Email: editor@ijerst.com or editor.ijerst@gmail.com

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

Predicting Flight Tickets using Gradient Boosting Regressor vs. Linear Regression

[1] Mrs.Zunaira Begum,
[2] MOHAMMED SAMEE UDDIN, [3] MOHAMMED FAZIL UZ ZAMA, [4] SYED RAHMAN PASHA,
[1] Assistant Professor, Department of AIML, Lords Institute of Engineering & Technology.
[234] Student, Department of AIML, Lords Institute of Engineering & Technology.

Abstract-

This project's overarching goal is to use Gradient Boosting to make airfare predictions for future ticket purchases. In contrast to fresh linear regression, regression device learning is a collection of guidelines. Here are the materials and methods used: the New Linear Regression Algorithm and Gradient Boosted Regression. The sample size was 10. Two companies use this image to compute these algorithms, which are based on a total of twenty algorithm samples. A G Power value of 80% was used to compare the sample to a control group, and a sample size of 10 was calculated. Gradient Boosting regression (82.5% accuracy) trumps new Linear Regression (62.5%) in terms of attained values. 8%. Reason being, compared to new linear regression, Gradient Boosting regression yields better results. A one-tailed test revealed that the new linear regression technique has a 0.00% statistically significant difference from the Gradient Boosting Regressor. The significance level at which this result was reached was p 0.05. Finally, after analyzing all of the approaches, it was found that the airfare prediction outperformed the new linear regression. After completing all of the steps, this became apparent.
Keywords: Reliability, Machine Learning, Gradient Boosting Regression, Novel Linear Regression, Flight Ticket Prediction, Flight Fare, and Flight Fare Prediction.

INTRODUCTION

Right now, flying is among the most costly and quickest ways to go from point A to point B. Here is an example: Here is an example: An example of this Machine learning algorithms are expected to keep improving the accuracy of their value estimates (National Research Council et al. 2003). A lot of the effort that goes into creating predictions is done by machine learning and the algorithms that power it. This is due to the fact that the algorithms give efficient and reliable results by analyzing shopping chart trends and prior airfare data (Zhao et al., 2021). (Yu 2021). Consequently, we will be implementing a system that uses machine learning to predict airfare in order to do this task. These algorithms learn from the patterns and data of previous match prices, allowing them to provide accurate and efficient outcomes. In order to determine whether algorithm is better at forecasting future travel expenditures, this study employs two separate methods, namely Gradient Boosting Regression and Comparison with New Linear Regression (Lok 2018). Google Scholar published about 200 publications and IEEE published approximately 600 articles on future airfare forecasts throughout the last five years, from 2017 to 2021. A significant amount of investment has been done in this field. A new linear regression algorithm is developed and an analysis of the Gradient Boosting Regressor technique is carried out using an empirical methodology (Purey and Patidar 2018). Both of these are assessed based on how well they perform. The new model outperforms its predecessor in terms of accuracy of output. Full execution of the experiment confirmed that the proposed algorithmic strategy improved accuracy on par with that of the previous model (Tavana, Nedjah, and Alhajj 2020). Our goal is to identify a workable solution, and to that end, we will evaluate the new linear regression algorithm against the time-tested Gradient Boosted Regression technique. After discussing the two machine learning algorithms in the prior section, this portion of our poll will focus on their pros and cons (Deepak, John Justin Thangaraj and Rajesh Khanna 2020). on compare the two approaches, one must first apply gradient boosting regression on the data set, then observe the data using the new linear regression, and lastly plot the findings. Specifically, you may do this by first running the dataset with gradient boosting regression, then seeing the results through the new linear regression, and lastly Now that we have a mechanism, we can more accurately predict how much aircraft tickets will cost in the future.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

