JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Flights Price Detection using Machine Learning

Ankita Dhumal[1], Namrata Patole[2], Shivani Dulal[3], Pragati Nikam[4]

Department of Electronics & Telecommunication Engineering, Smt.Kashibai Navale College of Engineering

[1]dhumalankita1@gmail.com

[2]namrata.nikam_skncoe@sinhagad.edu

[3]shivanidulal28@gmail.com

[4]pragatinikam32@gmail.com

Abstract— **The airline implements dynamic pricing for flight tickets. According to the survey, flight ticket prices change during the morning and evening time of the day. Also, it changes with the holidays or festival season. There are several different factors on which the price of the flight ticket depends. The seller has information about all the factors, but buyers can access limited information only which is not enough to predict the airfare prices. Considering the features such as departure time, the number of days left for departure, and time of the day it will give the best time to buy the ticket. The proposed system can help buyers whether to buy a ticket or not.**

***Keywords*—Machine learning Algorithm, predictor, SVM, Random Forest.**

I. INTRODUCTION

Flight ticket prices increase or decrease every now and then depending on various factors like the timing of the flights, destination, and duration of flights. In the proposed system a predictive model will be created by applying machine learning algorithms to the collected historical data of flights. With the explosive growth of the Internet and e-commerce, air passengers nowadays can check the airfare and availability of any airline around the world easily. When satisfied with airfare, these customers can purchase their desired tickets online through official airline or agent websites. To help customers to buy the most inexpensive airfare, there have been a number of prediction models to predict airfare prices.

Since the airline industry's deregulation, airfare pricing strategy has evolved into a complex structure of sophisticated rules and mathematical models that drive airfare pricing strategies. Even though these rules are still largely unknown, studies have revealed that they are influenced by a variety of factors. Traditional variables such as distance, while still important, are no longer the sole determinants of pricing strategy. Economic, marketing, and societal trends have all played a growing role in determining airfare prices. Flight Ticket prices for the same flight fluctuate dramatically and significantly from day to day. It is extremely difficult for a customer to purchase an airline ticket at the lowest possible price because the price changes dynamically. Customers can also use septimate analysis (reviews) to help them make decisions about which airlines to fly by analyzing the opinions of other customers. The goal of this research is to better understand the factors that influence airfare and to develop and finetune models that can predict airfare well in advance.

II. MACHINE LEARNING

Machine Learning is one of the hottest research topics in computer science and engineering, which is applicable in many disciplines. It provides a collection of algorithms, methods, and tools able to embody some kind of intelligence to machines. The power of ML is the provided modeling tools, which are able to be trained, via a learning procedure, with a set of data describing a certain problem and to respond to similar unseen data in a common way. Some well-known ML models are Multilayer Perceptron's (MLPs), Radial Basis Function (RBF) and Generalized Regression (GRNN) neural networks, Support Vector Machines (SVMs), Decision Trees (DTs), Extreme Learning Machines (ELMs), etc.

One of the reasons that ML has attracted scientists from several disciplines is its ability to provide humanlike intelligence to machines as the amount of data used during learning increases. However, the increase of the training data needs parallel implementations of the ML algorithms using specialized software and/or hardware platforms. In the context of machine learning, there are two possible alternatives for handling the problem of airfare pricing prediction. The first approach tackles the prediction of air ticket prices as a regression problem, while the second one transforms it into a classification task. The former strategy is usually applied for the prediction of the exact air ticket price since the regression models try to approximate a function that describes the mapping law between data features and airfare prices. The latter approach cannot predict the exact air ticket prices but can provide decisions regarding the range of a price or a decision to buy or not the ticket with the specific price.

