A Detailed Investigation on Drift Detection by using Modified Light Gradient Boost Model from Machine Learning Algorithm

A PROJECT REPORT

Submitted to

SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES

In partial fulfillment for the award of the degree of

BACHELOR OF COMPUTER SCIENCE AND ENGINEERING

By

N. Raja Likitha (191912152)

Supervisor

Dr.T.J.Nagalakshmi

SAVEETHA SCHOOL OF ENGINEERING,

SIMATS, CHENNAI - 602105

April-2023

SAVEETHA SCHOOL OF ENGINEERING
SAVEETHA INSTITUE OF MEDICAL AND TECHNICAL SCIENCES,CHENNAI - 602105

BONAFIDE CERTIFICATE

Certified that this project report "**A Detailed Investigation on Drift Detection by using Modified Light Gradient Boost Model from Machine Learning Algorithm**" is the Bonafide work of "**N. Raja Likitha**" who carried out the project work under my supervision.

SIGNATURE
Dr.T.J.Nagalakshmi

PROJECT GUIDE

PROFESSOR

Department of ECE

Saveetha School of

EngineeringSIMATS

UNIVERSITY EXAM DATE:

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

This project work would not have been possible without the contribution of many people. It gives me immense pleasure to express my profound gratitude to our Honorable Chancellor **Dr. N. M. Veeraiyan**, Saveetha Institute of Medical and Technical Sciences, for his blessings and for being a source of inspiration. I sincerely thank our Director of Academics **Dr. Deepak Nallaswamy,** SIMATS, for his visionary thoughts and support. I am indebted to extend my gratitude to our Director **Mrs. Ramya Deepak,** Saveetha School of Engineering, for facilitating us all the facilities and extended support to gain valuable education and learning experience.

I register my special thanks to **Dr. B. Ramesh,** Principal, Saveetha School of Engineering, Institute of Electronics and Communication Engineering, for the support given to me in the successful conduct of this project. I wish to express my sincere gratitude to my supervisor **Dr.T.J. Nagalakshmi**, for his inspiring guidance, personal involvement and constant encouragement during the entire course of this work.

I am grateful to Project Coordinators, Review Panel External and Internal Members and the entire faculty of the Department of Electronics and Communication Engineering, for their constructive criticisms and valuable suggestions which have been a rich source to improve the quality of this work.

N. Raja Likitha

TABLE OF CONTENTS

CHAPTER-1

Title page:

Improving Prediction Accuracy in Drift Detection Using Logistic Regression in Comparing with Modified Light Gradient Boost Model

N. Raja Likitha[1], T.J.Nagalakshmi[2]

N. Raja Likitha[1]
Research Scholar,
Department of Electronics and Communication Engineering,
Saveetha School of Engineering,
Saveetha University of Medical and Technical Sciences,
Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.
natharajalikitha19@saveetha.com

T.J.Nagalakshmi[2]
Project Guide, Corresponding Author,
Department of Electronics and Communication Engineering,
Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences,
Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.
nagalakshmitj.sse@saveetha.com

Keywords: Innovative Concept Drift, Data Stream mining, Drift Detection, Classifier, Technology, Deep Learning, Modified Light Gradient Boost Model, Logistic Regression.

ABSTRACT

Aim: The goal of the proposed work is improving prediction accuracy in drift detection using Logistic Regression compared with modified light gradient boost model. **Materials and Methods:** The collection of 40 samples were taken by varying test and training data set size. These samples are divided into Two groups (Group 1 - Logistic Regression, Group 2 - Modified Light Gradient Boost Model) each having 20 samples and the accuracy was calculated to quantify the improving prediction accuracy in drift detection using Logistic Regression and Modified Light Gradient Boost Model. The G power is taken as 80%. **Result:** The Mean accuracy results for the simulation is 0.646 for Logistic Regression, and the Modified Light Gradient Boost Model provides results with an Mean accuracy of 0.96708. The obtained 2 tailed significance is 0.0 which is less than ($p<0.05$). Therefore here statistical significance is observed between 2 groups. **Conclusion:** For the given dataset Logistic Regression Performs significantly less than the Modified Light Gradient Boost Model in the prediction.

Keywords: Innovative Concept Drift, Data Stream mining, Drift Detection, Classifier, Technology, Deep Learning, Prediction, Modified Light Gradient Boost Model, Logistic Regression.

INTRODUCTION

Drifting in data involves unforeseen changes in the statistical features of the target variable over time. The difficulty with this is that as time goes on, the predictions get less accurate. Lack of a suitable approach for handling the constant flow of data is a major issue with data stream analysis (Prentice and Zhao 2019). It is a challenging field for data stream mining since the method needs to identify changes quickly in order to obtain useful data from it (Song and Zhang 2008). It must maintain accurate statistics on this changing data. Similarly Innovative concept drift, as input changes, the predictive model should be updated(Yadav, Suresh Yadav, and Shyamala Bharathi 2022). The phrase "ensemble" is used to describe multiple models of ground methods that are implemented as an advice to enhance overall performance (Madigan, Gavrin, and Raftery 1995). A similar idea is used in data Stream mining, where ensembles are built using Bagging, Boosting, and Stacking (Bharathi* et al. 2019). Because of the impact on prediction model accuracy, changes in data ideas must be managed carefully. Such changes are detected by diffusion detection methods, letting this same predictive algorithm keep updating at the diffusion percentage (Levin and Yehudai 2017; Nagalakshmi et al. 2022). There some applications are drift mobile, drift life, drift connect in deep learning (Sammut and Harries 2017).

Google Scholar has almost ten thousand articles, Springer and IEEE Xplore in recent years related to this research. The most cited articles in IEEE Xplore are 1-25 of 412 and in Google Scholar 18,600 (Lughofer et al. 2015) In this citation they have done using classification-based stream mining in order to alert operators in the event of unintended system changes (Thomas 1992) here this site uses a radical drift time expansion chamber (Becker and Ebner 2019) here in this site they used collision detection on acceleration data (Shelmerdine et al. 2018) here by doing surveys in hospitals they give digital radiography systems with wireless detectors (Chua, Jordan, and Muller 2020) in this site by using online they use to update initially trained with offline labs.

Innovative Concept drift is the unanticipated change in the target variable's statistical characteristics over time that the model is attempting to predict in machine learning and predictive analytics. The difficulty with this is that as time goes on, the predictions get less accurate.

MATERIALS AND METHODS

A computer lab in the Department of VLSI of the Saveetha School of Engineering at the Saveetha Institute of Medical and Technical Sciences served as the setting for the study. The work is divided into two groups. The Logistic Regression model is in Group 1, and the Modified Light Gradient Boost

Model is in Group 2; each group has a sample size of 20, and the pretest power is 0.80.The samples of group 1 and 2 are taken by changing the test and training dataset size.

