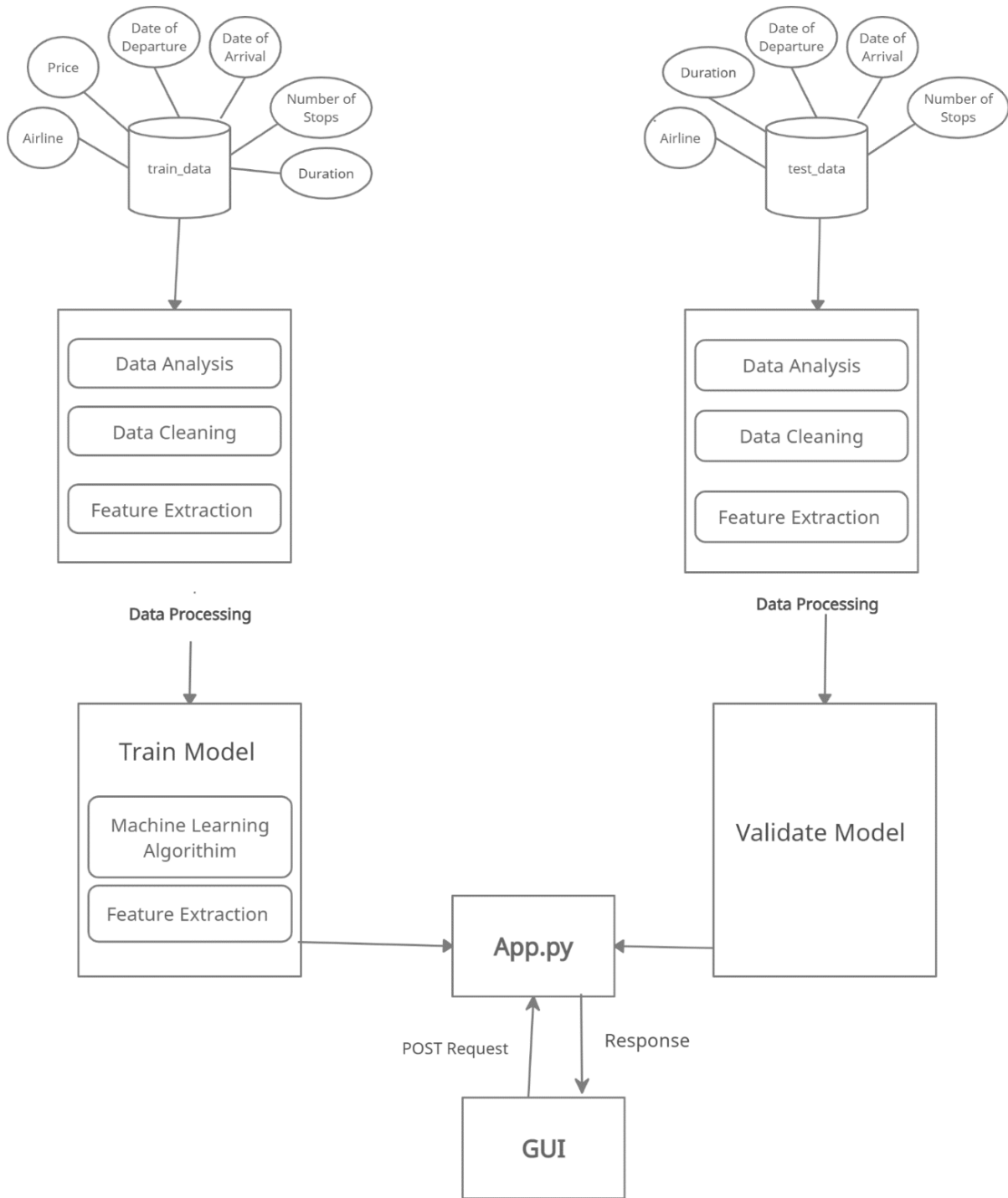
Feature 6: Destination Location

Figure 2 Extracted Features for training the model.

Phase 4: Machine Learning Model Selection

Machine learning is a science that uses statistical techniques to give computer system ability to learn from the given dataset without being explicitly programmed. The supervised learning algorithm deals with labeled data set training for predicting the results. Our system will be provided with label dataset and it is expected to predict the new input data. Therefore, we will use supervised machine learning algorithm.