MATERIALS AND METHODS

Currently, the study is taking place in the Machine Learning Laboratory at the Saveetha Institute of Technology in Chennai. This institute also houses the Saveetha Institute of Medical and Engineering Sciences. In a supervised learning setting, we compared the efficiency of two controllers to determine the optimal sample size using the G Power program. Twenty sets are used in this study, with ten sets used for each of the two methods of novel linear regression and gradient boosting. This sum of sets is called the "sets chosen." G Power 3.1 (G power settings: Statistical Test difference) is used to calculate the pre-test T-power values. Two algorithms—Gradient Boosting Regressor and a novel linear regression algorithm—implemented using technical analysis software; between the two means, p=0.05, and p=0.80. [G power settings: two algorithms (Gradient Boosting Regressor and a new linear regression) with a significance level of 0.05 and a power of 0.80. This research did not need an ethical committee's approval as it did not utilize any human or animal products. Assumption-Based Regression with Gradient Boosting from Sklearn.ensemble Import precision_score from sklearn.metrics into the Random GradientBoostingRegressor. Fit the gradient to the data set X and y using the GradientBoostingRegressor function. The predicted gradient is equal to the gradient.predict(X test score). gradient[X train, y train] = score print('accuracy score in whole:', score) while rounding(score*100,2) prints "accuracy score in percent:" In order to arrive at a final prediction, Gradient Boosting Machines (GBMs) combine the results of many decision trees (Rafique et al. 2020). Always keep in mind that the Gradient Boosting Machine uses decision trees for its underperforming pupils. You may anticipate the targets' continuous and categorical features with the help of gradient-boosting methods (Harrington 2012). The cost function for the variable is the logarithmic loss when it is used as a classifier, while it is the mean squared error (MSE) when it is used as a regressor. An Unofficial Program for New Linear Regression Given by Sklearn. importing a linear model Predictive Models The result of a linear regression is linear. Assign the values of X and y to the train's fit function.

One might say that linear prediction is just that: linear. X exam prediction line of best fit. Finding the score for the X train and the y train. print "total accuracy score:", print "accuracy score процент:", then start a new loop iterating from 100 to 2. According to Chandler, Pachter, and Mears (1993), contemporary linear regression models provide equations that describe the relationship between independent and dependent variables. Here is the simplest way to express a regression equation with a single dependent variable and an independent one:

$$Y=M*X+C \ldots \quad (1)$$

The regression coefficient, denoted as m, is used to estimate the dependent variable Y. (also referred to as the slope), where c is a constant and x is an independent variable. The line's slope is another name for the slope. To rephrase, y is the result of combining m, x, and care, which are the inputs. The most recent linear regression technique attempts to predict trends and values in the future.

$$j(\theta) = i/2m\ {}^{m}\Sigma i=1\ (h_\theta(x^{(i)})-y^{(i)})^2 \ldots \quad (2)$$

This function is also known as the "Error Squared Function" or the "Root Mean Squared Error." Since the derivative of a quadratic function eliminates half of the calculations required for gradient descent, this is the case. The hardware configuration includes 8 GB of RAM and an Intel Core I5 CPU. There was a 1 terabyte hard drive and a 64-bit operating system. The operating system used was Windows 10, and the implementation was carried out using Jupyter Notebook, which contains Python. Kaggle, a free online community for data science and machine learning enthusiasts, is where the material was retrieved. The source from which this data was extracted is: The given URL is https://www.kaggle.com/nikhilmittal/flight-fare prediction-mh. Identified and described for each of the eleven variables comprising the dataset. 1. The name of the airline that was used for the journey. 2. Trip Date: The trip's actual date. Third, the origin: where the flight will begin. Fourthly, the location of the flight's last stop. 5. Itinerary: an airline-required standard document detailing the start and finish points of a trip. 6. Dep_Time: When the flight leaves from the origin. When the plane lands at its destination is the seventh component of the arrival time. 7. Length: Time spent in flight (in hours or minutes). 9. Total Stops: The sum of all the times the aircraft landed before taking off. 10.1. Additional Info: It shows further flight details. Eleventh, Cost: How much the plane ticket does cost. Data Analysis by Statistic One may do an interactive or batch analysis of the data's statistical significance with the

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

help of the SPSS Statistics software package. It is now feasible to analyze the statistical significance of methods based on linear regression and gradient regression with the help of the SPSS software package. The independent variables are the price of the ticket, the route, and the date, while the dependent variables are the number of flights and the flight numbers themselves. Group statistics and independent sample tests were conducted based on the experimental data, and graphs representing the two parameters under investigation were created.

RESULTS

The results of a 500-person simulation using the Proposed Gradient Boosting Regressor technique and the current Novel linear regression system ran in parallel in a Jupyter notebook are detailed in Table 1. Finding out which of the two systems would work better was the primary motivation for doing the simulation. Multiple periods throughout the day were used to conduct the simulation. Table 1 shows that whereas the new linear regression method only managed a 62.8% overall accuracy, the Gradient Boosting Regressor approach averaged an 82.5% success rate. With a dataset size of 9650, 70% training data, and 30% testing data, Linear Regression achieved an accuracy of 82.52% in sample 1 and Gradient Boosting Regression achieved an accuracy of 62.81% in sample 1. The results are shown in Table 1.