The data set is implemented using Google Collab together with code that was stimulated by Collab. The data is imported, and the data visualisation is completed. Following visualisation, the data set goes through a level of data preparation where the error numbers from the Google Collab drive are checked against the mounted code. The precision of the current classifier, Logistic Regression, is contrasted with the Modified Light Gradient Boost Model with deep learning.

Logistic Regression: Predicting a binary outcome, such as yes or no, with the help of previous observations from a data set is the goal of the logistic regression statistical analysis method. By examining the correlation between one or more pre-existing independent variables, a logistic regression model forecasts a complete data variable of deep learning. A logistic regression could be used in Innovative Concept drift, for instance, to forecast whether a candidate for office will win or lose, or if a high school student can be considered into a specific institution or not. These straightforward decisions among two alternatives enable for binary outcomes.

Modified Light Gradient Boost Model: The Train Using AutoML tool employs in this Technology the Modified Light Gradient Boost Model, a gradient boosting ensemble technique that is based on Logistic Regression. Light Gradient Boost Model is a decision tree-based technique that may be applied to both classification and regression problems. For excellent performance with dispersed systems, Modified Light Gradient Boost Model has been specially designed.

Testing Procedure

Figure 1. depicts the workflow process, in which Google Collab-generated code is used to implement the data set. The data has been imported, and the data visualisation is finished. Following visualisation, the data set is subjected to data preparation, in which the error numbers from the Google Collab drive are compared to the mounted code. The accuracy of the Logistic Regression was then compared to that of the current classifier, the Modified Light Gradient Boost Model.

STATISTICAL ANALYSIS

The proposed study work and the preceding work method's accuracy are both evaluated using the SPSS software package. Testing is carried out using a separate sample T-test. Group 1 has been taken as Logistic Regression . Group 2 is taken as a Modified Light Gradient Boost Model. The independent variables are baseline information, properties of images. The dependent variable is accuracy (Sayed-Mouchaweh, Zaytoon, and Billaudel 2011) .

RESULTS

Compare the accuracy percentage values of the Logistic Regression with Modified Light Gradient Boost Model both algorithms provide different Mean accuracy Modified Light Gradient Boost Model 0.96708 and Logistic Regression 0.646

Figure 2 shows the bar chart comparing accuracy values (Logistic Regression and Modified Light Gradient Boost Model), in that Logistic Regression gives accuracy about 0.646 where the Modified Light Gradient Boost Model gives about 0.96708. Here the error bar is +/-1 Standard Error. Figure 3: is the graphical representation with +/- 1 standard deviation.

Table 1 demonstrates the T-test tables. The Logistic Regression (0.646) classifier has a higher mean value when compared to the Modified Light Gradient Boost Model (0.96708) classifier, which was evaluated with the 20 number of samples per group. The means that were different for both of the classifiers present also revealed the standard deviation.

Table 2 shows the independent sample tests have been used in both divisions and it has discovered that precision (t = -47.46) & Mean Difference = (-0.321075) and it has its same criteria deviation variance of 0.006764. Between the two groups, there is a huge variance (means difference is -0.321075) ($p<0.05$).

DISCUSSION

The findings of the project study proved that the Modified Light Gradient Boost Model classifier outperformed the Logistic Regression algorithm with a 0.96708 accuracy rate in predicting the data with the accurate values that have the greater accuracy which is deemed to be better work. The prior findings demonstrate that when determining the Modified Light Gradient Boost Model's accuracy, the Logistic Regression support is not superior to it.

This site for Logistic regression nearly the accuracy is 74% using statistical significance and sample test are the technology used in this (Krishna, Vamsi Krishna, and Praveenchandar 2022); (Levin and Yehudai 2017; Nagalakshmi et al. 2022; Harinath* et al. 2019) . (Harini and Sashi Rekha 2022) In this article based on machine learning the accuracy of the Modified Light Gradient Boost Model is 91% (Batchu and Seetha 2022), this site gives a different set of accuracy in Modified Light Gradient Boost Model (MLGBM) the accuracy is 98% (Shobeyri 2023), in this article the Modified Light Gradient Boost Model algorithm accuracy percentage is nearly 59.7% even though using the technology (Terado and Hayashida 2020) in this site the average accuracy of Modified Light Gradient Boost Model is 96% and which supports the Modified Light Gradient Boost Model to improve in researching with deep learning.

In this site they have used different techniques like data stream mining to propose and utilise and detect the Modified Light Gradient Boost Model. They have tried to get the accuracy to 98% (Mahanty, Panda, and Mishra 2021) . (Felix, Yovan Felix, and Sasipraba 2019) in this site they have proposed the Modified Light Gradient Boost Model to get the dataset by using machine learning and accuracy of 95% (Brzezinski and Stefanowski 2014) This gives the accuracy of 95% to develop and change the accuracy of innovative Concept drift.

The problems experienced in this research work are sampling mismatch, data quality issues can occur more frequently than expected. There are numerous data quality issues, including incorrect input data and incorrect input processing steps and Anomalies can appear in both the training and target data sets. Anomalies can cause our data distributions to change.

Online learning is one technology for dealing with Innovative concept drift. where the model is re-trained on every observation. It is critical to create a repeatable process for identifying data drift of data stream mining, defining thresholds for drift percentage, and configuring proactive alerting so that appropriate action is taken. Here some examples to develop to future scope are changes in the data due to seasonality, changes in consumer preferences. By this technology it can improve and make it easy for users to use their technology.

CONCLUSION

Results for the Drift Detection for Logistic Regression is 0.646 mean accuracy and 0.96708 mean accuracy produced by the Modified Light Gradient Boost Model. The Modified Light Gradient Boost Model algorithm gives higher accuracy as compared to the Logistic Regression algorithm.

DECLARATIONS

Conflict of Interest

No conflict of interest in this manuscript.

Author Contributions

RL played a role in the planning, simulation, acquisition, evaluation, and drafting of the manuscript. TJNL contributed to the conceptualization, validation of data, and critical evaluation of the manuscript.

Acknowledgment

The authors express their gratitude to the Saveetha School of Engineering and the Saveetha Institute of Medical and Technical Sciences (formerly known as Saveetha University) for supplying the necessary resources that allowed for the successful completion of this research.

Funding

We would like to express our gratitude to the following organisations for giving the funding necessary for us to finish this research project.

