A Mini Project Synopsis on

Flight Fare Prediction System

T.E. - I.T Engineering

Submitted By

Vedant Patil	**19104065**
Mudra Limbasia	**19104058**
Anvit Mirjurkar	**19104059**

Under The Guidance Of

Prof. Kiran Deshpande

DEPARTMENT OF INFORMATION TECHNOLOGY

A.P. SHAH INSTITUTE OF TECHNOLOGY

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

UNIVERSITY OF MUMBAI

Academic year: 2021-22

CERTIFICATE

This to certify that the Mini Project report on **Flight Fare Prediction System** has been submitted by Vedant Patil (19104065), Mudra Limbasia (19104058) and Anvit Mirjurkar (19104059) who are a Bonafede students of A. P. Shah Institute of Technology, Thane, Mumbai, as a partial fulfilment of the requirement for the degree in **Information Technology**, during the academic year **2021-2022** in the satisfactory manner as per the curriculum laid down by University of Mumbai.

Prod. Kiran Deshpande
Guide

Prof. Kiran Deshpande
Head Department of Information Technology

Dr. Uttam D.Kolekar
Principal

External Examiner(s)
1.
2.

Place: A.P.Shah Institute of Technology, Thane
Date:

TABLE OF CONTENTS

Chapter 1

Abstract

For flying tickets, the airline uses dynamic pricing. According to the poll, flight ticket costs fluctuate throughout the day, especially in the morning and evening. It also varies according to the holidays or festival season. The cost of a plane ticket is determined by a number of distinct factors. The supplier has access to all of the criteria, while purchasers only have access to a limited amount of information, which is insufficient to estimate flight pricing. Our project implementation will suggest the optimal time to buy the ticket based on features such as departure time, number of days till departure, and time of day. The goal of this work is to investigate the factors that drive airfare price fluctuations and how they are related to price changes through integration Machine Learning. Then, using this data, and using Artificial Intelligence, we have developed a system that can assist purchasers in deciding whether or not to purchase a ticket at particular time.

Introduction

Anyone who has ever purchased a plane ticket knows how quickly prices fluctuate. Aircraft use innovative revenue management tactics to implement a unique pricing plan. The cheapest accessible ticket fluctuates over time, and the price of a ticket can be high or low. This valuation approach adjusts the toll according to the time of day, such as morning, afternoon, or night. Seasonal variations in price, such as winter, summer, and holiday seasons, are also possible. The carrier's primary goal is to increase revenue, while the buyer is looking for the best deal. Purchasers typically try to acquire tickets well in advance of the departure date. They believe that airfare will be more expensive as the date of purchase approaches the departure date, although this is not always the case. A purchaser may end up paying more for a similar seat than they should.

According to a research, India's friendly aeronautics industry is on the rise. In 2020, India will be the third-largest avionics market, and by 2030, it would be the largest. By 2017, Indian aviation traffic is expected to reach 100 million passengers, up from 81 million in 2015. According to Google, "Cheap Air Tickets" is the most searched term in India. When the Indian white-collar class is introduced to air travel, customers are looking for low-cost options. The number of low-cost airplane tickets is increasing all the time. Hence developing an application, using Artificial Intelligence and Machine Learning to solve this issue becomes important.

1.1. Purpose:

The Flight Fare Prediction System makes it easier for the travelers to check prices of the flights they wish to buy. Being able to select source, destination, date, airline, etc. gives them a wide range of choices to choose from according to their preference and comfort. They don't need to search and navigate through the different websites and apps to compare the prices, and find the best one, it is hectic and time consuming. It also makes decision making a bit confusing. Checking the prices from the system gives them a fair idea as to what range they should be expecting to spend their money for the flight tickets.

1.2. Objectives:

The Flight Rate Prediction System is aimed to make decision making easier when it comes to booking flight tickets. The objectives of the project can broadly be laid down by the following statements –

1. To make travelling easy for the travelers and save their time.

2. To provide all the necessary flight details on one platform to minimize efforts and minimize decision making time.

3. To bring multiple Airlines, flights and their rates at the tip of your finger through unified web framework development.

