Original Article

Music Genre Classification using Linear Regression Compared with Extreme Gradient Boost Algorithm with Improved Accuracy

G. Roop Sumanth[1], Uma priyadarshini[2]

[1]Research Scholar, Department of Computer Science, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu. India. Pincode: 602105
[2]Project Guide, Corresponding Author, Department of Computer Science, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu. India. Pincode: 602105

Abstract

Aim: To classify music based on its genre Using Machine Learning Algorithm Extreme Gradient Boosting compared with Linear Regression.
Materials and Methods: The categorizing is performed by taking a Sample Size of n= 10 in Linear Regression and Sample size of n =10 in Gradient Boosting regression with the g-power value of 80% and Datasets containing recent study findings and a criterion of 0.05 percent, confidence interval of 95 % mean and standard deviation are collected from various web sources.
Results: The values obtained in terms of Accuracy are Identified based on the condiment Gradient Boosting Regression (59.0%) over Novel linear Regression (53.0%). The Novel linear Regression algorithm and Extreme Gradient Boosting were found to have a statistically significant difference ($p<0.05$).
Conclusion: The classification of music based on genre using Gradient Boosting Regression appears to be more accurate when compared to Linear Regression.

Keywords: Machine Learning Algorithm, Gradient Boosting Regression, Novel Linear Regression, Accuracy, Music Genre Classification, Decision tree, Binary classification.

https://doi.org/10.47750/pnr.2022.13.S04.198

INTRODUCTION

With the introduction of online music libraries and effortless access to music content at all times, The problem of managing the songs people listen to has risen. It is possible to organize and classify music based on its genre, which is defined by some musical characteristics such as harmonic content, rhythmic structure, and instrumentation. The ability to automatically classify and manage a user's music library makes audio streaming services like Spotify and iTunes beneficial(Wilkes, Vatolkin, and Müller 2021).

In the last 5 years 2017-2021, Google Scholar has published nearly 200 papers and the IEEE has published around 600 papers on Music Genre Classification which can be greatly invested. The analysis of the Extreme Gradient Boosting Regressor Algorithm and Linear Regression Algorithm in high-performance efficiency has been made using an experimental approach(Xie et al. 2021; Tian et al. 2019).

Our institution is passionate about high quality evidence based research and has excelled in various fields (Parakh et al. 2020; Pham et al. 2021; Perumal, Antony, and Muthuramalingam 2021; Sathiyamoorthi et al. 2021; Devarajan et al. 2021; Dhanraj and Rajeshkumar 2021; Uganya, Radhika, and Vijayaraj 2021; Tesfaye Jule et al. 2021; Nandhini, Ezhilarasan, and Rajeshkumar 2020; Kamath et al. 2020) The accuracy of the model is more than the existing model accuracy. The existence of the experiment is total and the improvement of accuracy of a proposed algorithm system compared to the existing model by improving. To overcome this problem, a Gradient Boosting Regression (Jome 2021)is used to improve the accuracy of the prediction and compare it with Novel linear Regression(Murphy 2017).Using the above two Machine Algorithms, utilizing their benefits and drawbacks in the current survey. On applying Gradient Boosting Regression to the dataset followed by performing observations using Novel linear Regression and the results were plotted on a graph the there two techniques are

compared based on the result. Finally Getting the best algorithm for the classification of upcoming music(Ajoodha 2014)

MATERIALS AND METHODS

The research work is conducted in the Machine Learning Laboratory at Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai. The sample size has been calculated using the SPSS software by comparing both of the controllers in Supervised learning. Each sample size is 10 sets for Gradient Boosting regressor and Novel linear Regression; a total of 20 sets is selected for this work. The pre-test power value is calculated using GPower 3.1 software (g power setting parameters: The difference of two independent means is taken as the statistical mean, α=0.05, power=0.80, Two algorithms (Gradient Boosting Regressor and Novel linear Regression algorithm) are implemented using Technical Analysis software.

CLASSIFIERS

Gradient Boosting Regression

One of the ensemble classifiers is boosting, which is achieved by merging several weak learners (such as decision trees). Boosting algorithms, on the other hand, are trained sequentially utilizing forward stagewise additive modeling, as opposed to RFs(Wade 2020).