1. Haneesh Sai Enterprises, Nellore.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

REFERENCES

Batchu, Raj Kumar, and Hari Seetha. 2022. “A Hybrid Detection System for DDoS Attacks Based on Deep Sparse Autoencoder and Light Gradient Boost Machine.” Journal of Information & Knowledge Management. https://doi.org/10.1142/s021964922250071x.

Becker, Felix, and Marc Ebner. 2019. “Collision Detection for a Mobile Robot Using Logistic Regression.” Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics. https://doi.org/10.5220/0007768601670173.

Bharathi*, Dr P. Shyamala, P. Shyamala Bharathi*, -Department of Electronics and Communication Engineering, Saveetha School of Engineering, SIMATS, Chennai, Tamilnadu, et al. 2019. “Resource Allocation by Demand Based Optimization and Machine.” International Journal of Innovative Technology and Exploring Engineering. https://doi.org/10.35940/ijitee.l3934.1081219.

Brzezinski, Dariusz, and Jerzy Stefanowski. 2014. “Reacting to Different Types of Concept Drift: The Accuracy Updated Ensemble Algorithm.” IEEE Transactions on Neural Networks and Learning Systems. https://doi.org/10.1109/tnnls.2013.2251352.

Chua, Adelson, Michael I. Jordan, and Rikky Muller. 2020. “Unsupervised Online Learning for Long-Term High Sensitivity Seizure Detection.” Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference 2020 (July): 528–31.

Felix, A. Yovan, A. Yovan Felix, and T. Sasipraba. 2019. “Flood Detection Using Gradient Boost Machine Learning Approach.” 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE). https://doi.org/10.1109/iccike47802.2019.9004419.

Harinath*, B., UG student, Department of ECE, Saveetha School of Engineering, SIMATS, Chennai, Tamilnadu, India., Sitrarasu, T. J. Nagalakshmi, student, Department of ECE, Saveetha School of Engineering, SIMATS, Chennai, Tamilnadu, India., and Asst. Prof., Department of ECE, Saveetha School of Engineering, SIMATS, Chennai, Tamilnadu, India. 2019. “Bone Fracture Detection System Using Image Processing and Matlab.” International Journal of Innovative Technology and Exploring Engineering 8 (12): 1459–61.

Harini, K., and K. K. Sashi Rekha. 2022. “Evaluating Performance Of Identifying At - Risk Students And Learning Achievement Model Using Accuracy And F-Measure by Comparing Logistic Regression, Generalized Linear Model And Gradient Boost Machine Algorithm.” 2022 International Conference for Advancement in Technology (ICONAT). https://doi.org/10.1109/iconat53423.2022.9725848.

Krishna, M. Vamsi, M. Vamsi Krishna, and J. Praveenchandar. 2022. “Comparative Analysis of Credit Card Fraud Detection Using Logistic Regression with Random Forest towards an Increase in Accuracy of Prediction.” 2022 International Conference on Edge Computing and Applications (ICECAA). https://doi.org/10.1109/icecaa55415.2022.9936488.

Levin, Stanislav, and Amiram Yehudai. 2017. “Boosting Automatic Commit Classification Into Maintenance Activities By Utilizing Source Code Changes.” Proceedings of the 13th International Conference on Predictive Models and Data Analytics in Software Engineering. https://doi.org/10.1145/3127005.3127016.

Lughofer, Edwin, Eva Weigl, Wolfgang Heidl, Christian Eitzinger, and Thomas Radauer. 2015. “Drift Detection in Data Stream Classification without Fully Labelled Instances.” 2015 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS). https://doi.org/10.1109/eais.2015.7368802.

Madigan, David, Jonathan Gavrin, and Adrian E. Raftery. 1995. “Eliciting Prior Information to Enhance the Predictive Performance of Bayesian Graphical Models.” Communications in Statistics - Theory and Methods. https://doi.org/10.1080/03610929508831616.

Mahanty, Chandrakanta, Devpriya Panda, and Brojo Kishore Mishra. 2021. “A Review of Different Data Mining Techniques Used in Big Data Applications.” Handbook of Research for Big Data. https://doi.org/10.1201/9781003144526-3.

Nagalakshmi, T. J., Sheshang D. Degadwala, B. Kannan, R. Prabha, and Ravi Kumar. 2022. “Auditory Model System to Recognise Alzheimer’s Diseases: Speech Signal Analysis.” International Journal of Medical Engineering and Informatics 1 (1): 1.

Prentice, Ross L., and Shanshan Zhao. 2019. “Trivariate Failure Time Data Modeling and Analysis.” The Statistical Analysis of Multivariate Failure Time Data. https://doi.org/10.1201/9780429162367-5.

Sammut, Claude, and Michael Harries. 2017. “Concept Drift.” Encyclopedia of Machine Learning and Data Mining. https://doi.org/10.1007/978-1-4899-7687-1_153.

Sayed-Mouchaweh, M., J. Zaytoon, and P. Billaudel. 2011. “Adaptive Time Window Size to Track Concept Drift.” 2011 10th International Conference on Machine Learning and Applications and Workshops. https://doi.org/10.1109/icmla.2011.26.

Shelmerdine, Susan C., Dean Langan, John C. Hutchinson, Melissa Hickson, Kerry Pawley, Joseph Suich, Liina Palm, et al. 2018. “Chest Radiographs versus CT for the Detection of Rib Fractures in Children (DRIFT): A Diagnostic Accuracy Observational Study.” The Lancet. Child & Adolescent Health 2 (11): 802–11.

Shobeyri, Gholamreza. 2023. “Using a Modified MPS Gradient Model to Improve Accuracy of SPH Method for Poisson Equations.” Computational Particle Mechanics. https://doi.org/10.1007/s40571-022-00549-8.

Song, Mingzhou, and Lin Zhang. 2008. “Comparison of Cluster Representations from Partial Second- to Full Fourth-Order Cross Moments for Data Stream Clustering.” 2008 Eighth IEEE International Conference on Data Mining. https://doi.org/10.1109/icdm.2008.143.

Terado, Ryosuke, and Morihiro Hayashida. 2020. “Improving Accuracy and Speed of Network-Based Intrusion Detection Using Gradient Boosting Trees.” Proceedings of the 6th International Conference on Information Systems Security and Privacy. https://doi.org/10.5220/0008963504900497.

Thomas, J. H. 1992. “Notes on Radial Drift for Dalitz Detection.” https://doi.org/10.2172/6789418.

Yadav, J. Suresh, J. Suresh Yadav, and P. Shyamala Bharathi. 2022. “Edge Detection of Images Using Prewitt Algorithm Comparing with Sobel Algorithm to Improve Accuracy.” 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM). https://doi.org/10.1109/iciem54221.2022.9853193.