4. To minimize consumer efforts by predicting rates of flight according to their choices, through Machine Learning.

5. To minimize time and efforts of the travelers, as they don't need to navigate through different websites, and apps to compare prices.

6. To give the consumers options on their travel preference, like date of departure, source destination, preferred Airline, etc. and predict the rates accordingly.

7. To predict fair rates for the flight tickets according to traveler preference.

8. To give multiple best suited flight options to the passengers.

1.3. Scope

The proposed software product is the Flight Fare Prediction System developed using Artificial Intelligence and Machine Learning. The website will be used for prediction purposed by the developers, and the travelers to check flight rates. The normal hectic process of buying flight tickets at the airport, through agency, or through different websites or apps maybe tiresome and time consuming; external agencies may also charge extra for their services. The intentions of building this system are to make traveling easy and stress free for the travelers. The system makes it easy to filter out the required options, according to the travelers traveling requirements, and predict the flight rate accordingly. The system gives the travelers a range of price around which they should be expecting to spend on the tickets. The consumers can register and login and avail the facility.

The benefits of the project can be represented through the following points listed:

- 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.

Chapter 2

Problem Definition

These are the main drawbacks in the existing systems referred:

- 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.

From a consumer's standpoint, determining the best time to buy plane tickets is difficult, mostly because consumers lack sufficient knowledge to make informed decisions regarding future price fluctuations.

A customer's ability to obtain an airline ticket at a low cost is quite challenging. Several strategies are used to determine the day when the cost of an airline ticket will be the lowest. The majority of these methods rely on advanced artificial intelligence (AI) research known as Machine Learning.

The normal traditional process of booking flights through agencies or in the airport or through websites and apps is very stressful.

Chapter 3

Proposed System

In our proposed system, we are going to provide solutions to all the problems listed in the previous section by automating the whole website by using an integrated software that handles the whole system. The proposed system provides, one platform for all the variety of airlines that the consumer may prefer. Fair rates are predicted for consumer benefit.

