

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Airfare Price Prediction Based on Reviews Using Machine Learning Techniques

Aakanksha V. Jain, Aanal S.Raval, Ruchi K. Oza

M. E Students, Dept. of Computer, Silver Oak College of Engineering and Technology, Ahmedabad, Gujarat, India.

ABSTRACT: This paper deals with the problem of correct airfare prices prediction and their reviews. Along with the prediction accuracy of each model, this paper studies the dependency of the accuracy on the feature set used to represent an airfare. For the experiment, dataset consisting for reviews, I take twitter and sky-traxwebsite(2017-2020). And for price dataset I take different Indian airlines' price database(2019).

KEYWORDS: prediction model,airfare price, pricing models,sentimental analysis,Twitter, Indian Airlines

I.INTRODUCTION

With the explosive growth of the Internet and ecommerce, air passengers nowadays can check airfare and availability of any airlines around the world easily. When satisfying with an airfare, these customers can purchase their desired tickets online through official airline or agent websites. To help the customers to buy the most inexpensive airfare, there have been a number of prediction models to predict the airfare prices.

Social media today is an integral part of people's daily routines and the livelihood of some. As a result, it is abundant in user opinions. The analysis of brand specific opinions can inform companies on the level of satisfaction within consumers.

Airline Ticket prices change very dynamically and significantly for the same flight day by day. It is very difficult a customer to purchase an air ticket in the lowest price, since the price changes dynamically.Sentimate analysis (Reviews) also helps to customers in decision making on selection of best airlines on analyzing other customer's opinion.The purpose of this study is to better analyze the features that affect airfare and to develop and tune models to predict the airfare well in advance.

The rest of this paper is organized as follows: Section II, presents Related work about this paper. Section III describes Methodology and Section IV discusses their results. Finally, Section V concludes conclusion.

II.RELATED WORK

One of the future directions that has great potential that improve the ticket price and demand prediction is to use the latest and advanced ML techniques (i.e. deep learning) in conjunction with valuable social media-based data. Sentiment analysis of different twitter hash tags could convey the presence of some Consumer loyalty prediction is performed using Twitter data. Three classifiers are employed, namely, Random Forest, Decision Tree and Logistic

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Regression. A maximum classification accuracy of 99.05% is observed for Random Forest in previous papers. And for price prediction Bagging Regression tree gives highest result that is 87.42% in paper[1].

Performance improvements on a variety of traditional problems. Within deep learning, convolutional neural networks (CNNs) have received much attention recently. Also, XG-Boost is an implementation of gradient boosted decision trees designed for speed and performance.

III. METHODOLOGY

METHODS FOR PRICE PREDICTION

A. Feature Selection

For price prediction following features are considered: Arrival date, Departure date, Price, Airline name.

B. Dataset Collection

For price prediction, here we consider Indian airlines dataset for testing [Indian Airline dataset-March 2019-July 2019, 10683 rows].

C. (ML Models Selection): Here different ML models are considered to predict the impact of features selected earlier on users' preferences.

The ML models compared in this work are the following:

Bagging Regression tree
Random Forest
Multilayer Perceptron (MLP)
XG-Boost

METHODS FOR AIRLINE'S REVIEWS ANALYSIS

A. Dataset and Preprocessing

For reviews, Twitter and sky-trax sites are considered. (2017-2020) Preprocessing includes conversion to lower case, elimination of URLs, stop words, tokenizing, lemmatizing by WordNet Lemmatizer, and truncation by 5-fold cross validation.

(Feature Selection):

For every flight, Reviews of different Airlines is considered.

B. Sentiment Analysis

Twitter is an online platform that connects people through a social networking environment. Each user can create an account with a unique handle and post —Tweets. Sentiment analysis becomes important for identifying positive or negative tweets and determining the consumer voice. It discerns whether a particular text or piece of writing – tweet, for example – is positive, negative or neutral. It uses context of the piece as well as tone, emotion and vocabulary. Sentiment analysis can aid companies in marketing strategies by understanding a general public opinion and in

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

succession improving their customer service. A company can determine the public opinion, analyze the customer satisfaction towards their products and be open to listening to the issues. This analysis can, not only help companies to know how their customers think, but can also aid them in competitor advantage.

For this we use Sonnetword function to count the unique words (positive and negative words). From each Tweet, its sentiment score is computed and the tweet is segregated into positive type or negative type as given. Positive words such as happy, great, nice, thanks, enjoyed, helpful etc and Negative words such as sad, poor, useless etc are considered. SVM algorithm is used to sentiment positive and negative wods.

Positive words	Negative words
Happy	Avoid
Great	Hopeless
Comfortable	Worst
Safest	Uncool
Care	Careless

Table I

C. prediction:

(ML Models Selection):

XG-Boost

For classification stage, here we tested eXtreme Gradient Boosting [11],RegressionTree,MultilayerPerceptron. As this algorithm is more efficient in text classification than any other currently present for handling sparsity patterns.