The decision trees that are learned during the early iterations are quite simple. Because it is built to focus on the occasions where previous learners made mistakes, the classifier becomes more strong as training advances. The final prediction is a weighted linear combination of the individual learners' output after training. Extreme Gradient Boosting, or XGB, is a boosting implementation that allows the model to be trained quickly and in parallel.

Pseudocode

```
%%time
xgb_classifier = xgb.XGBClassifier(max_depth=5, learning_rate=0.1, objective='multi:softprob',
n_estimators=500, sub_sample = 0.8, num_class = len(label_dict))
xgb_classifier.fit(x_train, y_train)
pred_probs = xgb_classifier.predict_proba(x_test)
display_results(y_test, pred_probs)
with open('pred_probas/xgb_classifier.pkl', 'wb') as f:
   pickle.dump(pred_probs, f)
```

Novel linear Regression

For binary classification tasks, this linear classifier is commonly employed. The LR is implemented as a one-vs-rest approach for this multi-class classification assignment. That is, seven binary classifiers are trained separately. During testing, the predicted class is picked from among the seven classifiers with the highest probability(Montgomery, Peck, and Geoffrey Vining 2015; Ghildiyal, Singh, and Sharma 2020). The simplest form of the regression is defined in Equation 1, with one dependent and one independent variable is defined by the equation,

$$y=m*x+c \text{ —- } (1)$$

y is the estimated dependent variable, m is the regression coefficient or slope, x is the independent variable, and c is a constant. In other words, when m, x, and care are utilized as inputs, y is the output. The goal of novel linear regression is to forecast trends and future values.

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

This technique is also known and called the “Squared error function” or “Mean squared error”(Dua, Ghotra, and Pentreath 2016). In Equation 2, Because the derivative term of the square function will cancel out the 1/2 term, the mean is halved (1/2m) for the purpose of computing the gradient descent. The hardware configuration is an INTEL I7 core 9th generation processor with RAM 16GB. The system type used 64-bit and 1TB HardDisk. The operating system used was Windows 10 and the tool used for implementation was Jupyter Notebook with Python programming language.

Data was used from Kaggle, which is a freely available platform for data scientists and machine learning enthusiasts.

Pseudocode

```
%%time
lr_classifier = LogisticRegression(penalty='l2')
lr_classifier.fit(x_train, y_train)
pred_probs = lr_classifier.predict_proba(x_test)
display_results(y_test, pred_probs)
with open('pred_probas/lr_classifier.pkl', 'wb') as f:
pickle.dump(pred_probs, f)
```

Statistical Analysis

SPSS Statistics is a statistical analysis software package that allows you to perform interactively or batch statistical analysis. Novel Linear Regression (Wade 2020) and Gradient Boosting Regression-based Algorithms are now statistically analyzed using the software. The independent variable is Gradient Boosting Regression Accuracy and the Dependent variable is efficiency. The independent test analyses and calculates the accuracy of the Gradient Boosting Regression for both Methods.

RESULTS

Table 1 shows the simulation result of the proposed algorithm Gradient Boosting Regressor algorithm and the existing system Novel linear Regression were run at different times in the Jupyter notebook with a sample size of 500. From table 1, it was observed that the mean accuracy of the Gradient Boosting Regressor algorithm was 59.7% and the Novel linear Regression algorithm was 53.4%.

Using an independent variable T-test among the study groups, the Mean, Standard Deviation, and Standard Error Mean were determined. The Gradient Boosting Regressor algorithm produces a significant difference from the Novel linear Regression algorithm with a value of 0.826 and effect size=0.050.

Table 2 represents the Mean of Gradient Boosting Regressor algorithm which is better compared with the Novel linear Regression algorithm with a standard deviation of 1.21607 and 1.32783 respectively. From the algorithm and Novel linear Regression algorithm in terms of mean and accuracy. The mean results, the Gradient Boosting Regressor algorithm (59.7%) gives better accuracy than the Novel linear Regression algorithm (53.4%). Figure 1 gives the comparison chart of the Gradient Boosting algorithm which is better than Novel Linear Regression. Hence it is evident, that the conclusion of Gradient Boosting Regressor performs better than Novel Linear Regression. The plots that resulted are depicted in the picture below. The figure has been placed at the end of the paper. Table 3. Linear Regression and Extreme Gradient Boost findings with two-tailed significant values p= 0.01. Independent Sample T-test of the significance level Linear Regression and Extreme Gradient Boost results with two-tailed significant values p= 0.01.
With a 95 % confidence interval, both the Linear Regression and the Extreme Gradient Boost have a significance level of less than 0.05.