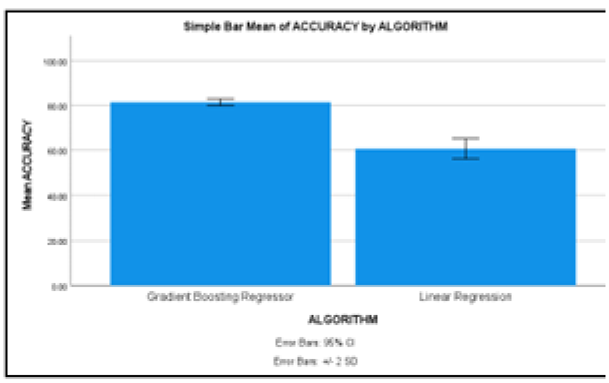
Sample (N)	Dataset Size	Gradient Boosting Regressor Accuracy in %	Novel linear Regression Accuracy in %
1	9650	82.52	62.81
2	8500	82.49	62.58
3	7900	82.05	62.24
4	7000	81.68	62.05
5	6500	81.52	61.85
6	5500	81.25	61.68
7	4000	81.02	61.25
8	3500	80.96	58.96
9	1500	80.65	57.02
10	1000	80.32	56.85

Table 2 shows that the mean gradient boosting regression technique also has a standard deviation of 1.32783, which is the same as the new linear regression strategy. There is a mean gradient boosting regression technique shown in Table 2. With a standard deviation of 1.21607, the mean gradient boosting regression approach is on par with the newly-developed linear regression algorithm in terms of accuracy and precision. The findings show that the new Linear Regressor approach only has a degree of accuracy of 81.5%, whereas the Gradient Boosting Regressor algorithm has a far greater degree of accuracy of 82.5%. of The average, dispersion, and margin of error for all the data points are shown in Table 3. The mean values for the data obtained from each study group were determined using independent Student's t-tests. The effect size of the Gradient Boosting Regressor approach is 0.826, whereas that of the new linear regression method is 0.050. This difference is the main reason why these two algorithms are so dissimilar. Verification of Means Equivalence using T-test Third Table. The significance and standard error are determined using T-tests on independent samples. The value of wet basis P is less than 0.05.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

Accuracy	Levene's test for equality of variances		T-test of Equality of Means					95% of the confidence interval of the Difference	
	F	Sig.	t	df	Sig (2-tailed)	Mean Difference	Std.Error Difference	Lower	Upper
Equal variance assumed	.050	.826	34.876	18	.000	19.85800	.56938	18.66177	21.05423
Equal variances			34.876	17.863	.000	19.85800	.56938	18.66111	21.05489

Figure 1 includes a comparison chart that shows how the new Linear Regression approach pales in contrast to the Gradient Boosting method. Therefore, we may conclude that the new Linear Regressor is not nearly as effective as Gradient Boosting Regressor. Below in Figure 1 you can see the resulting graph. The pictures are below in the article's footer.



See Figure 1 for a mean-and-accuracy comparison of the Gradient Boosting Regressor and the Novel linear Regression method. These two approaches are examples of regression procedures. In terms of mean accuracy, the Gradient Boosting Regressor outperforms the Novel Linear Regression by a wide margin (4.998 x 0.56938 = 18.66111 x 21.05489). On one side, we can see a side-by-side comparison of Innovative Linear Regression and Gradient Boosting Regression, and on the other, we can see the mean detection accuracy relative to two standard deviations.

In order to improve the efficiency of the models and provide more accurate predictions about flight costs, additional regressors based on linear regression and gradient boosting have been added to the previously constructed models. Our research has shown, however, that when it comes to large data sets, Gradient Boosting Regressor outperforms the new linear regression in terms of accuracy and effectiveness. The fact that we found the Gradient Boosting Regressor to be more accurate and effective led us to this conclusion. Using the Gradient Boosting Regressor approach to enhance ticket prediction was suggested in a recent research report. The method's execution may benefit from this. With gradient boosting regressors, models can handle massive amounts of data with ease. Here is an example: In order to forecast a company's ticket price, we created a technique called Gradient Boosting Regressor that looks at the daily movement of tickets bought from many airlines. Any business may use this strategy to predict how much a ticket will cost. Furthermore, enhancing the accuracy of ticket prediction is not a job that is well-suited to the most recent linear regression algorithms. The suggested Gradient Boosting Regressor and a novel linear regression technique (Ataman and Kahraman 2021) are the only two articles in the previous debate that are guaranteed to give improved performance when it comes to improving the accuracy of aircraft ticket prediction. Ataman and Kahraman were the researchers behind these two studies. However, improving the accuracy of airline ticket prediction is the shared goal of the newly constructed linear regression technique and the suggested Gradient Boosting Regressor. Furthermore, the latest cost estimates do not account for any hidden costs, an issue that has recently gained a lot of attention because of how prominent it is in people's thoughts. ticket Panwar et al. (2021) suggest using a new linear regression method and the Gradient Boosting Regressor technique to improve aircraft prediction accuracy. Despite the limitations of airline forecasting in terms of price prediction, these abilities are reliant on large margins of future pricing, which opens the door to more accurate price forecasting down the road. In order to implement deep learning, the underlying algorithms may factor in predictions.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

