ISSN: 0974-5823 Vol. 7 No. 6 June, 2022

International Journal of Mechanical Engineering

Prediction of Airfare Prices Using Machine Learning

[1]Mrs. Sowjanya, [2]Saud Ikram, [3]Rayyan Khan, [4]Shaik Haseeb Jawad, [5]Saad ul arifeen

[1]Assistant Professor, Department of Computer Science and Engineering, Lords Institute of Engineering and Technology, Hyderabad.

[2,3,4,5]Research Scholar, Department of Computer Science and Engineering, Lords Institute of Engineering and Technology, Hyderabad.

ABSTRACT— This study focuses on the issue of predicting flight prices. Assuming that these factors have an impact on ticket prices, a standard set of characteristics for a typical flight is selected for the survey. For the purpose of forecasting airline prices for tickets, these features are fed into eight cutting-edge machine learning (ML) algorithms [6]. The models' results are then compared. This study examines the relationship between the performance of a predictions and the feature set used to describe an airline, as well as the precision of every model's predictions. Information from Aegean Airlines flights from Thessaloniki to Stuttgart is used in the trials as a novel database for training each machine learning model. When it comes to this particular prediction problem, experimental results demonstrated that machine learning techniques are capable of handling it almost 88% of the time.

Index Terms— predictions, pricing models, and machine learning.

I.INTRODUCTION

In today's world, airlines employ a variety of complex strategies and procedures to dynamically allocate airfare pricing [1], [2]. A number of financial, marketing, commercial, and social considerations all play a role in the final cost of a flight. The dynamic nature of airline pricing makes it nearly impossible for passengers to find the best deal on a plane ticket due to the complexities of the pricing methods used by the airlines. As a result, a number of methods [3], [4] that can estimate the price of flight and give the customer with the appropriate time to acquire a ticket have been presented recently. Predictions from the evolutionary computing research topic widely used In machine Learning are used in most of these methods (ML). With 75.3 percent accuracy, Groves and Gini [4] used a PLS regression model to optimise the purchase of airline tickets (acc.). Ripple Drop Rules Learner (74.5 percent accuracy), Logistic Regression (69.9 percent accuracy), and Linear SVM (69.4 percent accuracy) ML models were used to forecast if the ticket price would drop in the future. Air ticket prices can be predicted with satisfactory accuracy using a linear autoregressive distributed lag mixture linear regression suggested by Janssen [6]. Linear Regression (77.06 percent accuracy), Nave Bayes (73.06 percent accuracy), Softmax Regression (76.84 percent accuracy) and SVM (80.6% accuracy for two bins) models were tested for their ability to forecast plane ticket costs. Some classical models, like as GARCH, were used in all of the previous studies to estimate airline ticket prices around the globe. The performance of current ML models on this task is, to the knowledge of the researcher' understanding, still undiscovered.

The following points outline the contributions of the recommendations section: For the first time, ticket price predictions in Greece has been attempted, as well as a research into the influence of characteristics on flight costs and a comparison of the current state-of-the-art Ml algorithms. Here are the rest of the sections: Section II provides an introduction to machine learning and how it can be used to solve the challenge of predicting airfare prices. From a theoretical viewpoint, Section III identifies the current study, whereas Section IV focuses on the experimental approach and outcomes of the models employed in the study. Finally, Section V summarises the findings of the study and suggests future research avenues.

II.RELATED WORK

As an instance of low-cost airlines pricing methods, Ryanair is examined

Ryanair and the low-cost operating model are examined in this chapter, as well as variable pricing approaches, in light of these two areas of aviation literature: The success of Europe's low-cost leader has been built on a foundation of low fares, and price formation is a hot topic of discussion among academics and industry professionals alike. Scientists have spent a lot of time studying the reduced business model's expense strategy. But the low-cost model's viability depends on the delicate balance between price levels, loadings and operational expenses [12]; the significance of the mixture of different decisions made by operators indicates that other parts of the low-cost marketing strategy should be examined. Even more crucial than minimising expenses is a company's ability to generate revenue and set prices, which require further research. While little is known about airline pricing fixing at the micro level, research has shown that the outcomes are strikingly different. The differences are based on the difficulty of taking into account the micro structure of low-cost pricing rather than the average fare and the limited collection of existing evidence collected by the researcher (most of the studies limited the extension of the sample, few fixed departing data, only one departing airport, a limited set of advancing booking price offered). This chapter tries to define the fundamental aspects of Ryanair's marketing strategy, as well as the competition and context factors that influence the selection of the average rates and their relative dynamics, in this framework.