III. LITERATURE SURVEY

Evaluating the algorithmic rule, a dataset is collected, and pre-processed, performed data modelling, and studied a value difference for the number of restricted days by the passengers traveling. Machine Learning algorithms with square measures for forecasting the accurate fare of airlines and it gives accurate value of plane price ticket at limited and highest value.[1] Gathered airfare data from a specific Greek airline corporation (Aegean Airlines) from the web and showed that it is feasible

to predict prices for flights based on historical fare data. The experimental results show that Models are a satisfactory tool for predicting airfare prices. .[2]

A context-aware ensemble algorithm for airfare price forecasting is proposed. Airfare forecasting can be regarded as a time series problem and machine learning techniques have been used to solve this problem. We combine context modeling, clustering, and ensemble techniques and propose ACER, an adaptive context-aware ensemble regression framework. ACER finds feature.[3]

A machine learning framework was developed to predict the quarterly average airfare price on the market segment level. We combined the U.S. domestic airline ticket sales data and non-stop segment data from two public datasets (DB1B and T-100). With the help of the feature selection techniques, our proposed model is able to predict the quarterly average airfare price with an adjusted R-squared score of 0.869.[4]

According to economic theory, the pass-through of airline cost changes strongly depends on the type of cost change (firm-specific or sector-wide) and market conditions. In monopolistic markets, most or all of a cost change may be passed through, depending on the shape of the relationship between prices and demand.[5]

IV. SYSTEM ARCHITECTURE

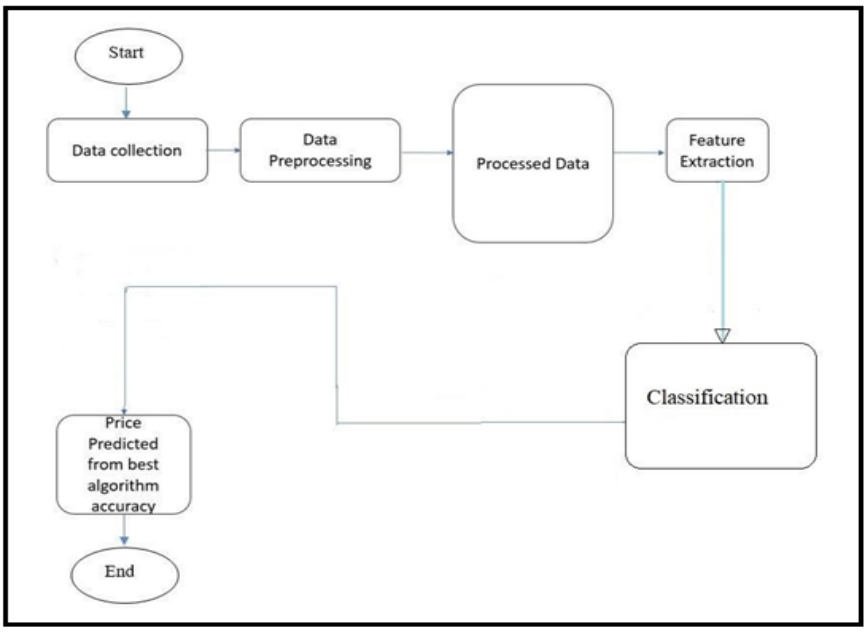

Fig 1: Block Diagram of Flight Price Detection

Data pre-processing is a data mining approach that details putting raw data into comprehensible format. Data cleansing technique include filing in missing number, smearing noisy data, and resolving discrepancies in the data. The data set is cleaned & decimal values are changed into appropriate folate values because it has some missing values. The data attribute that where utilise to train machine learning modules have a significant impact on the model's performance. Model performance maybe adversely affected by irrelevant or only partially relevant features. Classification: by adjusting the training set to the classifier model, the model is trained. After testing the classifier model assigns the air quality a good or bad rating. The categorisation closely resembles the testing set. For this project, we used python libraries like Pandas, Numpy,Matplotlib,Seeborn & learn to implement the machine learning life cycle to create a simple web.

V. IMPLEMENTATION

Data pre-processing: It is a technique used in data mining that involves transforming raw data into an understandable format. The data is cleansed through processes such as filling in missing values, smoothing the noisy data, or resolving the inconsistencies in the data. As it contains some missing value, the dataset is cleaned, and decimal values are converted into proper float values.