In our Project we had carried out different Machine Learning Algorithms like Linear Regression, Decision Tree Regression, Random Forest Regression and thought about the precision of results dependent on our test informational index. In view of the different precision levels we observe that Random Forest Regression gives the most noteworthy exactness. In this manner we chose Random Forest Regression and made User Interface dependent on it.



Fig(8) Functionality working of the project chart

Chapter-3 Functionality/Working of Project

If we go deep in our project then we can understand the use of different ML algorithm and web-application framework.

Basic structure of use of things can be seen in this workflow.

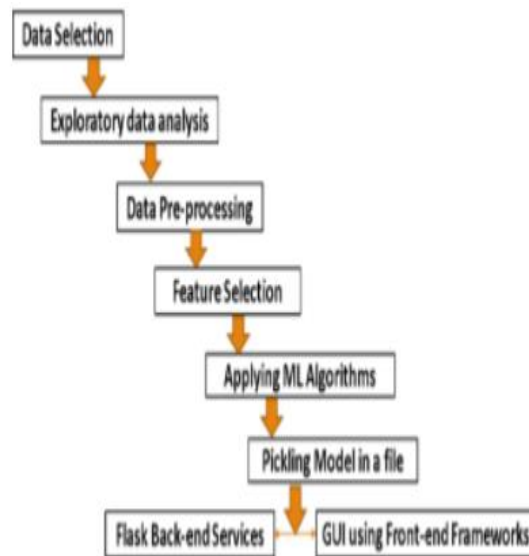


Fig (9) Solution Model

Data Selection:

Data selection is defined as the process of determining the appropriate data type and source, as well as suitable instruments to collect data. Data selection precedes the actual practice of data collection. The process of selecting suitable data for a research project can impact data integrity

Exploring Data:

Data exploration is the first step of data analysis used to explore and visualize data to uncover insights from the start or identify areas or patterns to dig into more. Using interactive dashboards and point-and-click data exploration, users can better understand the bigger picture and get to insights faster.

Data Pre-Processing:

Data preprocessing is the process of transforming raw data into an understandable format. ... The quality of the data should be checked before applying machine learning or data mining algorithms.

ML Algorithms:

At its most basic, machine learning uses programmed algorithms that receive and analyse input data to predict output values within an acceptable range. As new data is fed to these algorithms, they learn and optimise their operations to improve performance, developing 'intelligence' over time.

Picking Model:

The general selection model (GSM) is a model of population genetics that describes how a population's allele frequencies will change when acted upon by natural selection.

Back-end Services:

Wamp servers: To store the query data and details of user.

SQL: To store user and file metadata received by the User.

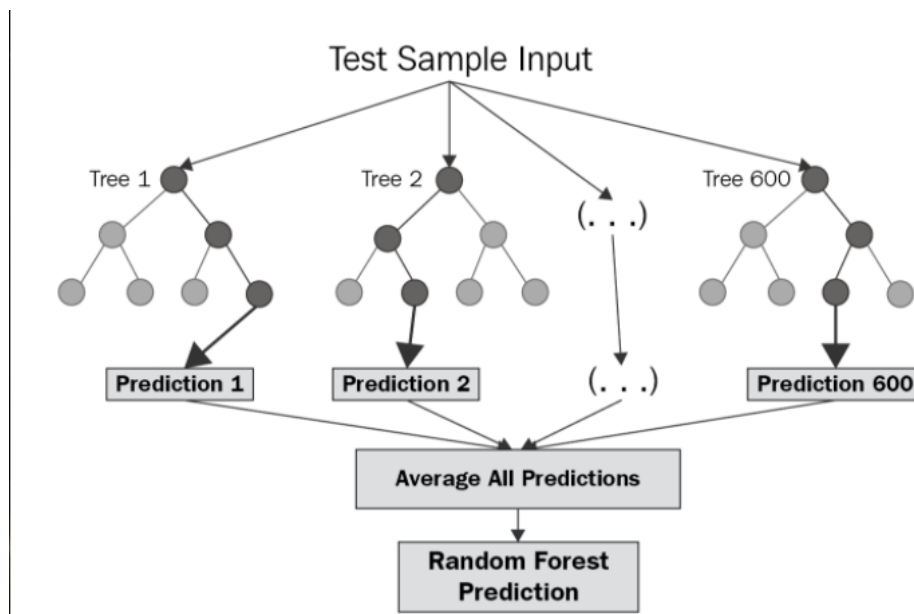
Front end Services:

HTML/CSS: For building client-side web application.