Figures and Tables

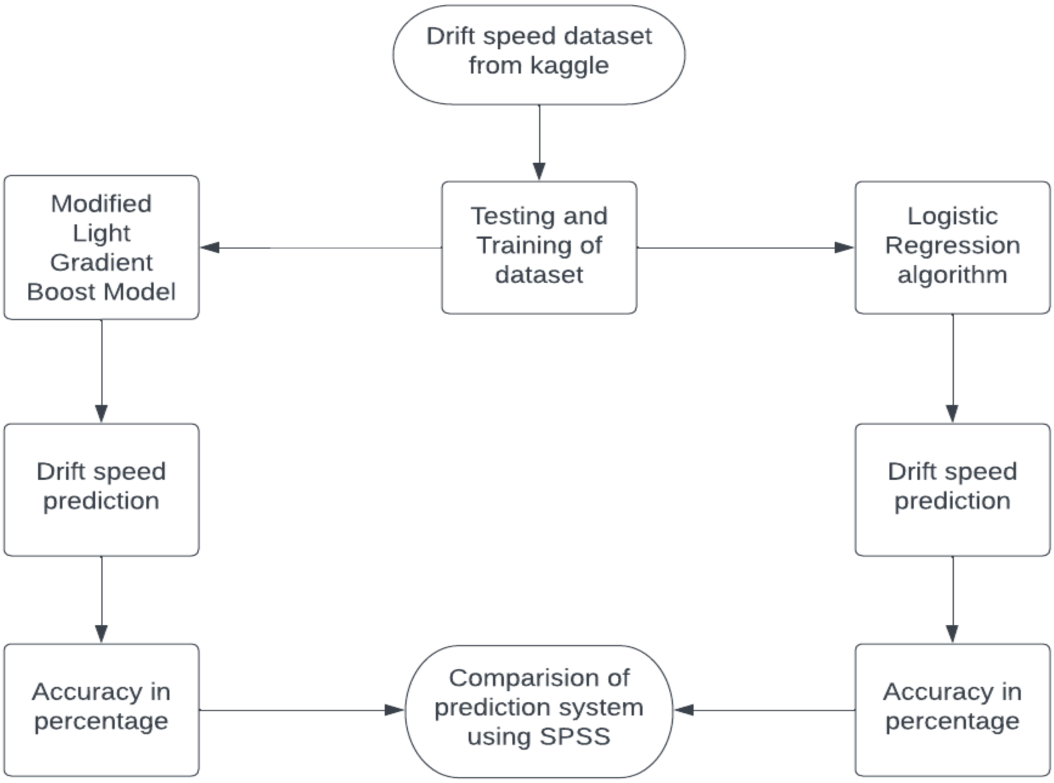

Drift speed dataset from kaggle
Testing and Training of dataset
Modified Light Gradient Boost Model
Logistic Regression algorithm
Drift speed prediction
Drift speed prediction
Accuracy in percentage
Comparision of prediction system using SPSS
Accuracy in percentage

Fig. 1. Process flow the accuracy finding using modified Logistic Regression which is starting from the basic process of importing the data to the program to giving results of the accuracy.

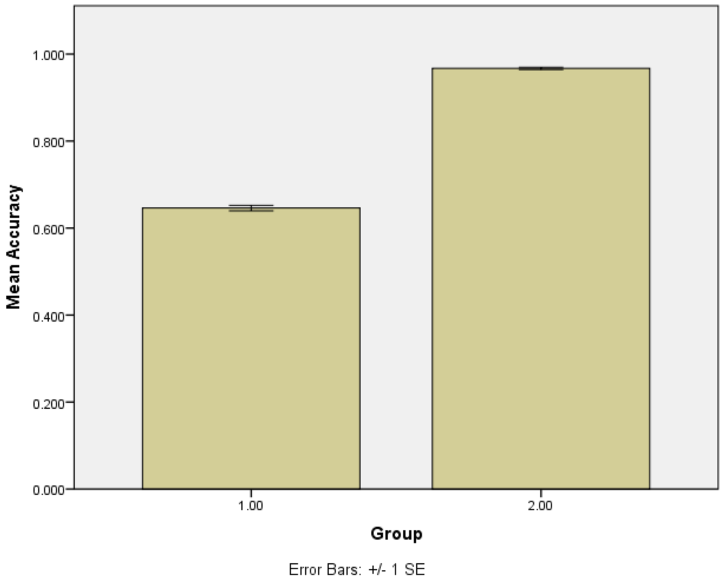


Fig. 2. The Logistic Regression and Modified Light Gradient Boost Model Algorithms were designed and analysed by comparing mean accuracy X axis: Logistic Regression (Group1), Modified Light Gradient Boost Model (Group2) and Y axis: Mean Accuracy, with +/-1 SE.

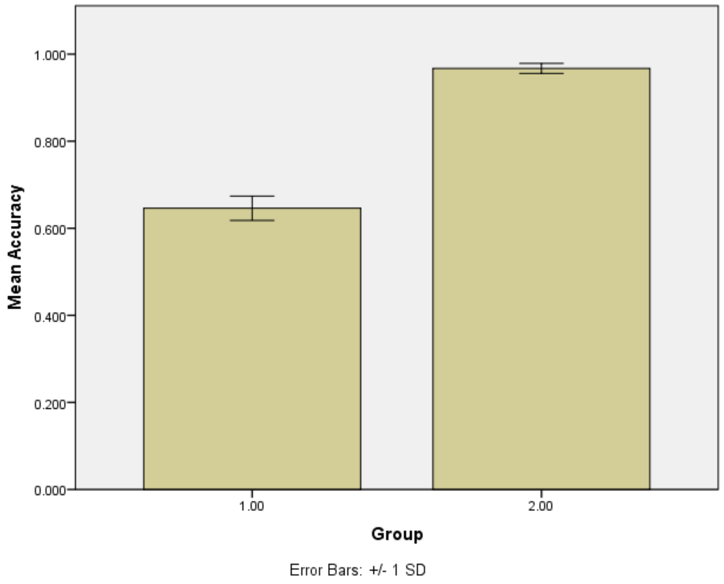


Fig. 3. The Logistic Regression and Modified Light Gradient Boost Model algorithms were designed and analysed by comparing mean accuracy X axis:LogisticRegression (Group1), Modified Light Gradient Boost Model (Group2) and Y axis: Mean Accuracy, with +/- 1SD.

Table 1. Group statistics of Logistic Regression and Modified Light Gradient Boost Model obtained for 20 samples each. The mean of Logistic Regression is 0.646 and forModified Light Gradient Boost Model is 0.96708.

