

A Novel Machine Learning Methodology to Increase Sales in Business Services

Tadvi Shabana^{1*}, Shaikh Afifa², Sayyed Naziya³, Khan Mariya⁴

^{1,2,3,4}Computer Engineering, M.H Saboo Siddik College of Engineering, Mumbai University, Mumbai, India

**Corresponding Author: shabanatadvi15@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 27/Dec/2018, Published: 31/Dec/2018

Abstract— Ticket purchasing in advance is a well- known traditional approach but it entirely depends on the Airline industry to change the fare according to factors whether the travel is during the holidays, the number of free seats in the plane etc. Some of the features are seen, but some of them remained hidden. We are using Indian Domestic Airline Dataset which contains multiple columns so over a period as the data increases (approx. 1 year) we will be able to extract few more hidden features to increase the efficiency and accuracy of the system. The goal is to use machine learning techniques to model the behaviour of flight ticket prices over the time. In other words system will be able to provide a general idea to the clients when to increase or decrease the fares i.e. prediction of Airfare. For that after collecting the dataset the proposed system will extract important features from dataset, cleaning of data and using Regression Machine Learning Algorithms multiple models will be trained and the accuracy of those models will be compared and prediction report will be given to client.

Keywords— Airfare, Feature Extraction, Cleaning data, Regression, Machine Learning, Data Analytics

I. INTRODUCTION

Business Services provides us intangible products such as accounting, banking, consulting and finance etc. Airline industry is one of the sectors which provide us with the above given services. It also provides air transport services for travelling passenger. These airline services can be categorized as being intercontinental, domestic, regional or international. The goal of improving business sales (revenue) in airline industry depends on the flight ticket selling.

Now-a-days the airline industries are using complex algorithms and strategies for the airfare prices in a dynamic way to regulate seats demand and maximize their revenue. Though the strategies developed works properly but aren't that accurate. The airfare price depends on many factors like base fare (airfare), crude oil, meal and route. There are so many airline agencies through which we can also book the tickets and sometimes we find their prices lower than direct booking. This happens because those agencies have their own commissions from the flight sales. Though they get all fare related information from connected airline and try to adjust their airfare prices so that more customers should use them for booking flight ticket online. If the journey is not too long then passenger can probably skip meal, Wi-Fi or other facilities charges in order to decrease the fare price [1]. But the **Base Fare** will be common for all the passengers containing **Air Charges**. So airfare prices can be dynamically changed by considering many factors. For

generating more revenue it is necessary for an airline industry to create some good strategy to predict airfare prices so that it increases industry's profit as well as customer's welfare.

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. 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. Random forest regression, linear regression, K-nearest neighbours Regression etc. [8]. 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**".

Section II contains the related work of the system and it also explains the limitations, Section III briefs about the procedures and methods followed to achieve the desired goal with block diagram, Section IV describes results and discussion, section V concludes research work with future directions.

II. RELATED WORK

In the existing Airfare prediction systems, for predicting dynamically increasing or decreasing prices of the air different behavioural characteristics of the timing factors are considered and are given much priority. Although, different features are also taken into consideration but the increasing price of crude oil is not taken into consideration [4] which indeed have a high tendency of affecting the airfare and that too at a higher level. The systems taken into considerations are all mentioned in the literature survey.

Limitations:

The existing systems doesn't provide any severe drawbacks but it did have certain limitations

- a. The system doesn't have sufficient data for better prediction.
- b. The system changes accuracy with changing algorithm and so it becomes a bit confusing though the accuracy only changes much when features important features are removed.
- c. The system doesn't include the behavioural changing pattern of the crude oil.

III. METHODOLOGY

To train and test the data and check the accuracy of all models

- f. To generate prediction report that will predict AIRFARE prices.

According to the research the existing approach doesn't focus on the Indian domestic airlines dataset. It moreover focuses on the international airline dataset. They have not considered factors like crude oil or number of intermediate stops etc.