Get the Dataset: To create a machine learning model, the first thing we required is a dataset as a machine learning model completely works on data. The collected data for a particular problem in a proper format is known as the dataset. Dataset may be of different formats for different purposes, such as, if we want to create a machine learning model for business purpose, then dataset will be different with the dataset required for a liver patient. So, each dataset is different from another dataset.

Importing Libraries: In order to perform data pre-processing using Python, we need to import some predefined Python libraries. These libraries are used to perform some specific jobs. There are three specific libraries that we will use for data pre-processing.

Importing the Datasets: Now we need to import the datasets which we have collected for our machine learning project. But before importing a dataset, we need to set the current directory as a working directory. To set a working directory in Spyder IDE.

Handling Missing data: The next step of data pre-processing is to handle missing data in the datasets. If our dataset contains some missing data, then it may create a huge problem for our machine learning model. Hence it is necessary to handle missing values present in the dataset.

The next step of data pre-processing is to handle missing data in the datasets. If our dataset contains some missing data, then it may create a huge problem for our machine learning model. Hence it is necessary to handle missing values present in the dataset accuracy. The testing set is never used for training, which could otherwise lead to overfitting the mode

Train Set: The train set would contain the data which will be fed into the model. In simple terms, our model would learn from this data. For instance, a Regression model would use the examples in this data to find gradients in order to reduce the cost function.

Dev Set: The development set is used to validate the trained model. This is the most important setting as it will form the basis of our model evaluation. If the difference between error on the training set and error on the dev set is huge, it means the model as high variance and hence, a case of over-fitting.

Test Set: The test set contains the data on which we test the trained and validated model. It tells us how efficient our overall model is and how likely is it going to predict something which does not make sense. There are a plethora of evaluation metrics which can be used to measure the performance of our model.