Accuracy	Group	N	Mean	Std.Deviation	Std.Error Mean
	1	20	0.646	0.027985	0.006258
	2	20	0.96708	0.011483	0.002568

Table 2. Group of some independent sample tests for Equality of Variances. The significance taken for this research is 0. The mean difference obtained is determined as 0.321075

Accuracy		Levene's Test for Equality of Variances				T-test for Equality of means				
		F	Sig.	t	df	Sig(2-tailed)	Mean Difference	Std.Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	18.813	0	-47.468	38	0	0.321075	0.006764	0.334768	-0.307382
	Equal variances not assumed			-47.468	25.222	0	0.321075	0.006764	0.334999	-0.307151

Algorithm, Y Axis: Mean accuracy of detection ± 1 SD.

CHAPTER-2

Title page:

Improving Prediction Accuracy in Drift Detection Using Random Forest in Comparing with Modified Light Gradient Boost Model

N.Raja Likitha[1], T. J. Nagalakshmi[2]

N.Raja Likitha[1]
Research Scholar,
Department of Electronics and Communication Engineering,
Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences,
Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.
natharajalikitha19@saveetha.com

T. J. Nagalakshmi[2]
Project Guide, Corresponding Author,
Department of VLSI MicroElectronics,
Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences,
Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.
nagalakshmitj.sse@saveetha.com

Keywords: Innovative Concept Drift, Data Stream mining, Drift Detection, Classifier, Technology, Deep Learning, Modified Light Gradient Boost Model, Random Forest.

ABSTRACT

Aim: The ultimate aim of this proposed research work is to improve the prediction accuracy in drift detection using Random Forest compared with modified light gradient boost model. **Materials and Methods:** A total number of 40 samples were taken by varying tests and training according to data set size. The accuracy was measured to quantify the increasing prediction accuracy in drift detection using Random Forest and modified light gradient boost model. These samples are divided into two groups (Group 1- Random Forest, Group 2- Modified Light Gradient Boost Model) each with 20 samples. The G power is assumed to be 80%. **Result:** The mean accuracy of the Random Forest results for the simulation was 0.71 and the mean accuracy of the Modified Light Gradient Boost Model results was 0.96708. The obtained 2 tailed significance is 0.0 which is less than ($p<0.05$). This shows that there is a statistical significance between the two groups. **Conclusion:** When measuring accuracy for the given dataset Random Forest performs significantly below the Modified Light Gradient Boost Model.

Keywords: Innovative Concept Drift, Data Stream mining, Drift Detection, Classifier, Technology, Deep Learning, Modified Light Gradient Boost Model, Random Forest

INTRODUCTION

Drift is the statistical term representing sudden change in the parametric properties of a goal over time. The problem with this is that the prediction gets less accurate over time. A major issue is the absence of an appropriate method for handling the constant flow of data (Quinonero-Candela et al. 2022). Data stream mining in this area is difficult since it must swiftly spot changes in order to extract any valuable information. On these altering facts, it must continue to preserve correct statistics. In a similar vein, the predictive model needs to be updated when input changes. When several models of fundamental approaches are employed to enhance performance overall,the word "ensemble" is employed. The Similar idea is applied in various data extraction, from which algorithms have been constructed to use Classifier and Enhancing and Loading approaches (Koonin and Galperin 2013). Changes in data ideas need to be controlled carefully because they have an effect on prediction model accuracy. These changes are detected by drift detection algorithms, letting its proposed method to review at the drift income level deep learning (Casillas, Wang, and Yao 2018). Deep learning applications include drift mobile, drift life, and drift connect (al., Nagalakshmi, and al. 2021) (Sammut and Harries 2017).

In recent years, more than a thousand publications have been published in Google Scholar, Springer, and IEEE Xplore. 1–25 of 412 papers in IEEE Xplore and 18,600 in Google Scholar have received the most citations (Lughofer et al. 2015), in this citation they have done using classification-based stream mining in order to alert operators in the event of unintended system changes (Thomas 1992), here this site uses a radical drift time expansion chamber (Becker and Ebner 2019), here in this site they used collision detection on acceleration data (Shelmerdine et al. 2018) here by doing surveys in hospitals they give digital radiography systems with wireless detectors (Chua, Jordan, and Muller 2020), in this site by using online they use to update initially trained with offline labs.

The problem with the previous work is that the predictions get less accurate over time. It is a term used in machine learning and predictive analytics. Innovative Concept drift is the unanticipated change over time in the statistical properties of the target variable that the model is attempting to forecast.

MATERIALS AND METHODS

The testing was completed in a computer lab at the Saveetha School of Engineering's Department of VLSI at the Saveetha Institute of Medical and Technical Sciences. There are two accuracy divisions in the project. Each group has a sample size of 20 data points, and the G power is 80%. Group 1 is the Random Forest model, while Group 2 is the Modified Light Gradient Boost Model model. Google Colab has been used to compare the accuracy of the results and the necessary algorithm.

The data set is implemented with Google Collab and code that was inspired by Collab. The data stream mining has been imported, and the data visualization has been completed. Following the visualization, the data set goes through a data preparation stage in which the error numbers from the Google Collab drive are compared to the mounted code. The current classifier, Random Forest is compared to the Modified Light Gradient Boost Model as for regards to precision.

Random Forest: For classification purposes, monitored methods and algorithms such as random forest are frequently used with deep learning. It builds tree structure on samples collected and utilizes those average for identification as well as significant number poll for return. One of the Random Forest Algorithm's most important characteristics is its capability to deal with collected data with these time series of Innovative concept drift, just like correlation and relative frequency, as it is in evaluation. It produces superior results when it comes to classification tasks.

Modified Light Gradient Boost Model: The Train Using AutoML tool employs the Modified Light Gradient Boost Model, a gradient boosting ensemble technique that is based on Random Forest. Light Gradient Boost Model is a decision tree-based technique that may be applied to both classification and regression problems. For excellent performance with dispersed systems, Modified Light Gradient Boost Model has been specially designed.

Test Procedure

Figure. 1. demonstrates the workflow process, in which the code stimulated by Google Collab is used to implement the data set. The Innovative concept drift of data is imported, and the data visualization is completed. Following visualization, the data set goes through a level of data stream mining preparation where the error numbers from the Google Collab drive are checked against the mounted code. The Random Forest's accuracy was then compared against the Modified Light Gradient Boost Model, the current classifier, for accuracy.