Have the pricing policies of Ryanair shifted over time? Based on its 2006–2007 flights

Whether or not Ryanair's pricing models have evolved over time is a topic of discussion in this research paper. Over a two time periods, from 1 January 2006 to 31 December 2007, we compile a panel dataset of fares for all European flights operated by Ryanair [3]. When comparing "similar" flights, we determine the average rate for a 90-day period previous to departure, as well as the strength of variable pricing for each flight. Average rates and the degree to which dynamic pricing was utilised both declined in 2007 according to our findings [15]. There was an overall decrease in airfare of more around 10% for over one-third of the flights that were affected by the price cut. Ryanair appears to have softened its dynamic pricing actions on existing routes, which are generally deployed to promote more touristic demand, now that it's become the leading low-cost carrier in Europe. Booking in advance, on the other hand, is more expensive than booking on the spot.

Airline ticket purchase timing can be predicted using a regression model

Flight ticket purchasing from a customer viewpoint is problematic due to a lack of information regarding future price changes. Future prices can be predicted using a model that takes into account the probability of price changes [5]. Flights on specified routes and dates can be predicted based on a historical price quote corpus using the suggested approach. With our algorithm, we also use it to anticipate the costs of airfares with particular admirable characteristics, including tickets from a different airline or nonstop or multi-segment trips. Buyers can estimate the potential cost of their choices by evaluating models with various desirable attributes. On the basis of two high-volume routes, we calculated the estimated expenses of a variety of different options. These models' performance is improved by incorporating examples of night before going to bed features, classifying the raw features according to their degree of similarity, and then removing underperforming feature classes [11]. Using these models, we found that advising customers about purchase policies in the two months leading up to an anticipated resignation decreased the estimated price of purchases. When compared to an existing commercial web site, the proposed technique performs better than expected.

III.METHODOLOGY

Thessaloniki-Stadt flight Aegean Airlines [12] is initially chosen as a case subject for our investigation. First, a set of flight characteristics that have an impact on fares will be identified, followed by the gathering of sufficient flight data to train and test the machine learning models, followed by the selection of regression ML models for comparison and the experimental evaluation of those models. The following sections go into greater depth about each step in the processing process: There are two phases to the selection process: Stage 1 (Feature Extraction) and Phase 2 (Air Ticket Pricing). This phase is critical because it lays out the nature of the issue at hand. Flight departure and arrival times, as well as the amount of free bags that were available, were taken into account for each flight. A holiday day, an overnight flight, and the day of the week are all options. There are certain essential aspects according to the above list that would be assessed using a "one leave-out" rule to determine their impact. Please note that the characteristic F4 tells how many days there are between when you book your ticket and when you board your plane.

Data Collection - In this research, we're primarily concerned with estimating the cost of a single round-trip flight. Between December and July, flights from Thessaloniki to Stuttgart will be collected for the purposes of the tests. 1814 flights were documented and made available in [13] for the eight features (F1-F8) that were carefully gathered for each one of them. Selecting ML Models in Phase 3 - The same flight data was used to test eight cutting-edge regression machine learning models [8, [10], [14], [15]. In this study, the following ML models are compared:

Multiple-Layer Perception (MLP). Neural network for generalised regression. Extremely Powerful Learning Tool (ELM). Regression Tree using Random Forests. Tree of Regression. Regression Tree with Bagging. SVM with regressions (Polynomial and Linear). Regression Linearity (LR). Phase 2 flight data was used in a cross-validation technique to train the abovementioned Ml techniques using the 1814 flights recorded in phase 1. Prediction accuracy (percent - MSE between the intended and projected prices) and training time (in seconds) are the effectiveness indicators were using to provide a graphical representation.

This project has necessitated the development of the following modules:

Upload Airfare Prices Dataset: We could upload information to the system using this module.