Data Model Used:

Forest Regression Model:

Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model.



Fig(10) Model figure

General Points:

1. Pick at random k data points from the training set.
2. Build a decision tree associated to these k data points.
3. Choose the number N of trees you want to build and repeat steps 1 and 2.

4. For a new data point, make each one of your N -tree trees predict the value of y for the data point in question and assign the new data point to the average across all of the predicted y values.

Steps to specify the good algorithm to use.

Step 1: Identify your dependent (y) and independent variables (X)

Our dependent variable will be *prices* while our independent variables are the remaining columns left in the dataset.

Step 2: Split the dataset into the Training set and Test set.

The importance of the training and test split is that the training set contains known output from which the model learns off of. The test set then tests the model's predictions based on what it learned from the training set.

Step 3: Training the Random Forest Regression model on the whole dataset.

From the sklearn package containing ensemble learning, we import the class **RandomForestRegressor**, create an instance of it, and assign it to a variable.

Step 4: Predicting the Test set results

Now that we've successfully created a Random Forest Regression model, we must assess its performance.

- **Simple Linear Regression: 50%**
- **Multiple Linear Regression: 65%**
- **Decision Tree Regression: 65%**
- **Support Vector Regression: 71%**

- **Random Forest Regression: 81%**

Data set for the project:

	A	B	C	D	E	F	G	H	I	J	K
1	Airline	Date of Journey	Source	Destination	Route	Dep Time	Arrival Time	Duration	Total Stops	Additional Info	Price
2	IndiGo	24/03/2019	Banglore	New Delhi	BLR → DEL	22:20	01:10 22 Mar	2h 50m	non-stop	No info	3897
3	Air India	1/05/2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h 25m	2 stops	No info	7662
4	Jet Airways	9/06/2019	Delhi	Cochin	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	19h	2 stops	No info	13882
5	IndiGo	12/05/2019	Kolkata	Banglore	CCU → NAG → BLR	18:05	23:30	5h 25m	1 stop	No info	6218
6	IndiGo	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h 45m	1 stop	No info	13302
7	SpiceJet	24/06/2019	Kolkata	Banglore	CCU → BLR	09:00	11:25	2h 25m	non-stop	No info	3873
8	Jet Airways	12/03/2019	Banglore	New Delhi	BLR → BOM → DEL	18:55	10:25 13 Mar	15h 30m	1 stop	In-flight meal not included	11087
9	Jet Airways	01/03/2019	Banglore	New Delhi	BLR → BOM → DEL	08:00	05:05 02 Mar	21h 5m	1 stop	No info	22270
10	Jet Airways	12/03/2019	Banglore	New Delhi	BLR → BOM → DEL	08:55	10:25 13 Mar	25h 30m	1 stop	In-flight meal not included	11087
11	Multiple carriers	27/05/2019	Delhi	Cochin	DEL → BOM → COK	11:25	19:15	7h 50m	1 stop	No info	8625
12	Air India	1/06/2019	Delhi	Cochin	DEL → BLR → COK	09:45	23:00	13h 15m	1 stop	No info	8907
13	IndiGo	18/04/2019	Kolkata	Banglore	CCU → BLR	20:20	22:55	2h 35m	non-stop	No info	4174
14	Air India	24/06/2019	Chennai	Kolkata	MAA → CCU	11:40	13:55	2h 15m	non-stop	No info	4667
15	Jet Airways	9/05/2019	Kolkata	Banglore	CCU → BOM → BLR	21:10	09:20 10 May	12h 10m	1 stop	In-flight meal not included	9663
16	IndiGo	24/04/2019	Kolkata	Banglore	CCU → BLR	17:15	19:50	2h 35m	non-stop	No info	4804
17	Air India	3/03/2019	Delhi	Cochin	DEL → AMD → BOM → COK	16:40	19:15 04 Mar	26h 35m	2 stops	No info	14011
18	SpiceJet	15/04/2019	Delhi	Cochin	DEL → PNQ → COK	08:45	13:15	4h 30m	1 stop	No info	5830
19	Jet Airways	12/06/2019	Delhi	Cochin	DEL → BOM → COK	14:00	12:35 13 Jun	22h 35m	1 stop	In-flight meal not included	10262
20	Air India	12/06/2019	Delhi	Cochin	DEL → CCU → BOM → COK	20:15	19:15 13 Jun	23h	2 stops	No info	13381
21	Jet Airways	27/05/2019	Delhi	Cochin	DEL → BOM → COK	16:00	12:35 28 May	20h 35m	1 stop	In-flight meal not included	12898