STATISTICAL ANALYSIS

The suggested study work as well as the preceding work method's accuracy are both evaluated using the SPSS software package. Testing is carried out using a separate sample T-test. Group 1 has been taken as Random Forest. Group 2 is taken as a Modified Light Gradient Boost Model. The independent variables are baseline information, properties of images. The dependent variable is accuracy (Soni and Yadav 2022)**.**

RESULTS

Compare the accuracy percentage values of the Random forest with Modified Light Gradient Boost Model both algorithms provide different Mean accuracy Modified Light Gradient Boost Model 0.96708 and Random Forest 0.71.

In Fig. 2. compares accuracy results using a flow chart (Random Forest and Modified Light Gradient Boost Model), in that Random Forest Provides greater precision regarding 0.71 while the Modified Light Gradient Boost Model provides approximately 0.96708. Here the error bar is +/-1 Standard Error. Fig. 3. is the graphical representation with +/- 1 standard deviation.

Table 1. demonstrates the T-test tables. The Random Forest (0.71) classifier has a higher mean accuracy value when compared to the Modified Light Gradient Boost Model with accuracy 0.96708 classifier, which was evaluated with the N=20 for a group, in the table comparing the two classifiers. The means that were different for both of the classifiers present also revealed the standard deviation.

Table 2. independent sample test.This table shows the sample group the T-test is used for both groups and has revealed that precision as (t = -37.451) & Mean Difference = (-0.257075) and The only variation and also has the relatively similar sampling error 0.006764. Between the two groups, there is a statistical difference (means difference is -0.257075) (p<0.05).

DISCUSSION

The data from the research project that was carried out showed that the Random Forest algorithm outperforms the Random Forest classifier with 0.96708 better when it comes to predicting the data stream mining with the accurate values that have the higher accuracy (p<0.05, Independent variable test, SPSS IBM tool), and with the significance that is considered as better work. The prior technology findings demonstrate that when determining the Modified Light Gradient Boost Model's accuracy, the Random Forest support is not superior to it.

It aims to identify frauds committed with credit or debit cards as well as a test carried out to determine the most appropriate method. Here this site gives a Random Forest accuracy value of 76% (Krishna, Vamsi Krishna, and Praveenchandar 2022). (Susmitha, Femila, and Sivasamy 2022) this site uses machine learning technology to detect the Random Forest for getting accuracy. The accuracy percentage of Random Forest is 73% (M S et al. 2022), in this site by keeping an eye on the performance of the random forest model, they were able to detect outlier detection in finance datasets. So the accuracy of Random Forest is 71% (McCarthy and Sen Gupta 2021), in this article the Random Forest algorithm accuracy percent is nearly 75% (S. Kumar et al. 2021), in this site the average accuracy of Random Forest is 70% and which supports the Random Forest to improve in research.

In (M. R. Kumar and Malathi 2022)(K. S. Kumar and Nagalakshmi 2022) suggested the Random Forest algorithm for detection in this article using machine learning and its accuracy percent is 99%. Whereas in (Hemanthkumar and Shyamala Bharathi 2022) gives the Random Forest accuracy of 88%. To develop and change the accuracy with deep learning (Bénitière, Necsulea, and Duret, n.d.; Bharathi, Vijaya Bharathi, and Sireesha 2016), in this site they tried to use Random Forest for detection to make it a challenging role to get the accuracy. It increases accuracy rate nearly 80%.

The problems experienced in this research work are Random forest classifiers may not be suitable for real-time predictions due to their slow rate, Because it uses multiple decision trees to make predictions, this algorithm is significantly slower than other classification algorithms and The Random Forest algorithm can be significantly altered by a small data change.

Further development of Random Forest is trying to increase the speed in getting predictions in real-time. And make the significance faster than the other classifier algorithms. Changes in the data as a result of seasonality, shifts in consumer preferences, and other such factors are some examples that can expand upon in the future of Innovative concept drift. This can make technology easier to use and improve with this technology.

CONCLUSION

Results for the Drift Detection for Random Forest is 0.71 mean accuracy and 0.96708 mean accuracy produced by the Modified Light Gradient Boost Model. The Modified Light Gradient Boost Model algorithm gives higher accuracy as compared to the Random Forest algorithm.

DECLARATIONS

Conflict of Interest

No conflict of interest in this manuscript.

Author Contributions

RL played a role in the planning, simulation, acquisition, evaluation, and drafting of the manuscript. TJNL contributed to the conceptualization, validation of data, and critical evaluation of the manuscript.

Acknowledgment

The authors express their gratitude to the Saveetha School of Engineering , Saveetha Institute of Medical and Technical Sciences (formerly known as Saveetha University) for supplying the necessary resources that allowed for the successful completion of this research.

Funding

We would like to express our gratitude to the following organizations for giving the funding necessary for us to finish this research project.

1. Haneesh Sai Enterprises, Nellore.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

REFERENCES

al., Dr T. J. Nagalakshmi Et, T. J. Nagalakshmi, and Et al. 2021. “Intelligent Door Knocking Security System Using IOT.” Turkish Journal of Computer and Mathematics Education (TURCOMAT). https://doi.org/10.17762/turcomat.v12i2.2205.

Becker, Felix, and Marc Ebner. 2019. “Collision Detection for a Mobile Robot Using Logistic Regression.” Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics. https://doi.org/10.5220/0007768601670173.

Bénitière, Florian, Anamaria Necsulea, and Laurent Duret. n.d. “Random Genetic Drift Sets an Upper Limit on mRNA Splicing Accuracy in Metazoans.” https://doi.org/10.1101/2022.12.09.519597.

Bharathi, M. Vijaya, M. Vijaya Bharathi, and Rodda Sireesha. 2016. “Enhancing Fault Divination Accuracy Using Naïve Bayes Classifier with PYTHON and PHP.” International Journal of Software Engineering and Its Applications. https://doi.org/10.14257/ijseia.2016.10.8.07.

Casillas, Jorge, Shuo Wang, and Xin Yao. 2018. “Concept Drift Detection in Histogram-Based Straightforward Data Stream Prediction.” 2018 IEEE International Conference on Data Mining Workshops (ICDMW). https://doi.org/10.1109/icdmw.2018.00129.

Chua, Adelson, Michael I. Jordan, and Rikky Muller. 2020. “Unsupervised Online Learning for Long-Term High Sensitivity Seizure Detection.” Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference 2020 (July): 528–31.

Hemanthkumar, K. A., and P. Shyamala Bharathi. 2022. “Improved Accuracy of Plant Leaf Classification Using Random Forest Classifier over K-Nearest Neighbours.” 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES). https://doi.org/10.1109/icses55317.2022.9914269.

Koonin, Eugene V., and Michael Galperin. 2013. Sequence — Evolution — Function: Computational Approaches in Comparative Genomics. Springer Science & Business Media.