DISCUSSION

Novel Linear Regression and Gradient Boosting Regressor both were Implemented to classify music based on genre and to Improve the Accuracy of the existing Model. But from the obtained results in our paper, it is concluded that the Gradient Boosting Regressor is more efficient and accurate in classification compared with Novel linear Regression for the larger datasets(Ajoodha 2014).

In the recent survey, the Proposed Gradient Boosting Regressor Algorithm is a Promising option for Music Genre Classification. Gradient Boosting Regressor-based models have a lower error level than the Larger data. Proposed Gradient Boosting Regressor Algorithm for classification of music of selected genres by comparing different musical characteristics. Further, the Novel linear Regression algorithm is not suitable for Improving the Accuracy of Music genre classification. (Silver, Lee, and Childress 2016)

From the above discussion, only a few articles ensure that they provide better performance than the proposed Gradient Boosting Regressor and Novel linear Regression algorithm for improving the accuracy of Music genre classification. So that the proposed Gradient Boosting Regressor algorithm and Novel linear Regression Algorithm can be used to improve the Music genre classification accuracy. (Davis 2018).

CONCLUSION

The primary goal of the study is to observe the values and calculate the accuracy of Music Genre Classification. In this research paper, a GradientBoosting Regression model with Novel Linear Regression. The results obtained show that the GradientBoosting Regression has found 59.7% of accuracy in the Music Genre Classification than the 53.4% of the Novel Linear Regression.

DECLARATIONS

Conflict of Interest
No conflict of interest in this manuscript.

Author Contributions

Author GRS was Involved in data collection, data analysis, and manuscript writing. Author UP was involved in the conceptualization, data validation, and critical review of the manuscript.

Acknowledgement

The author would like to express their sincere gratitude to Saveetha School of Engineering, and Saveetha Institute of Medical and Technical Sciences for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Kaizentric Technologies Pvt.Ltd, Chennai
2. Saveetha University
3. Saveetha Institute of Medical And Technical Sciences
4. Saveetha School of Engineering

REFERENCES

1. Ajoodha, Ritesh. 2014. Automatic Music Genre Classification.
2. Davis, Rosalind M. 2018. Feature Extraction and Machine Learning Techniques for Musical Genre Determination.
3. Devarajan, Yuvarajan, Beemkumar Nagappan, Gautam Choubey, Suresh Vellaiyan, and Kulmani Mehar. 2021. “Renewable Pathway and Twin Fueling Approach on Ignition Analysis of a Dual-Fuelled Compression Ignition Engine.” Energy & Fuels: An American Chemical Society Journal 35 (12): 9930–36.
4. Dhanraj, Ganapathy, and Shanmugam Rajeshkumar. 2021. “Anticariogenic Effect of Selenium Nanoparticles Synthesized Using Brassica Oleracea.” Journal of Nanomaterials 2021 (July). https://doi.org/10.1155/2021/8115585.
5. Dua, Rajdeep, Manpreet Singh Ghotra, and Nick Pentreath. 2016. Machine Learning with Spark - Second Edition.
6. Ghildiyal, Anirudh, Komal Singh, and Sachin Sharma. 2020. “Music Genre Classification Using Machine Learning.” 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). https://doi.org/10.1109/iceca49313.2020.9297444.
7. Jome, Kaitlyn. 2021. Gradient Boosting Trees: A Beginner’s Guide For Gradient Boosting: Decision Tree Machine Learning Projects.
8. Kamath, S. Manjunath, K. Sridhar, D. Jaison, V. Gopinath, B. K. Mohamed Ibrahim, Nilkantha Gupta, A. Sundaram, P. Sivaperumal, S. Padmapriya, and S. Shantanu Patil. 2020. “Fabrication of Tri-Layered Electrospun Polycaprolactone Mats with Improved Sustained Drug Release Profile.” Scientific Reports 10 (1): 18179.
9. Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. 2015. Introduction to Linear Regression Analysis. John Wiley & Sons.
10. Murphy, Andrew. 2017. “Linear Regression (machine Learning).” Radiopaedia.org. https://doi.org/10.53347/rid-57103.
11. Nandhini, Joseph T., Devaraj Ezhilarasan, and Shanmugam Rajeshkumar. 2020. “An Ecofriendly Synthesized Gold Nanoparticles Induces Cytotoxicity via Apoptosis in HepG2 Cells.” Environmental Toxicology, August. https://doi.org/10.1002/tox.23007.
12. Parakh, Mayank K., Shriraam Ulaganambi, Nisha Ashifa, Reshma Premkumar, and Amit L. Jain. 2020. “Oral Potentially Malignant Disorders: Clinical Diagnosis and Current Screening Aids: A Narrative Review.” European Journal of Cancer Prevention: The Official Journal of the European Cancer Prevention Organisation 29 (1): 65–72.
13. Perumal, Karthikeyan, Joseph Antony, and Subagunasekar Muthuramalingam. 2021. “Heavy Metal Pollutants and Their Spatial Distribution in Surface Sediments from Thondi Coast, Palk Bay, South India.” Environmental Sciences Europe 33 (1). https://doi.org/10.1186/s12302-021-00501-2.
14. Pham, Quoc Hoa, Supat Chupradit, Gunawan Widjaja, Muataz S. Alhassan, Rustem Magizov, Yasser Fakri Mustafa, Aravindhan