Fig(11) Sample data set

	A	B	C	D	E	F	G	H	I	J	K
1	Airline	Date of Journey	Source	Destination	Route	Dep Time	Arrival Time	Duration	Total Stops	Additional Info	Price
2	IndiGo	24/03/2019	Banglore	New Delhi	BLR → DEL	22:20	01:10 22 Mar	2h 50m	non-stop	No info	3897
3	Air India	1/05/2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h 25m	2 stops	No info	7662
4	Jet Airways	9/06/2019	Delhi	Cochin	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	19h	2 stops	No info	13882
5	IndiGo	12/05/2019	Kolkata	Banglore	CCU → NAG → BLR	18:05	23:30	5h 25m	1 stop	No info	6218
6	IndiGo	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h 45m	1 stop	No info	13302
7	SpiceJet	24/06/2019	Kolkata	Banglore	CCU → BLR	09:00	11:25	2h 25m	non-stop	No info	3873
8	Jet Airways	12/03/2019	Banglore	New Delhi	BLR → BOM → DEL	18:55	10:25 13 Mar	15h 30m	1 stop	In-flight meal not included	11087
9	Jet Airways	01/03/2019	Banglore	New Delhi	BLR → BOM → DEL	08:00	05:05 02 Mar	21h 5m	1 stop	No info	22270
10	Jet Airways	12/03/2019	Banglore	New Delhi	BLR → BOM → DEL	08:55	10:25 13 Mar	25h 30m	1 stop	In-flight meal not included	11087
11	Multiple carriers	27/05/2019	Delhi	Cochin	DEL → BOM → COK	11:25	19:15	7h 50m	1 stop	No info	8625
12	Air India	1/06/2019	Delhi	Cochin	DEL → BLR → COK	09:45	23:00	13h 15m	1 stop	No info	8907
13	IndiGo	18/04/2019	Kolkata	Banglore	CCU → BLR	20:20	22:55	2h 35m	non-stop	No info	4174
14	Air India	24/06/2019	Chennai	Kolkata	MAA → CCU	11:40	13:55	2h 15m	non-stop	No info	4667
15	Jet Airways	9/05/2019	Kolkata	Banglore	CCU → BOM → BLR	21:10	09:20 10 May	12h 10m	1 stop	In-flight meal not included	9663
16	IndiGo	24/04/2019	Kolkata	Banglore	CCU → BLR	17:15	19:50	2h 35m	non-stop	No info	4804
17	Air India	3/03/2019	Delhi	Cochin	DEL → AMD → BOM → COK	16:40	19:15 04 Mar	26h 35m	2 stops	No info	14011
18	SpiceJet	15/04/2019	Delhi	Cochin	DEL → PNQ → COK	08:45	13:15	4h 30m	1 stop	No info	5830
19	Jet Airways	12/06/2019	Delhi	Cochin	DEL → BOM → COK	14:00	12:35 13 Jun	22h 35m	1 stop	In-flight meal not included	10262
20	Air India	12/06/2019	Delhi	Cochin	DEL → CCU → BOM → COK	20:15	19:15 13 Jun	23h	2 stops	No info	13381
21	Jet Airways	27/05/2019	Delhi	Cochin	DEL → BOM → COK	16:00	12:35 28 May	20h 35m	1 stop	In-flight meal not included	12898

Fig(12) Sample training data set

Modules Descriptions

Module 1:

Home Page

Basically, this page gives the outline of our web application and can also be called as the secondary page of our project.