Krishna, M. Vamsi, M. Vamsi Krishna, and J. Praveenchandar. 2022. “Comparative Analysis of Credit Card Fraud Detection Using Logistic Regression with Random Forest towards an Increase in Accuracy of Prediction.” 2022 International Conference on Edge Computing and Applications (ICECAA). https://doi.org/10.1109/icecaa55415.2022.9936488.

Kumar, K. Saketh, and T. J. Nagalakshmi. 2022. “Design of Intrusion Detection System for Wireless Ad Hoc Network in the Detection of Man In The Middle Attack Using Principal Component Analysis Classifier Method Comparing with ANN Classifier.” In 2022 14th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS), 1–6.

Kumar, Marri Ranjith, and K. Malathi. 2022. “An Innovative Method in Improving the Accuracy in Intrusion Detection by Comparing Random Forest over Support Vector Machine.” 2022 International Conference on Business Analytics for Technology and Security (ICBATS). https://doi.org/10.1109/icbats54253.2022.9759062.

Kumar, Sanjeev, Ravendra Singh, Mohammad Zubair Khan, and Abdulfattah Noorwali. 2021. “Design of Adaptive Ensemble Classifier for Online Sentiment Analysis and Opinion Mining.” PeerJ. Computer Science 7 (August): e660.

Lughofer, Edwin, Eva Weigl, Wolfgang Heidl, Christian Eitzinger, and Thomas Radauer. 2015. “Drift Detection in Data Stream Classification without Fully Labelled Instances.” 2015 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS). https://doi.org/10.1109/eais.2015.7368802.

McCarthy, Ryan A., and Ananya Sen Gupta. 2021. “Employing and Interpreting a Machine Learning Target-Cognizant Technique for Analysis of Unknown Signals in Multiple Reaction Monitoring.” IEEE Access : Practical Innovations, Open Solutions 9 (February): 24727–37.

M S, Abdul Razak, C. R. Nirmala, Maha Aljohani, and B. R. Sreenivasa. 2022. “A Novel Technique for Detecting Sudden Concept Drift in Healthcare Data Using Multi-Linear Artificial Intelligence Techniques.” Frontiers in Artificial Intelligence 5 (August): 950659.

Quinonero-Candela, Joaquin, Masashi Sugiyama, Anton Schwaighofer, and Neil D. Lawrence. 2022. Dataset Shift in Machine Learning. MIT Press.

Sammut, Claude, and Michael Harries. 2017. “Concept Drift.” Encyclopedia of Machine Learning and Data Mining. https://doi.org/10.1007/978-1-4899-7687-1_153.

Shelmerdine, Susan C., Dean Langan, John C. Hutchinson, Melissa Hickson, Kerry Pawley, Joseph Suich, Liina Palm, et al. 2018. “Chest Radiographs versus CT for the Detection of Rib Fractures in Children (DRIFT): A Diagnostic Accuracy Observational Study.” The Lancet. Child & Adolescent Health 2 (11): 802–11.

Soni, Sheetal, and Usha Yadav. 2022. “COVID-19 Impact on Online Learning: A Statistical and Machine Learning Model Analysis for Stress Detection.” Predictive Analytics of Psychological Disorders in Healthcare. https://doi.org/10.1007/978-981-19-1724-0_7.

Susmitha, Inturi, Roseline J. Femila, and Vinay Sivasamy. 2022. “Detection of Forest Fire Using Support Vector Machine in Comparison with Random Forest to Measure Accuracy, Precision and Recall.” 2022 International Conference on Cyber Resilience (ICCR). https://doi.org/10.1109/iccr56254.2022.9995895.

Thomas, J. H. 1992. “Notes on Radial Drift for Dalitz Detection.” https://doi.org/10.2172/6789418.

Figure and Tables

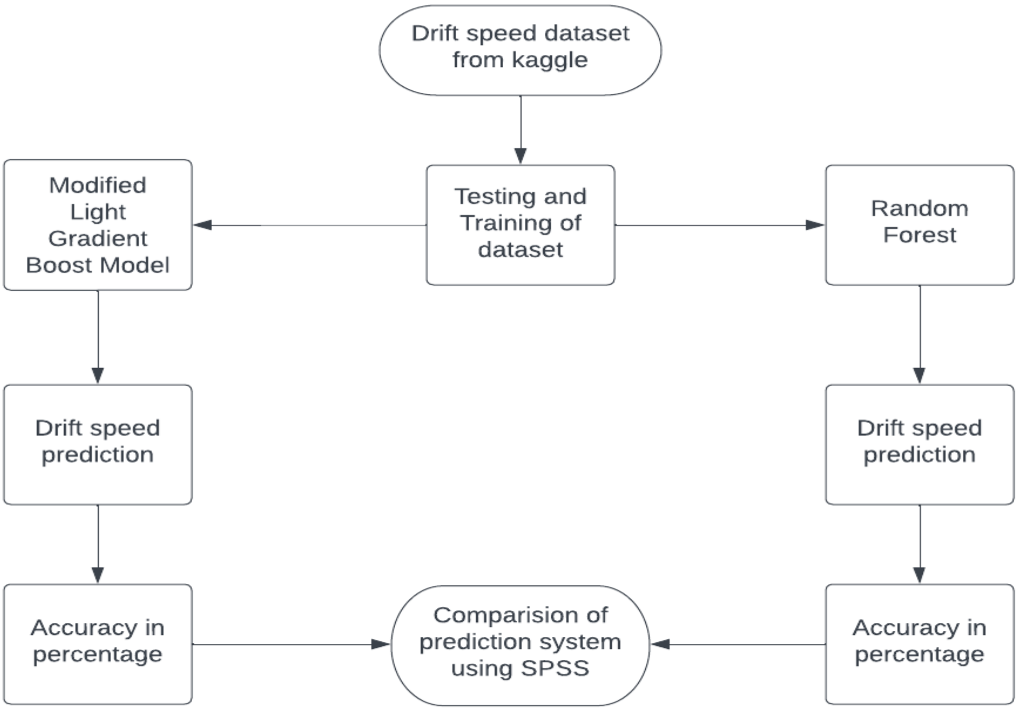

Drift speed dataset from kaggle
Testing and Training of dataset
Modified Light Gradient Boost Model
Random Forest
Drift speed prediction
Drift speed prediction
Accuracy in percentage
Accuracy in percentage
Comparision of prediction system using SPSS

Fig. 1. Process flow the accuracy finding using modified Random Forest which is starting from the basic process of importing the data to the program to giving results of the accuracy.