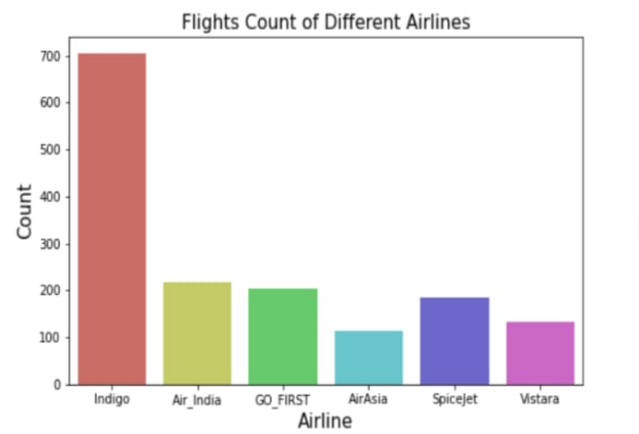

Fig 1: Flights count of Different Airlines

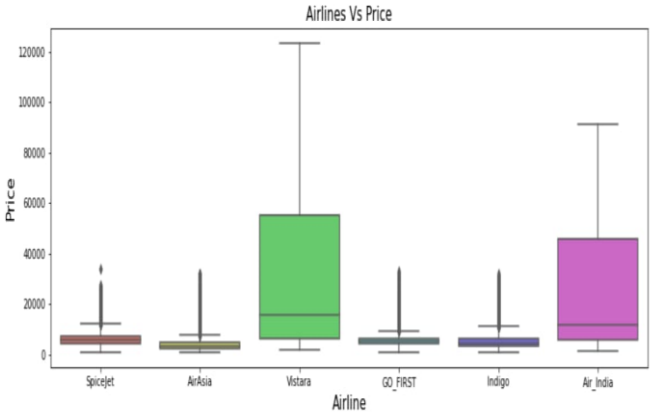

Fig 2: Airline Vs Price

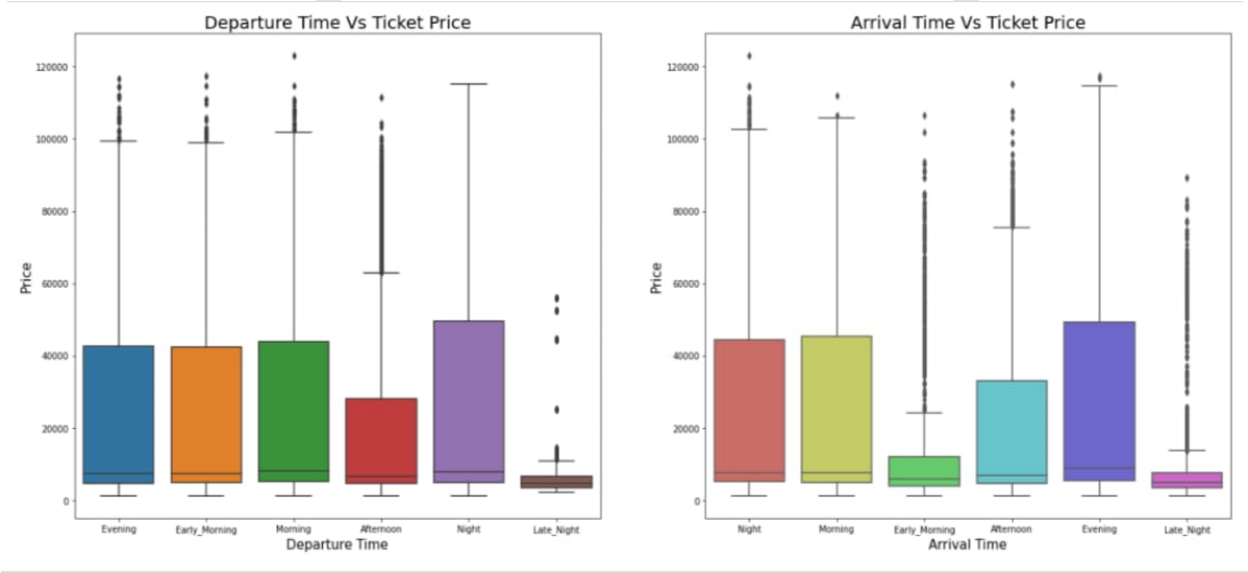

Fig 3: Departure time Vs ticket price, Arrival time Vs ticket price

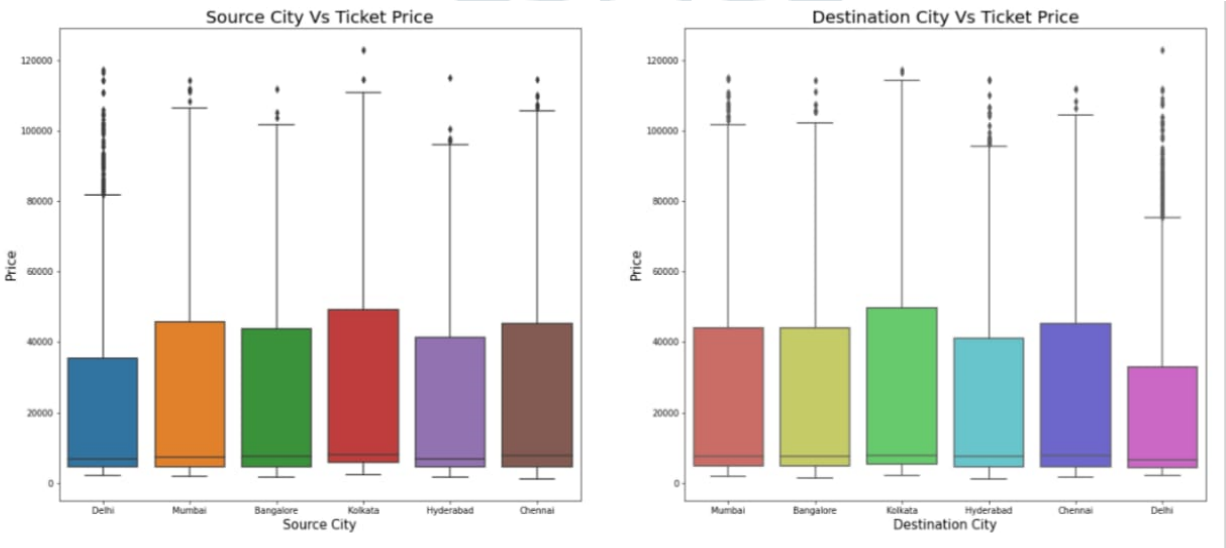

Fig 4: Source City Vs Ticket Price & Destination city Vs Ticket Price

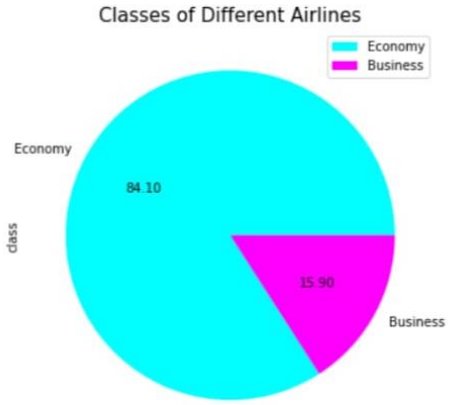

Fig 5: Classes of Different Airline

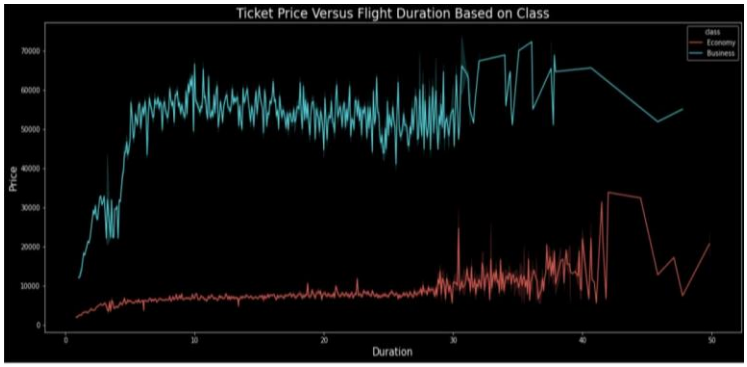

Fig 6: Ticket Price Vs Flight Duration Based on Class

The Fig 1 Matplotlib Graph Shows different airlines eg: Indigo,Air-india, Go-First,AirAsia,Spicejet,Spicejet,Vistara Flights Counts Based On Graphical Representation.in above Fig Shows Indigo airline Count is largest among all Airlines.Fig 2 Matplotlib Graph Shows Airline Ticket class Vs Price Count graph.in which 2 classes include Business & Economy. Business Tickets have largest price Count than economy Class. The Fig 3 Circular Matplotlib Graph Shows Percentage of Classes based on different Airline. Economy class have largest Percentage 84.10% & Business Class contain 15.90% out of 100%. Fig 4 Matplotlib graph represent. Departure Time Vs Ticket Prearrival Time Ticket Price. Which include Evening, early morning, Morning, afternoon, late night in Departure time & Night,morning,every morning Afternoon, late night include Arrival time. In Fig 5There are Two Matplotlib Graph Which include source City Vs Ticket Price & Destination city Vs Ticket Price.Sorce city vs Ticket price graph include delhi,Mumbai,banglore,Kolkata,Hydrabad,Chennai Source city. In Fig 6Matplotlib Graph Represent Ticket Price Vs Flight Duration Based on their classes.in which Economy classes represented by red lines & Business classes Represented by Blue lines. Business class has highest ticket price.