This Page is just to make the web application more attractive and user friendly so that more traffic can be attracted in our application.

The Page contains all the hyperlink to the other respective pages.

- Home
- Packages
- Gallery
- Contact
- Discount

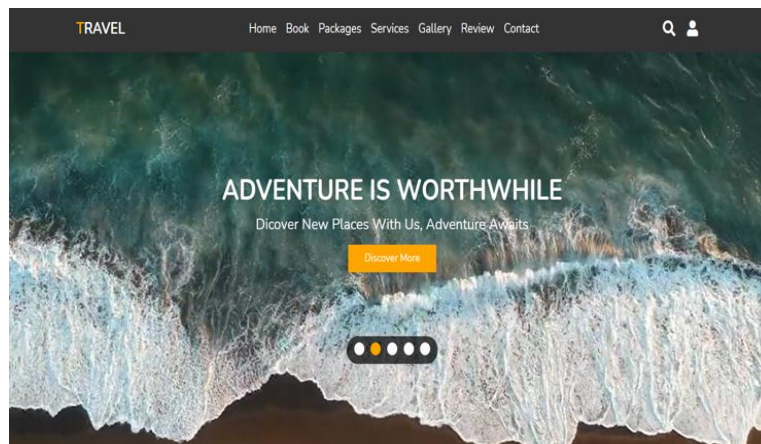



Fig (13) Home Page

Module 2:

Book Now Page

This Page is one of the key pages of our application which tells the different types of price prediction of the flights. Basically, our approach was to make a web application where different algorithms are been used so that we can take out the fare of the price of the flight and then manipulate our discount page which will bring the traffic to our web application.

BOOK NOW



Where To
place name

How Many
number of guests

Arrivals
dd-mm-yyyy

Leaving
dd-mm-yyyy

Book Now

Fig(14) Book Now

Module 3:

Packages Page

This Page defines different types of packages which will tell the different packages price which will give the more attractiveness to the web application. In these different types of cost saving packages are been given which helps the user to know their best use from it.

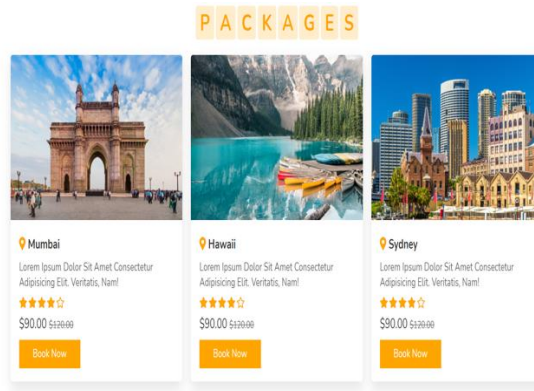


Fig (15) Packages Page

Chapter-4

Results and Discussion

This paper is adapted to the proposed program in VS Code IDE which make daily shifts of services such as chamber services, registering of New Patient, Discharge, appoint a doctor, and finally lists bill etc., Internet properties in file for multiple users.

a) Our paper provides a quick overview of the Automatic AI and web frameworks.

b) Manual record keeping takes time and is prone to make mistakes. Improvement implementation of the flight prediction system, an electronic management system will be used. The paper of an electronic flight prediction is completely computerized and it is useful even for any user members who can see the report and therefore the discount report and many other options.

Efficiency: The most necessary requirement for system performance. The system should work well for whenever new user accepted, discount should be given automatically and the user is given all the details regarding the flight. And if there is another record then new data will be stored in the databases.

Controls: The complete control of the system is in the hands of authorized person. All control is under the controller and members you have the right to simply see the records not to change anything transaction or entry.

Security: Operating systems used. Since illegal access can damage the database and will affect not only application abuttal. So, it also affects the worth of the time and money. So, security has who will be assigned to this task.

1. EXPERIMENTAL RESULTS-

In our project we had implemented various Machine Learning Algorithms such as Linear Regression, Decision Tree Regression, Random Forest Regression and compared the accuracy of results based on our test data set.