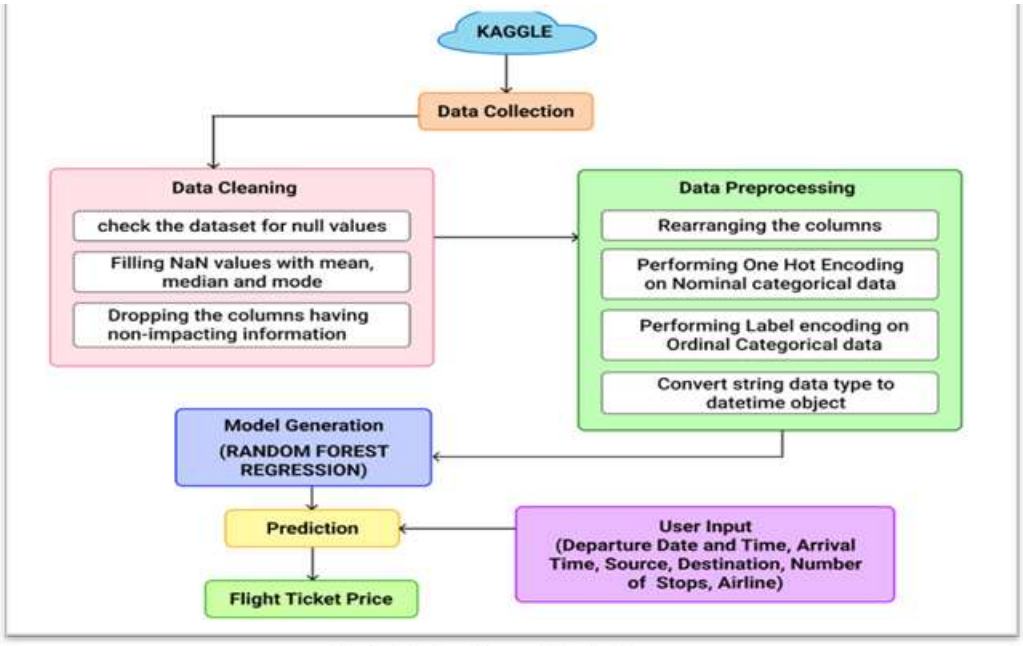


Fig1. Methodology block diagram

3.1 Feature and Functionality

- **Flight rate prediction:** Rates of Airlines and flights are predicted as per the passengers' requirements. Passengers can check flight rates sometime prior so that they can book whenever the rates are feasible. The Sklearn library is used to train the prediction model we created. Some part of the data is also used by the library to test the accuracy of the prediction model, so that the user is provided with the data with utmost accuracy.

- **Flight Halts:** 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. Sklearn is used to train model according to the dataset we have chosen to implement the feature of choosing a flight with stops.

Chapter 4

Project Outcomes

This project aims to develop an application which will predict the flight prices for various flights using machine learning model. The user will get the predicted values and with its reference the user can decide to book their tickets accordingly. The project will be trained using various machine learning algorithms like Extra Trees Regressor and Random Forest Regressor.

The prediction model will be trained according to the dataset that we have selected and will use machine learning algorithms stated above to get an approximate value of the flight rate that the user wants to check. The user will have to select various fields like source, destination, departure date, arrival date, time of flight and the preferred airline, according to which the user will be presented with a predicted value for the flight ticket. Thus, the consumer efforts are minimized as they will get a predicted value and also, they are able to check the prices for various flights at one place which would save their time that they would have spent in searching the flight fare in various websites.

Chapter 5

Software Requirements

Software	Version
PYTHON	**3.9.8**
HTML	**5**
CSS	**3**
FLASK	**2.1.1**
JUPYTER NOTEBOOK	**6.4.10**
GITHUB & GIT	**2.35.1**
MYSQL	**15.1**

1. **Python** – Python is the main component of our project, it is used as a tool for data cleaning, data pre-processing as well as to train and test our model. It is also used in Flask to develop our webpage which is our frontend application.

2. **HTML and CSS** – Html and CSS are used in our frontend application to build and style the website respectively.

3. **Flask** – Flask is used as the base of our graphical user interface (GUI). We have used it to build the frontend of our website along with HTML and CSS. It is a micro web framework that is used to deploy a website on a server.

4. **Jupyter Notebook** – Jupyter Notebook in our project plays a key part. We have used it as an IDE to develop our code, train our machine as well as to test the prediction model that we have developed.