CONCLUSION

The data analysis will help the research achieve its primary goal of determining the accuracy of ticket estimations. This research article compares and contrasts two popular regression models: GradientBoosting and the new linear regression. In comparison to the new linear regression's accuracy of 62.8%, the findings show that GradientBoosting regression achieves an impressive 82.5%. DECLARATIONS Potential Forgery: This paper does not include any conflicts of interest. Works Cited by the Author All stages of data collection, analysis, and publishing were supervised by the author NSSVS. Among JJT's many contributions to the work are the study's conception, data validation, and critical evaluation of the text. Praise and Honor The author would like to use this opportunity to express their deepest appreciation to the Saveetha Institute of Medical and Technological Sciences and the Saveetha School of Engineering for providing the necessary resources to complete this project. Financial Support The following organizations generously provided the funding that allowed us to complete the research. We are really grateful to them. 1. The Chennai-based Qbec Infosol. 2. Saveetha Faculty of Law #3. Saveetha Institute of Health and Technology 4. Engineering at Saveetha School

REFERENCES

[1]. Ataman, Görkem, and Serpil Kahraman. 2021. "STOCK MARKET PREDICTION IN BRICS COUNTRIES USING LINEAR REGRESSION AND ARTIFICIAL NEURAL NETWORK HYBRID MODELS." The Singapore Economic Review. https://doi.org/10.1142/s0217590821500521 .

[2]. Chandler, P. R., M. Pachter, and M. Mears. 1993. "Constrained Linear Regression for Flight Control System Failure Identification." 1993 American Control Conference. https://doi.org/10.23919/acc.1993.4793484.

[3]. Deepak., S. John Justin Thangaraj, and M. Rajesh Khanna. 2020. "An Improved Early Detection Method of Autism Spectrum Anarchy Using Euclidean Method." In 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). https://doi.org/10.1109/i-smac49090.2020.9243361.

[4]. [4] IEEE. Harrington, Peter. 2012. Machine Learning in Action. Simon and Schuster. [5] 5 Lok, Johnny Ch. 2018. Prediction Factors Influence Airline Fuel

[5]. Price Changing Reasons. National Research Council, Transportation Research Board, Studies and Information Services, Division on Engineering and Physical Sciences, Aeronautics and Space Engineering Board, and Committee on Aeronautics Research and Technology for Vision 2050. 2003. Securing the Future of U.S. Air Transportation: A System in Peril. National Academies Press.

[6]. Panwar, Bhawna, Gaurav Dhuriya, Prashant Johri, Sudeept Singh Yadav, and Nitin Gaur. 2021. "Stock Market Prediction Using Linear Regression and SVM." 2021 International Conference on Advance Computing and Innovative Technologies in Engineering https://doi.org/10.1109/icacite51222.2021.9404733. (ICACITE).

[7]. Purey, Prateek, and Anil Patidar. 2018. "Stock Market Close Price Prediction Using Neural Network and Regression Analysis." International Journal of Computer Sciences and Engineering. https://doi.org/10.26438/ijcse/v6i8.266271.

[8]. Rafique, Muhammad, Aleem Dad Khan Tareen, Adil Aslim Mir, Malik Sajjad Ahmed Nadeem, Khawaja M. Asim, and Kimberlee Jane Kearfott. 2020. "Delegated Regressor, A Robust Approach for Automated Anomaly Detection in the Soil Radon Time Series Data." Scientific Reports 10 (1): 3004. Tavana, Madjid, Nadia Nedjah, and Reda Alhajj. 2020. Emerging Trends in Intelligent and Interactive Systems and Applications: Proceedings of the 5th International Conference on Intelligent, Interactive Systems and Applications (IISA2020). Springer.

[9]. 6 Yu, Jiangni. 2021. "A New Way of Airline Traffic Prediction Based on GCN-LSTM." Frontiers in Neurorobotics 15 (December): 661037.

[10]. Zhao, Zhichao, Jinguo You, Guoyu Gan, Xiaowu Li, and Jiaman Ding. 2021. "Civil Airline Fare Prediction with a Multi Attribute Dual-Stage Attention Mechanism." Applied Intelligence(Dordrecht, Netherlands), August, 1–16.