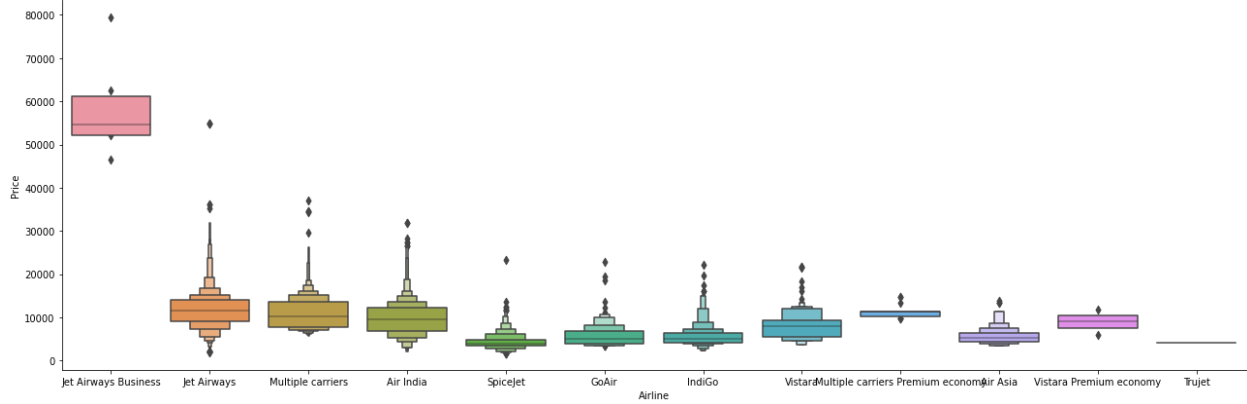


Fig 16 Box Plot for Airline VS Price

As we can see the name of the airline matters. 'JetAirways Business' has the highest price range. Other airlines price also varies.