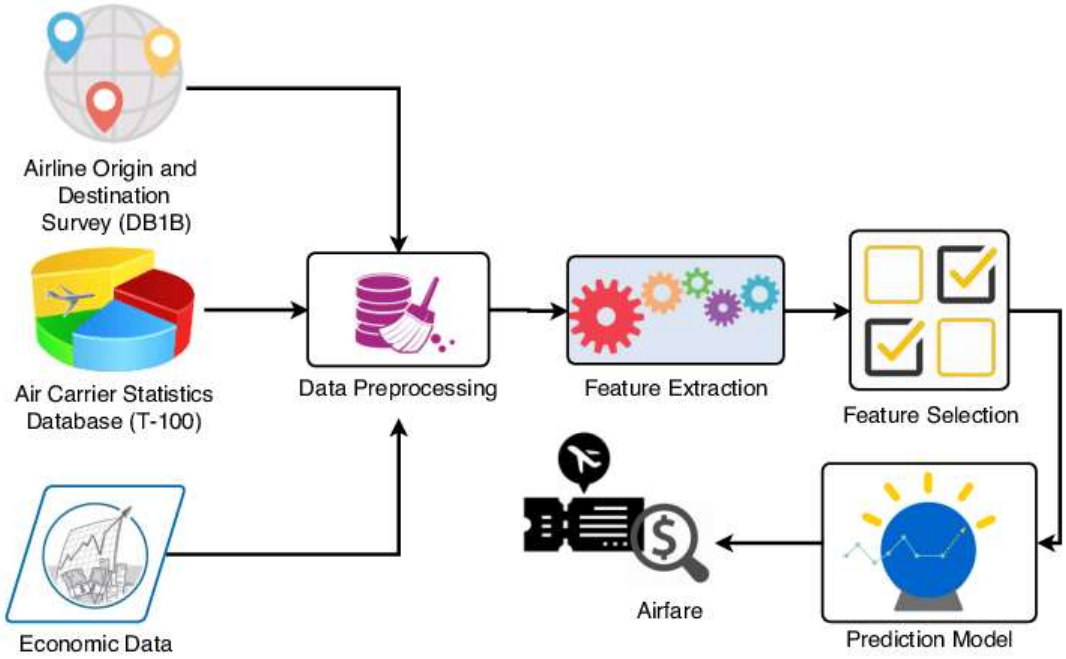
5. **GITHUB & GIT** – We have used Github to deploy our code on a web platform where all the group members can contribute their work done. While Git is a system that we have used to track the changes made in our Github repository by the team members. We also get the list of contributors in our project.

6. **MySQL** – MySQL is implemented in our project to store the registration data like username, email-id, password.

7. Data set used can be found at https://www.kaggle.com/datasets/nikhilmittal/flight-fare-prediction-mh

Chapter 6

Project Design



- In the initial phase various databases are collected, which are required for further implementation from trusted sources like 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 like Extra Trees Regressor and Random Forest Regressor. And finally, the air fare prediction system is built.

6.1. Implementation

For this project, we have integrated the machine learning life cycle to web framework which will predict the flight prices by applying machine learning algorithm to historical flight data using python libraries like Pandas, NumPy, Matplotlib, seaborn and sklearn. Figure.1 shows the steps that we followed from the life cycle:

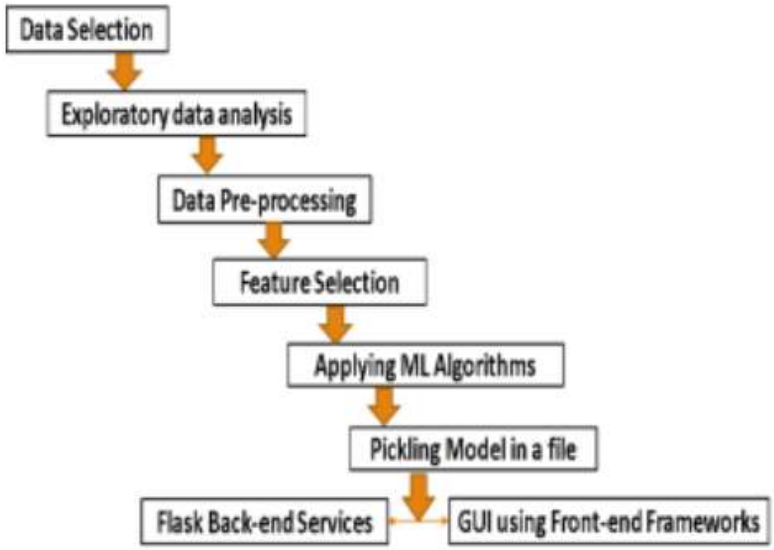


Fig. 1. Machine Learning Life Cycle

DATA COLLECTION The collection of data is the most important aspect of this project. There are various sources of the data on different websites which are used to train the models. Websites give information about the multiple routes, times, airlines and fare. Various sources from APIs to consumer travel websites are available for data scraping. In this section details of the various sources and parameters that are collected are discussed.

To implement this data is collected from a website “Makemytrip.com” and python is used for the implementation of the models and collection of the data.