Surendar, Amirzhan Kassenov, Zeinab Arzehgar, and Wanich Suksatan. 2021. "The Effects of Ni or Nb Additions on the Relaxation Behavior of Zr55Cu35Al10 Metallic Glass." Materials Today Communications 29 (December): 102909.
15. Sathiyamoorthi, Ramalingam, Gomathinayakam Sankaranarayanan, Dinesh Babu Munuswamy, and Yuvarajan Devarajan. 2021. "Experimental Study of Spray Analysis for Palmarosa Biodiesel-diesel Blends in a Constant Volume Chamber." Environmental Progress & Sustainable Energy 40 (6). https://doi.org/10.1002/ep.13696.
16. Silver, Daniel, Monica Lee, and C. Clayton Childress. 2016. "Genre Complexes in Popular Music." PloS One 11 (5): e0155471.
17. Tesfaye Jule, Leta, Krishnaraj Ramaswamy, Nagaraj Nagaprasad, Vigneshwaran Shanmugam, and Venkataraman Vignesh. 2021. "Design and Analysis of Serial Drilled Hole in Composite Material." Materials Today: Proceedings 45 (January): 5759–63.
18. Tian, Haoye, Haini Cai, Junhao Wen, Shun Li, and Yingqiao Li. 2019. "A Music Recommendation System Based on Logistic Regression and eXtreme Gradient Boosting." 2019 International Joint Conference on Neural Networks (IJCNN). https://doi.org/10.1109/ijcnn.2019.8852094.
19. Uganya, G., Radhika, and N. Vijayaraj. 2021. "A Survey on Internet of Things: Applications, Recent Issues, Attacks, and Security Mechanisms." Journal of Circuits Systems and Computers 30 (05): 2130006.
20. Wade, Corey. 2020. Hands-On Gradient Boosting with XGBoost and Scikit-Learn: Perform Accessible Machine Learning and Extreme Gradient Boosting with Python. Packt Publishing Ltd.
21. Wilkes, Ben, Igor Vatolkin, and Heinrich Müller. 2021. "Statistical and Visual Analysis of Audio, Text, and Image Features for Multi-Modal Music Genre Recognition." Entropy 23 (11). https://doi.org/10.3390/e23111502.
22. Xie, Xiangkun, Mingwei Yang, Shan Xie, Xiaoying Wu, Yuan Jiang, Zhaoyu Liu, Huiying Zhao, Yangxin Chen, Yuling Zhang, and Jingfeng Wang. 2021. "Early Prediction of Left Ventricular Reverse Remodeling in First-Diagnosed Idiopathic Dilated Cardiomyopathy: A Comparison of Linear Model, Random Forest, and Extreme Gradient Boosting." Frontiers in Cardiovascular Medicine 8 (August): 684004.