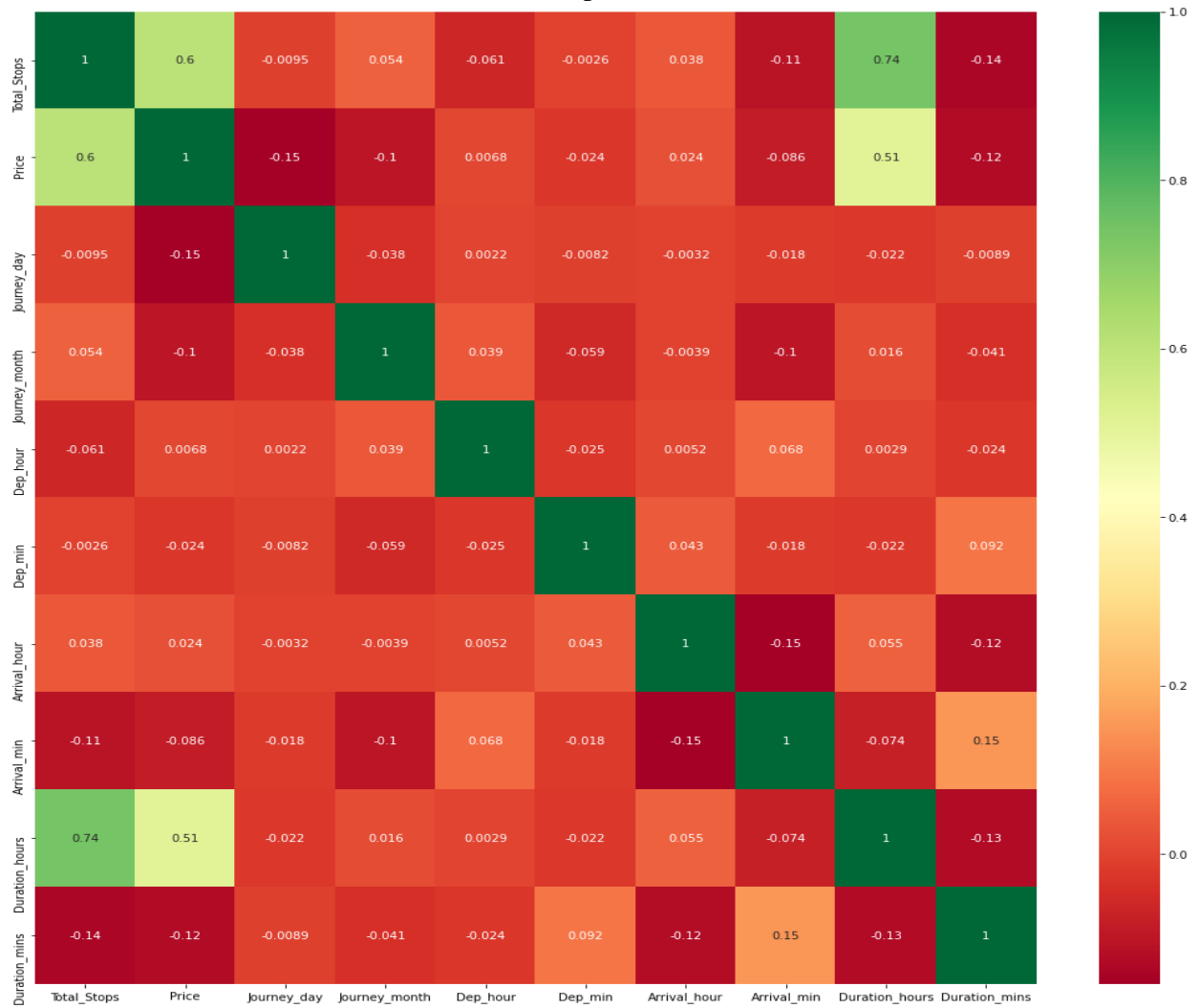


Fig 17 Heat Map

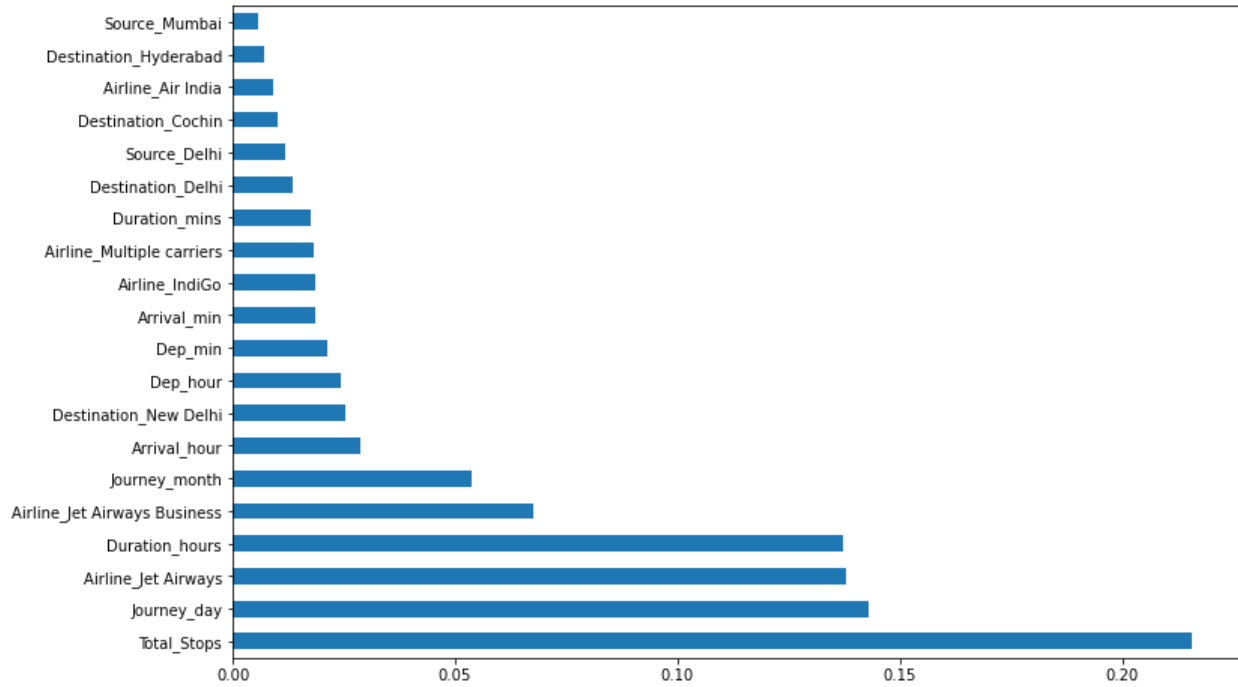


Fig 18 Importance of the Attributes

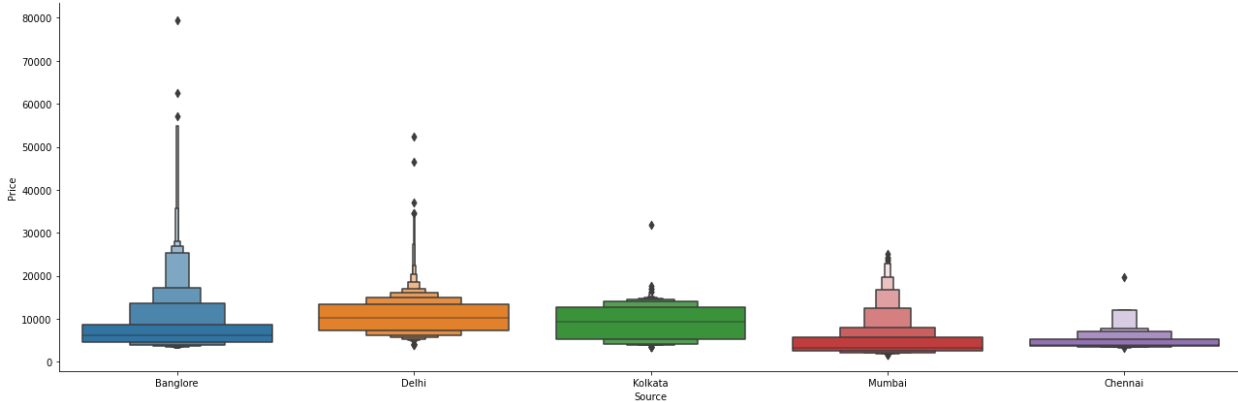


Fig 19 Box Plot for Destination VS Price

Chapter-5

Conclusion and Future Scope

Conclusion

Proper implementation of this project can lead to the savings of inexperienced people by providing them with information related to airline trends and giving them a predictable price, they use to decide whether to book a ticket now or later. I conclude that this type of service can be used with good predictive accuracy. As the predicted value is not completely accurate there is a huge scope for the development of this type of service.

Future Scope

Currently, there are many fields where prediction-based services are used such as stock price predictor tools used by stock brokers and service like Z-estimate which gives the estimated value of house prices. Therefore, there is requirement for service like this in the aviation industry which can help the customers in booking tickets. There are many researches works that have been done on this using various techniques and more research is needed to improve the accuracy of the prediction by using different algorithms. More accurate data with better features can be also be used to get more accurate results.