A. Collection of data

The script extracts the information from the website and creates a CSV file as output. This file contains the information with features and its details. Now an important aspect is to select the features that might be needed for the flight prediction algorithm. Output collected from the website contains numerous variable for each flight but not all are required, so only the following feature is considered.

• Origin

• Destination
• Departure Date
• Departure Time
• Arrival Time
• Total Fare
• Airways
• Taken Date

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 (BOM to DEL) over a period of three months that is from February to April. For each flight data with all the features collected manually.

B. 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.

C. Analyzing data

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.

IV. MACHINE LEARNING ALGORITHM

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, Multilayer

Perceptron[10], Support Vector Machine (SVM) and Gradient Boosting. All these models are implemented in the scikit learn. To evaluate the performance of this model, certain parameters are considered. They are as follows: R-squared value, Mean Absolute Error (MAE) and Mean Squared Error (MSE). The formulas for these three parameters are as follows:

$$R^2 = 1 - \frac{\sum_n^{t=1}(yi - \hat{yi})^2}{\sum_n^{t=1}(yi - \bar{yi})^2} \quad (1)$$

$$MAE = \frac{1}{n}\sum_{n}^{t=1}|yi - \hat{yi}| \quad (2)$$

$$MSE = \frac{1}{n}\sum_{n}^{t=1}(yi - \hat{yi})^2 \quad (3)$$

A. Linear Regression

Regression is a method of modeling a target value based on predictors that are independent. It is mostly based on the number of independent variables and the relationship between independent and dependent variables. linear regression is a type of analysis where the number of independent variables is one and the relationship between the dependent and independent variables vary linearly. The important concept to understand linear regressions are cost function and Gradient decent.

y(pred) = b0+b1 ∗ x (4)

B. Decision tree

The Decision tree calculation separates the informational collection into small subsets, at a similar same time it creates gradually. The last outcomes are the tree with the decision nodes, what's more, the leaf nodes. A decision hub may have at least two branches. In the beginning, consider the entire informational collection as root. Highlight esteems are wanted to be downright. On the off chance that the qualities are constant then they are discretized before structure the model. Based on characteristic qualities records are dispersed recursively. 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. Gini Index is a component that measures how frequently an arbitrarily picked component would be mistakenly distinguished. It implies a characteristic with a lower Gini index ought to be liked.

C. Random Forest

It is a supervised learning algorithm. The benefit of the random forest is, it very well may be utilized for both characterization and relapse issue which structure most of current machine learning framework. Random forest forms numerous decision trees, what's more, adds them together to get an increasingly exact and stable expectation. Random Forest has nearly the equivalent parameters as a decision tree or a stowing classifier model. It is very simple to discover the significance of each element on the expectation when contrasted with others in this calculation.

The regular component in these techniques is, for the kth tree, a random vector theta k is produced, autonomous of the past random vectors theta 1, ... , theta k-1 however with the equivalent distribution, while a tree is developed utilizing the preparation set and bringing about a classifier. x is an information vector. For 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. These methodologies are called as random forests.

D. Extra Tree Regression

Extra Trees is an ensemble machine learning algorithm that combines the predictions from many decision trees.

It is related to the widely used random forest algorithm. It can often achieve as-good or better performance than the random forest algorithm, although it uses a simpler algorithm to construct the decision trees used as members of the ensemble.

It is also easy to use given that it has few key hyperparameters and sensible heuristics for configuring these hyperparameters.

In this tutorial, you will discover how to develop Extra Trees ensembles for classification and regression.

Chapter 7

Project Scheduling

Sr. No	Group Member	Time duration	Work to be done
1	Vedant Patil	4th week of January	Implementing 1st module/ functionality
		2nd week of February	Testing 1st module
2	Mudra Limbasia	1st week of March	Implementing 2nd module
3	Anvit Mirjurkar	By the end of march month	Implementing 3rd module/ functionality

Chapter 8

Implementation snippets

This section represents some critical snippets addressing objectives discussed.