The proposed system is composed of four phases:

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

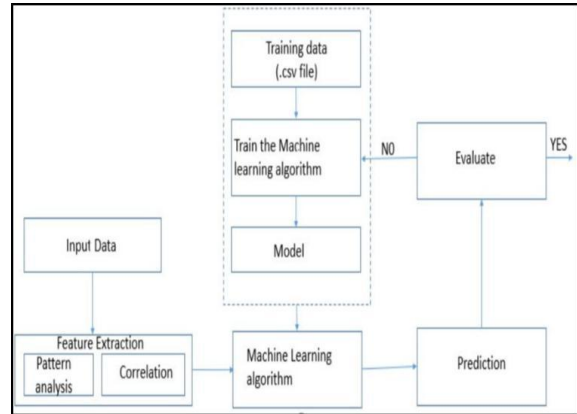


Figure 1: Block Diagram

To develop a web 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.

Phase 1: Data Input

The input data file is 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.

Phase 2: 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: departure time
- Feature 2: numbers of days till flight departure
- Feature 3: holiday day (yes or no)
- Feature 4: overnight flight (yes or no).
- Feature 5: day of the week
- Feature 6: crude oil

Phase 3: 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 labelled 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. [8]

For continuous flight fare changing data we will use regression machine learning ml models which are as follows:

1. Random Forest regression tree.
2. Logistic regression.
3. Decision Tree.
4. Begging regression tree.
5. Linear regression.

Phase 4: Prediction and Evaluation

When the input file or input data will be provided to trained ML model then it will predict some output which will be compared with the expected outcome. If outcome matches with the expected output then it will be accepted else it will again be given to the ML model [7].

IV. RESULTS AND DISCUSSION

The main objective of this paper is to analyse the dependency of air prices on different features and build a prediction model that could help Airline industry to predict price of Air ticket and gain maximum profit.

The following is the result of analysis done on the given dataset for domestic Airline using jupyter-notebook and it uses various python libraries like pandas, numpy and scikit-learn.

	RES_SEG_STATUS	STATECODE_G8	PAX_TYPE
RES_SEG_STATUS	1.000000	0.003345	-0.006332
STATECODE_G8	0.003345	1.000000	0.028819
PAX_TYPE	-0.006332	0.028819	1.000000
SEG_RES_CHANNEL	0.064907	-0.046235	0.012491

Figure 2: Correlation between columns

The term "correlation" refers to a mutual relationship or association between quantities. It decides prediction of one quantity from the other. Figure 2 shows dependencies among columns and this output gives a broader look in deciding the columns in model preparation.

V. CONCLUSION AND FUTURE SCOPE

This paper reported a study on airfare prediction using data analytics and machine learning. We gathered dataset from Domestic Indian Airlines of one month and applied data cleaning, features extraction and finding correlation [4].The experimental result shows what features are necessary for developing a prediction model and how they are interrelated with each other. To train the model different ML algorithms are used such as Random Forest, Linear regression and K-nearest algorithm.

Apart from selected features other factors can be also involved to improve the accuracy of the model. In future this proposed model can be trained for more than one year of data, if this is extended further then the processing speed as well as the power of the computer required will be more so this can be implemented using Big data analytics to improve sales in business services[3]

REFERENCES

- [1] A regression model for predicting optimal purchase timing for airline tickets.
- [2] "A Model of Optimal Consumer Search and Price Discrimination in the Airline Industry". David Li Sunday 15th November, 2015
- [3] International Journal of Computer Science and Mobile Computing "big data analysis of airline data set using hive" by p. swathi1, j. kumari2.
- [4] International journal of engineering science invention "airfare analysis and prediction using data mining and machine learning" by bhavuk chawla1,ms.chandandeep kaur2.
- [5] "Dynamic Pricing in the Airline Industry" By R. Preston McAfee and Vera Te Velde: California Institute of Technology.
- [6] Airline Data Set, United States Department of Transportation, Office of the Assistant Secretary for Research and Technology, Bureau of Transportation Statistics, http://www.transportation.gov/DL_SelectFields.asp?Table_ID=236
- [7] William Groves and Maria Gini, "On Optimizing Airline Ticket Purchase Timing", University of Minnesota, 2011
- [8] Manolis Papadakis, "Predicting Airfare Prices" in Stanford, 2013
- [9] Yuwen Chen, Jian Cao, Shanshan Feng and Yudong Tan, "An ensemble learning based approach for building airfare forecast service" Big Data (Big Data), 2015 IEEE International Conference, 29 Oct.-1 Nov. 2015.
- [10] WEKA Manual for Version 3-6-8, The University of Waikato, 2012