References

- [1] 2019 IEEE 20th International Conference on Integration data science (IRI).
- [2] Journal of King Saud University - Computer and Information Sciences, Volume 32, Issue 10, December 2020.
- [3] [2017 25th European Signal Processing Conference \(EUSIPCO\)](#).
- [4] Juhar Ahmed Abdella, Nazar Zaki, Khaled Shuaib, "Automatic Detection of Airline Ticket Price and Demand: A review", *Innovations in Information Technology (IIT) 2018 International Conference on*, pp. 169-174, 2018.
- [5] Tianyi Wang, Samira Pouyanfar, Haiman Tian, Yudong Tao, Miguel Alonso, Steven Luis, Shu-Ching Chen, "A Framework for Airfare Price Prediction: A Machine Learning Approach", *Information Reuse and Integration for Data Science (IRI) 2019 IEEE 20th International Conference on*, pp. 200-207, 2019.
- [6] INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH
VOLUME 8, ISSUE 12, DECEMBER 2019 ISSN 2277-8616
- [7] K. Tziridis, Th. Kalampokas, G.A. Papakotas and K.I. Diamantaras, "Airfare Prices Prediction Using Machine Learning Techniques", EUSIPCO 2017.
- [8] INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH Vol.
No. 09, Issue No. 03, March 2021