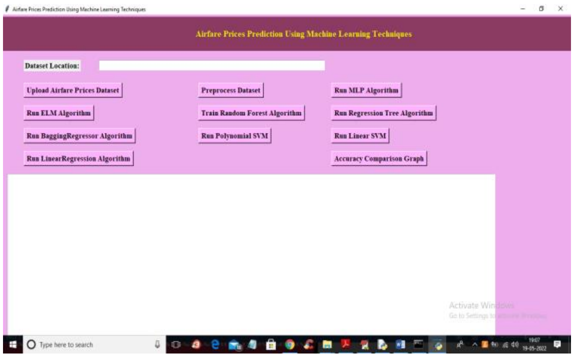
Pre-process Dataset: Our dataset comprises both numerical and non-numerical data, which will be converted to numerical data by applying Pre-processing in this module, which will then allow us to encode any non-numeric string data back into numerical data using the Label Encoder class.

Run MLP Algorithm: We'll partition the information into training and testing using this module, and then present this test data to the trained algorithm was applied its accuracy in prediction. Run all algorithms to see how accurate they are.

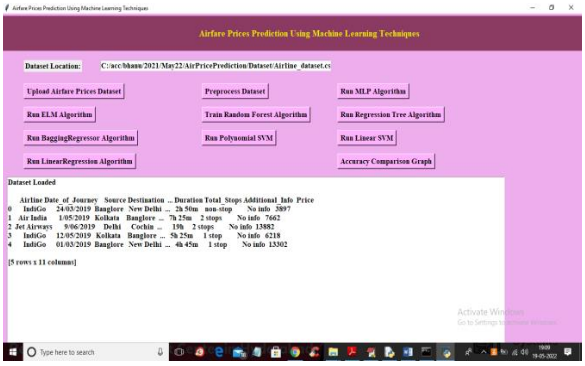
Accuracy Comparison Graph: With the help of this module, we can compare the accuracy of several methods.

IV.RESULT AND DISCUSSION

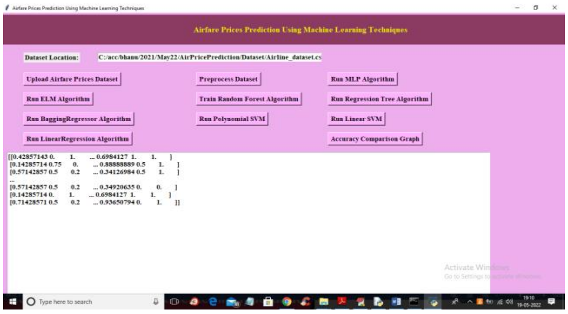
Run the project to get below output



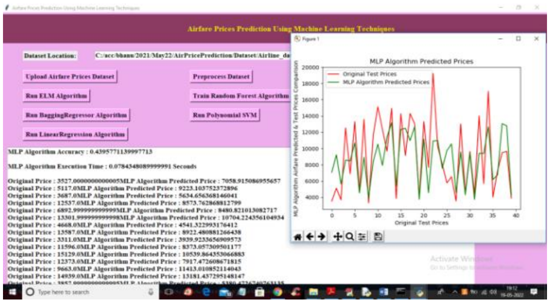
Upload the Airfare Prices Dataset after that. To load the dataset and obtain the following output, select and upload the database and then click 'Open.'



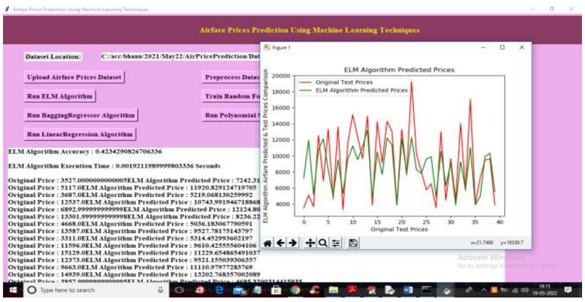
Clicking on "Preprocess Dataset" will encode any non-numeric data into numeric form, resulting in the output seen above



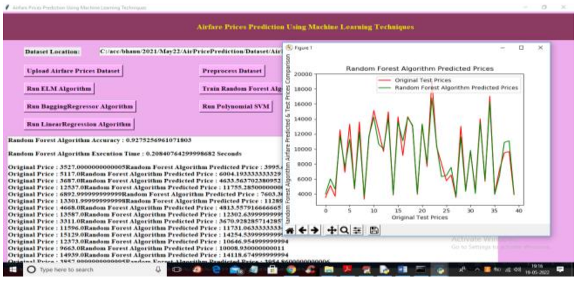
Click on the 'Run MLP Algorithm' tab to train MLP to apply learned model on test data to determine prices and estimate its prediction accuracy as shown in the above result.