TABLES AND FIGURES

Table 1. Comparison between Linear Regression and Gradient Boosting Regression with N=10 samples of the dataset with highest accuracy of respectively 53.4% and 59.7% in sample 1 (when N=1) using the dataset size= 9650 and the 70% of training and 30% of testing data

Sample (N)	Dataset Size	Gradient Boosting Regressor Accuracy in %	Novel linear Regression Accuracy in %
1	9650	59.72	53.41
2	8500	59.24	53.12
3	7900	58.93	52.89
4	7000	58.65	52.05
5	6500	58.21	51.85
6	5500	57.87	51.68
7	4000	57.59	51.25
8	3500	57.15	48.96
9	1500	56.86	47.02
10	1000	56.53	46.85

Table 2. Statistical analysis of GBR and Novel linear Regression Algorithm. Mean accuracy, Std. deviation, and std. error values are obtained for 20 sample data sets.

Sample Group	N	Value of Mean	Stnd. Deviation	Stnd. Error Mean

Measure of Accuracy	GBR	10	57.2270	1.16276	.36770
	LR	10	49.4000	1.79352	.56716

Table 3. For the two groups, An unbiased sample T-test is used to determine significance and standard error.. P<0.05 for wet basis.

	Levene's Test for Equality of Variances		**T-test of Equality of Means**					**95% of the confidence interval of the Difference**	
			t	**df**	**Sig (2-tailed)**	**Means Difference**	**Std. Error Difference**		
	F	**Sig.**						**Lower**	**Upper**
Accuracy									
Equal Variance Assumed	2.237	.152	11.580	18	.001	7.82700	.67592	6.40694	9.2470
Equal Variance Not Assumed			11.580	15.430	.001	7.82700	.67592	6.3897	9.2642

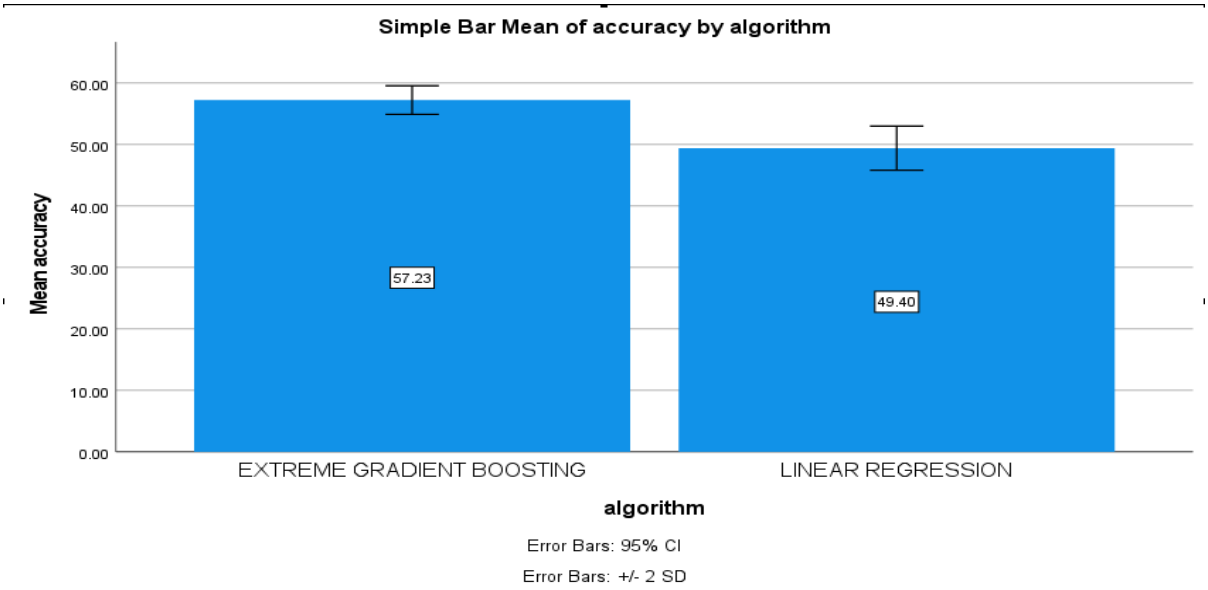

Fig. 1. Comparison of Gradient Boosting Regressor and Novel linear Regression algorithm in terms of mean and accuracy. The Gradient Boosting Regressor gave a better mean accuracy value than the Novel Linear Regression. X-axis: Gradient Boosting Regression vs Novel linear Regression, Y-axis: Mean accuracy of detection ±2 SD.