VI. RESULT

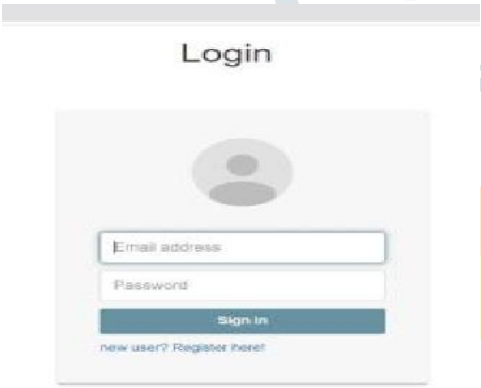

Fig 7: Login Page

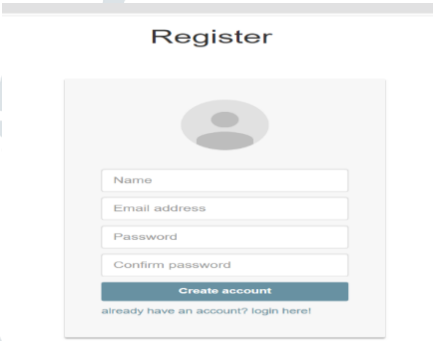

Fig 8 : Registration Page

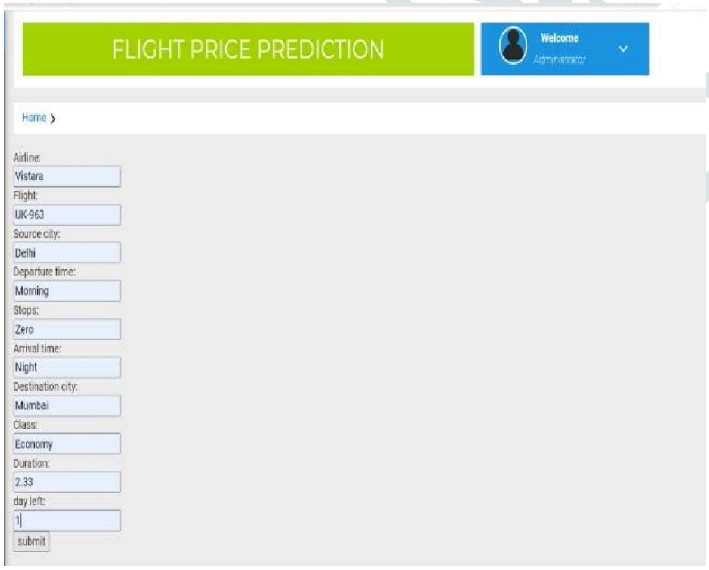

Fig 9: Description Page

Fig 10: Actual Result

Fig: 7 indicates it is a Login Page. The user will be able to login to the system after entering the correct email address i.e., username and password. It is developed using Flask and HTML. Fig 8 indicates it is a Registration Page. The user will be able to log in to the system only if he is a valid user or else, he has to register with the required system with various information for getting himself logged.

Fig 9 indicates the prediction page required details. Fig 10 indicates it is a Home Page where system results will be displayed. The user has to pass several inputs related to flight and click on submit button, after clicking on the submit button the system waits for the prediction result and once received it's been displayed on the page.

VII. CONCLUSION

This System Design for flight price predictions. It can give the predicted values of flight price to get a flight ticket at minimum cost. Data is collected from the websites know as Kaggle which has open access for data. assessed the structure for various datasets, with various parameters and machine learning techniques. Support Vector Machine and Random Forest as well as K-Nearest Neighbor Techniques are to be used in proposed system.

VIII. FUTURE SCOPE

Flight price prediction using machine learning has a wide range of future scope as it has the potential to revolutionize the travel industry. With the ever-increasing amount of data available, machine learning algorithms can analyse historical flight prices, weather patterns, seasonality, and other factors to predict future prices accurately. This technology can help airlines optimize pricing strategies, reduce revenue loss due to empty seats, and improve customer satisfaction. Additionally, travel booking platforms can use machine learning to provide personalized flight recommendations to their users based on their travel history and preferences. The future of flight price prediction using machine learning is promising, and it is expected to transform the travel industry by providing more affordable and convenient travel options for consumers.

REFERENCES

[1] Guan Gui and Fan Liu, *"A Framework for Airfare Price Prediction: A Machine learning Approach Economics and Statistics"*, vol. 83, no. 1, pp. 2022.

[2] N Lakshmi Kalyani and Bindu Sri Sai U, *"Machine Learning Model – based Prediction of Flight Delay"*, vol.45, no.6, pp. 1020-1029,2022.