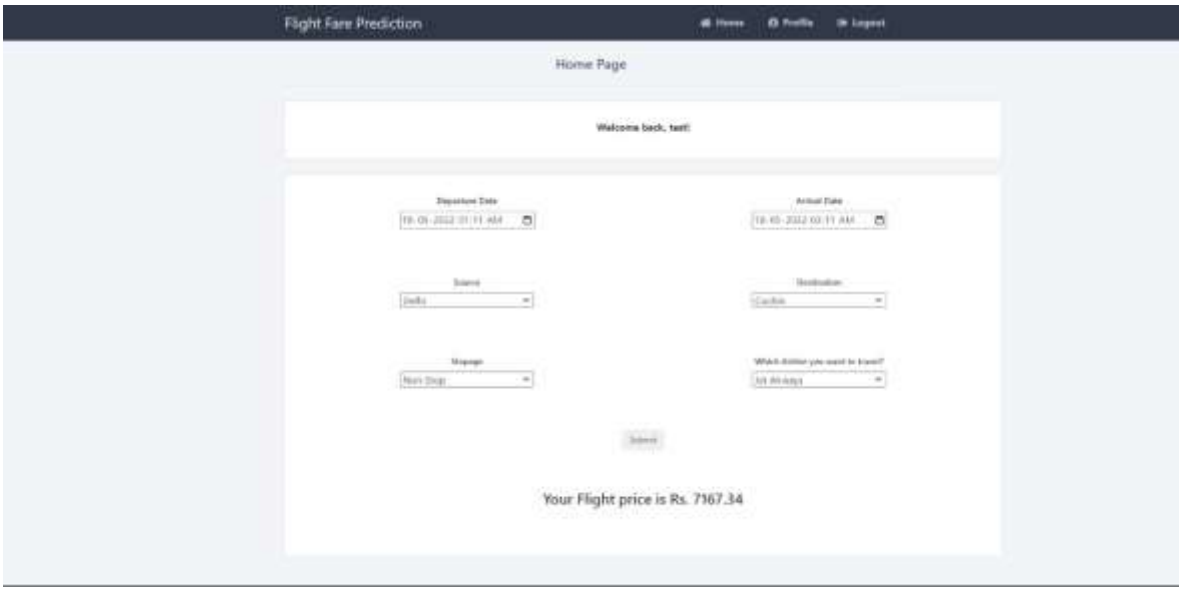


Figure 8.1: User Interface representing Flight Rate Prediction.

User Interface of Flight Fare Prediction System: This is the main page of the website where users can check the predicted flight rates for the flights, they wish to travel in. The user will have to choose the criteria like departure date, arrival date, source, destination, stoppage and the Airlines they wish to travel in. And the rates of the flight ticket will be predicted for the user after they click the submit button.

Chapter 9

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.

References

- K. Tziridis T. Kalampokas G. Papakostas and K. Diamantaras "Airfare price prediction using machine learning techniques" in European Signal Processing Conference (EUSIPCO), DOI: 10.23919/EUSIPCO.2017.8081365L.
- Li Y. Chen and Z. Li" Yawning detection for monitoring driver fatigue based on two cameras" Proc. 12th Int. IEEE Conf. Intell.Transp. Syst. pp. 1-6 Oct. 2009.
- Viet Hoang Vu, Quang Tran Minh and Phu H. Phung, ‖An Airfare Prediction Model for Developing Markets‖, IEEE paper 2018.

ACKNOWLEDGEMENT

This project would not have come to fruition without the invaluable help of our guide **Prof. Kiran Deshpande**. Expressing gratitude towards our HoD, **Prof. Kiran Deshpande**, and the Department of Information Technology for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our project co-ordinator Prof. Nahid Shaikh who gave us her valuable suggestions and ideas when we were in need of them. We would also like to thank our peers for their helpful suggestions.