Unlike random forests, XGBoost provides methodical and systematic solution and aggregates prediction power of multiple learners. Below diagram shows the complete flow of system.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

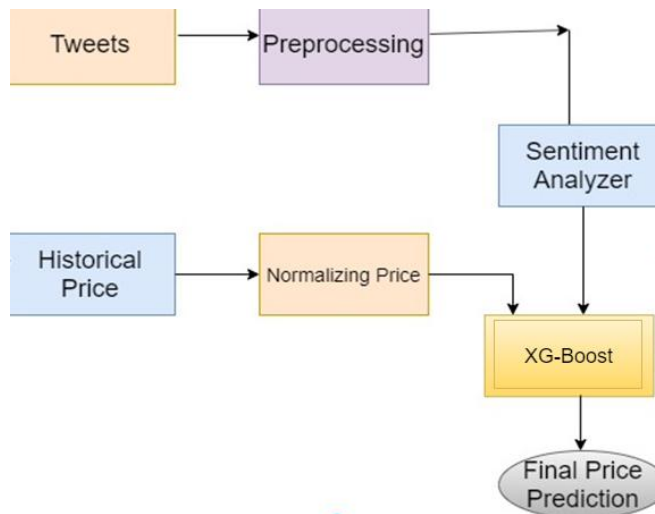


Fig: Flow Diagram

IV. RESULTS

REVIEW ANALYSIS RESULTS

XG-Boost gives higher results compare to other defines machine learning algorithms for analysis(Reviews) of different Indian airline, combined with SVM. XG-Boost gives 94.34% accuracyfor Airline Review Dataset.

ML Model	Accuracy
Decision tree	73%
Random Forests	82%
XGBoost	94%

TABLE-1

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

PRICE PREDICTION RESULTS

And for price prediction following results got, which is more result to previous one.

ML Model	Accuracy
Bagging Regression Tree	87.42%
MLP	80.28%
Random Forest	84.13%
XG-Boost	90.2%

TABLE-2

V.CONCLUSION

This paper reported on a preliminary study in “airfare prices prediction”. We gathered airfare data from a specific Indian airline from the web and showed that it is feasible to predict prices for flights based on historical fare data. We identified the compliments and complaints of customers, variations in sentiment over a period of time. This analysis provides a general opinion of passengers towards airlines. XG-boost gives better result compare to other machine learning models.

REFERENCES

- [1]K. Tziridis, Th. Kalampokas, G.A. Papakostas, “Airfare Prices Prediction Using Machine Learning Techniques”,978-1-5386-2290-2/18©2018 IEEE
- [2] Viet Hoang Vu, Quang Tran Minh, “An Airfare Prediction Model for Developing Markets”, Ho Chi Minh City University of Technology, VNU - HCM Ho Chi Minh City, Viet Nam, 978-1-5386-2290-2/18/ ©2018,IEEE
- [3]Rida Khan, SiddhalingUrolagin, “Airline Sentiment Visualization, Consumer Loyalty Measurement and Prediction using Twitter Data”,BITS Pilani, Dubai Campus, Academic City Dubai, United Arab Emirates, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol.9, No. 6, 2018
- [4] E. Prabhakar, “Sentiment Analysis of US Airline Twitter Data using New Adaboost Approach”, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181,2019
- [5] Juhar Ahmed Abdella, NazarZaki and Khaled Shuaib, “Automatic Detection of Airline Ticket Price and Demand: A review” , United Arab Emirates University, College of Information Technology Al Ain, UAE,2018 [6] Juhur Ahmed Abdella, NazarZaki, Khaled Shuaib, Fahad Khan, “Airline ticket price and demand prediction: A survey” ,Department of Information Systems and Security,UAEU,sciencedirect,2019
- [7] Yuling Li, Zhichao Li, “Design and implementation of ticket price forecasting system”, AIP Conference Proceedings, 23 May 2018
- [8] Deb Dutta Das, “Sentimental Analysis for Airline Twitter data”, 2017 IOP Conference
- [9] SupriyaRajankar. Neha Sakharkar, “A Survey on Flight Pricing Prediction using Machine Learning”, International Journal of Engineering Research & Technology (IJERT), June-2019
- [10] Bhavuk Chawla, Ms. Chandandeep Kaur, “Airfare Analysis And Prediction Using Data Mining And Machine Learning”, International Journal of Engineering Science Invention, November 2017
- [11]Priyadharshini P.,”PREDICTION OF DIABETES MELLITUS USING XGBOOST GRADIENT BOOSTING”, Sethu Institute of Technology, kariapatti, virudhunagar, Tamilnadu., International Journal of Advances in Science Engineering and Technology, ISSN(p): 2321 – 8991, ISSN(e): 2321 –9009 Vol-5, Dec.-2017
- [12] Jun Lu, “Machine learning modeling for time series problem: Predicting flight ticket prices”,Computer Science, EPFL,4 Feb 2018