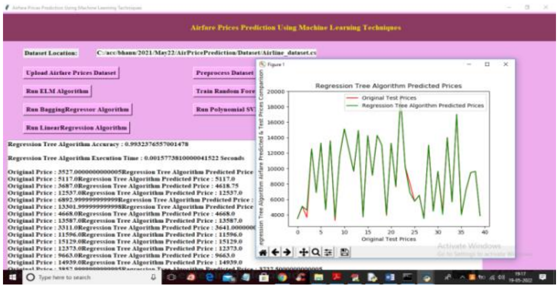
There is a 0.07 second execution time for MLP, and we can see the original TEST prices, as well as the values indicated by MLP and those same values that are plotted on the graph above. There is a large disparity between the red line representing initial Test prices and the green line representing predicted prices in this graph. Because of this, we have less accuracy, so we're going to shut the above graph and click the "Run ELM Algorithm" button to retrieve the output below.



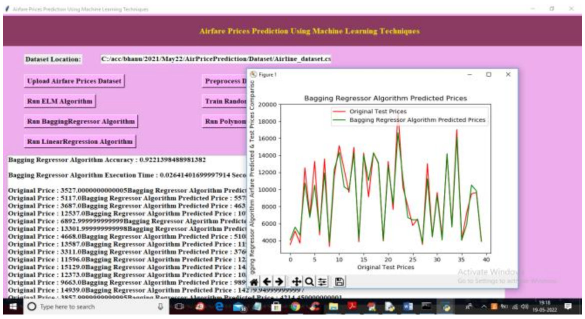
When we look at the ELM results above and click on Random Forest, we can see that ELM's accuracy is 0.42, and its predictions are also not very accurate.



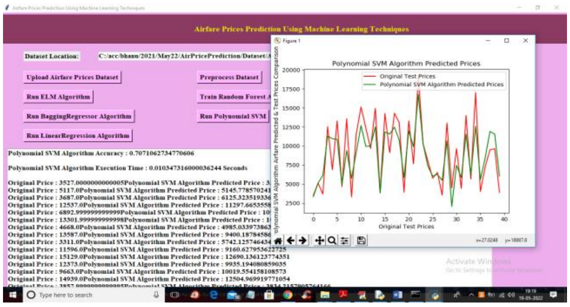
Random Forest yielded a prediction accuracy of 0.92 percent, and the graph shows that the prediction and test prices are nearly indistinguishable, so we can conclude that Random Forest's performance or prediction is excellent. To see the regression tree output, close the graph and click on the 'Regression Tree' button.



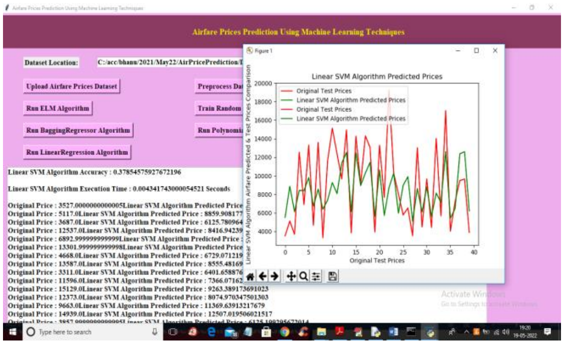
The 'Run Bagging Regression Algorithm' button will produce the following output, which demonstrates that our Regression Tree has an accuracy of 99.9 percent. The graph also shows that both lines closely match, indicating that the algorithm's performance is excellent.



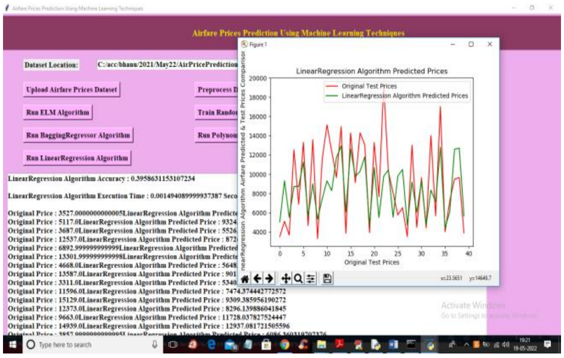
If you click on the 'Run Polynomial SVM' tab in the above results, you will see that we achieved a 0.92 percent accuracy rate.