[3] Jiage Huo and K. L. Keung, *"The Prediction of Flight Delay: Big Data-driven Machine Learning Approach"*, Tourism Management, vol. 31, no. 1, pp. 36-44, 2022.

[4] Tianvi Wang, Samira Pouvanfar, Haiman Tian,Miguel Alonso jr., Steven Luis and Shuc- Ching Chen *"flight Delay Prediction Based on Aviation Big Data and Machine Learning"* ,Journal of Transport Economics and Policy, vol. 7, no. 2, pp. 171-192, 2022.

[5] Viet Hoang Vu, Quang Tran Minh, Phu H. Phng, *"An Airfare Prediction Model for Developer Markets"*, of Revenue and Pricing Management, vol. 4, no. 1, pp. 39-53,2021.

[6] K.Tziridis, Th.Kalampokas, G.A. PapaKostas, K.I.Diamantars *"Air face Prices Prediction Using Machine Learning Techniques"* Journal of Transport Geography, vol. 14, no. 1,pp.15-22, 2021.

[7] vijayarangan Natarajan and Shubham Sinha, *"A Novel Approach: Airline Delay Prediction Using Machine Learning"*, *Studies*, vol. 42, no. 13, pp. 2371-2388, 2021.

[8] Suvojit Manna and Sanket biswas, *"A Statistical Approach to Predict flight delay using gradient boosted decision tree"*, the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1575-1583, 2021.

[9] Tao Liu, Jian Cao, Yudong Tan, quanwu Xiao, *"ACER:An Adaptive Context-Aware EnsSemble regression Model for airfare price Prediction"* , the 25th IEEE European signal processing conference, pp. 1036-1039, 2020.

[10] K. S. Gerardi and A. H. Shaoiro, *"Does competition reduce price dispersion? New evidence from the airline industry"*, the IEEE international conference on big data, pp. 964- 969, 2020.

[11] W. Groves and M. Gini, *"A regression model for predicting optimal purchase timing for Airline tickets"*, *IEEE* international conference on big data, pp. 964-969 2020.

[12] T. Wang et al, *"A Framework for Airfare Price Prediction: A Machine Learning Appro- Ach"*, *the* international conference on Progress in informatics and Computing, pp. 312- 317, 2020.

[13] T. Liu.J. Cao. S Feng and Y. Tan, *"An Adaptive context-aware ensemble regression model for Flight Fare Prediction Based on Kalman Filter"*, *Journal* of Air Transport Management, vol. 16, no. 4, pp. 169-177, 2019.

[14] M. S. Ryerson and H. Kim, *"Integrating airline operational practices into passenger airline hub definition"*, Journal of Transport Geography, vol. 31, pp. 84-93, 2019.

[15] H. Baik, A. A. Trani, N. Hinze, H. Swingle, S. Ashiabor and A. Seshadri, *"Forecasting model for air taxi commercial airline and automobile demand in the United States"* Transportation Research Record, vol. 2052, no. 1, pp. 9-20, 2019.

[16] T. Janssen, T. Dijkstra, S. Abbas and A. C. van Riel,*" A linear quantile mixed regression model for prediction of airline ticket prices"*, Radboud University, 2019.

[17] T. Wohlfarth, S. Clémençon, F. Roueff and X. Casellato, *"A data-mining approach to travel price forecasting"*, the 10th international conference on machine learning and Applications and workshops, vol. 1, pp. 84-89, 2018.

[18] H.-C. Huang, *"A hybrid neural network prediction model of air ticket sales"*, Telkomnika Indonesian Journal of Electrical Engineering, vol. 11, no. 11, pp. 6413-6419, 2017.

[19] E. J. Santana, S. M. Mastelini and S. Barbon, *"Deep regressor stacking for air ticket prices prediction"*, the XIII Brazilian symposium on information systems: information *systems* for participatory digital *governance*, pp. 25-31, 2017.

[20] S. Mumbower and L. A. Garrow, *"Data set – Online pricing data for multiple us carriers"*, Manufacturing & Service Operations Management, vol. 16, no. 2, pp. 198-203,2017.