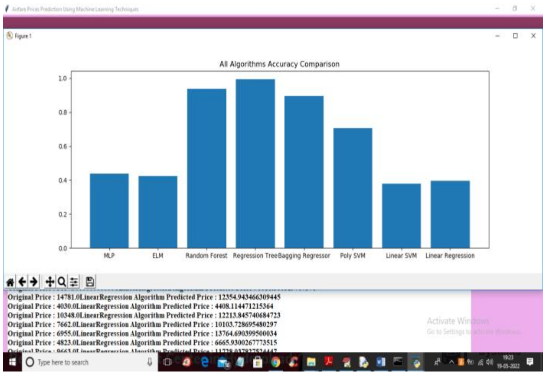


Run Linear SVM Algorithm tab at the bottom of the page to see what happens when we use Polynomial and then Linear SVM.



Using the Linear SVM, we obtained a 0.37 percent accuracy rate, and then clicking the 'Run Linear Regression Algorithm' tab, we received the following results.





There are x-axis labels for methodology names and y-axis values for algorithm accuracy in the above diagram, and Regression Tree had the highest accuracy across all algorithms.

V.CONCLUSION

A preliminary investigation in "airfare price forecasting" was published in this publication. On the internet, we found flight prices for the Greek airline Aegean Airline. Flight costs may be predicted based on historical data, according to the study historical prices. The test findings suggest that ML is effective. Algorithms are a useful tool for estimating the cost of airline tickets. Additionally, data plays a significant role in predicting airfares from which we took some of our information Conclusions that are worth noting. Which of the experiments led us to this conclusion? The most important factors in predicting airfare are characteristics.

There are additional features in addition to those selected. The accuracy of the forecast could be improved. This will be the case in the future. It is possible to modify the study to include predicting the cost of air travel to the airline's whole route map a few more experimentation on the initial pilot research is necessary, but larger data sets are required. Shows how machine learning techniques can be used to provide guidance customers to buy a plane ticket in the most advantageous market period.

VI.REFERENCES

[1] P. Malighetti, S. Paleari and R. Redondi, “Pricing strategies of low-cost airlines: The Ryanair case study,” Journal of Air Transport Management, vol. 15, no. 4, pp. 195-203, 2009.

[2] P. Malighetti, S. Paleari and R. Redondi, “Has Ryanair's pricing strategy changed over time? An empirical analysis of its 2006–2007 flights,” Tourism Management, vol. 31, no. 1, pp. 36-44, 2010.

[3] W. Groves and M. Gini, “A regression model for predicting optimal purchase timing for airline tickets,” Technical Report 11-025, University of Minnesota, Minneapolis, 2011.

[4] W. Groves and M. Gini, “An agent for optimizing airline ticket purchasing,” 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013), St. Paul, MN, May 06 - 10, 2013, pp. 1341-1342.

[5] M. Papadakis, “Predicting Airfare Prices,” 2014.

[6] T. Janssen, “A linear quantile mixed regression model for prediction of airline ticket prices,” Bachelor Thesis, Radboud University, 2014.

[7] R. Ren, Y. Yang and S. Yuan, “Prediction of airline ticket price,” Technical Report, Stanford Univerisy, 2015.

[8] S. Haykin, Neural Networks – A Comprehensive Foundation. Prentice Hall, 2nd Edition, 1999.

[9] S.B. Kotsiantis, “Decision trees: a recent overview,” Artificial Intelligence Review, vol. 39, no. 4, pp. 261-283, 2013.

[10] G.B. Huang, Q.Y. Zhu and C.K. Siew, “Extreme learning machine: Theory and applications,” Neurocomputing, vol. 70, no. 1-3, pp. 489- 501, 2009.

[11] G.A. Papakostas, K.I. Diamantaras and T. Papadimitriou, “Parallel pattern classification utilizing GPU-Based kernelized slackmin algorithm,” Journal of Parallel and Distributed Computing, vol. 99, pp. 90-99, 2017.

12] Aegean Airlines, https://en.aegeanair.com.

[13] https://github.com/humain-lab/airfare_prediction.

[14] L. Breiman, “Random forests,” Machine Learning, vol. 45, pp. 5-32, 2001.

[15] L. Breiman, J. Friedman, R. Olshen and C. Stone, Classification and Regression Trees. Boca Raton, FL: CRC Press, 1984.