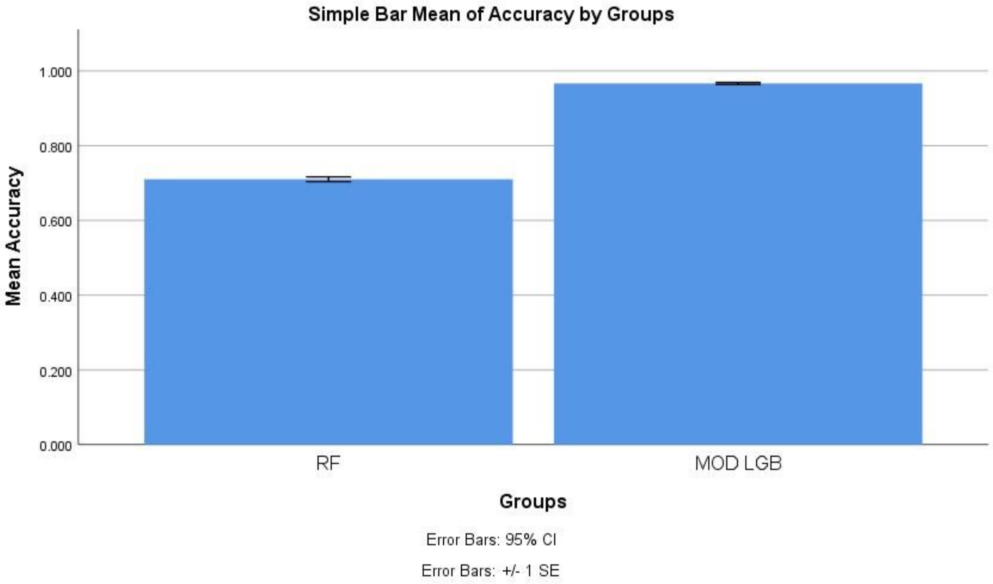


Fig. 2. The Random Forest and Modified Light Gradient Boost Model algorithms were designed and analyzed by comparing mean accuracy X axis: Random Forest (Group1), Modified Light Gradient Boost Model (Group2) and Y axis: Mean Accuracy, with +/-1 SE

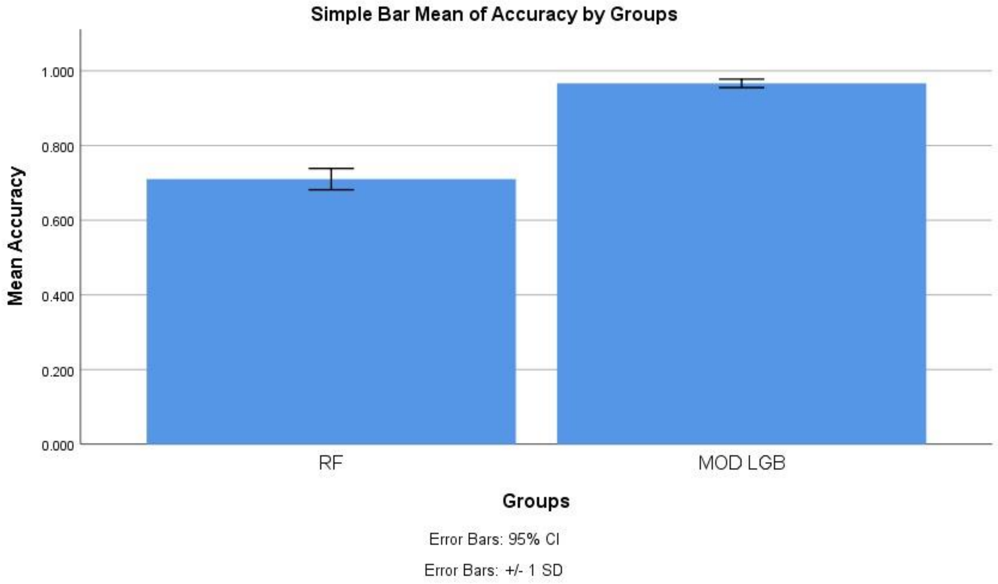


Fig. 3. The Random Forest and Modified Light Gradient Boost Model algorithms were designed and analyzed by comparing mean accuracy X axis: Random Forest (Group1), Modified Light Gradient Boost Model (Group2) and Y axis: Mean Accuracy, with +/- 1 SD.

Table 1. Group statistics of Random Forest and Modified Light Gradient Boost Model obtained for 20 samples each. The mean of Random Forest is 0.71 and forModified Light Gradient Boost Model is 0.96708

	Group	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	1	20	0.71	0.02847	0.006366
	2	20	0.96708	0.011483	0.002568

Table 2. Group of some independent sample tests for Equality of Variances. The significance value obtained for this study is 0 indicating statistically significant difference between the groups. The mean difference obtained is determined as -0.257075.

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	**Sig.**	**t**	**df**	**Sig. (2-tailed)**	**Mean Diff**	**Std. Error Diff**	**Lower**	**Upper**
Accuracy	**Equal variances assumed**	16.431	0	-37.451	38	0	-0.257075	0.006864	-0.270971	-0.243179
	Equal variances not assumed	-	-	-37.451	25.023	0	-0.257075	0.006864	-0.271212	-0.242938

CHAPTER-3

Title page:

Improving Prediction Accuracy in Drift Detection Using K-Nearest Neighbor in Comparing with Modified Light Gradient Boost Model

N.Raja Likitha[1], T.J.Nagalakshmi[2]

N.Raja Likitha[1]
Research Scholar,
Department of Electronics and Communication Engineering,
Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences,
Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.
natharajalikitha19@saveetha.com

T. J. Nagalakshmi[2]
Project Guide, Corresponding Author,
Department of VLSI MicroElectronics,
Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences,
Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.
nagalakshmitj.sse@saveetha.com

Keywords: Innovative Concept Drift, Data Stream Mining, Drift Detection, Classifier, Technology, Deep Learning, Modified Light Gradient Boost Model, K-Nearest Neighbor.

ABSTRACT

Aim: The aim of the proposed work is to improve prediction accuracy in drift detection technology using K-Nearest Neighbor and to compare with modified light gradient boost models. **Materials and Methods:** A total of 40 samples of drift were gathered using a range of predictions. These samples are divided into two groups (Group 1- K- Nearest Neighbor, Group 2- Modified Light Gradient Boost Model) each having 20 samples and the accuracy was calculated using KNN and a modified light gradient boost model to evaluate the increasing prediction accuracy in drift detection. The G power is considered to be 80%. **Result:** The K-Nearest Neighbor produced results for the simulation with a 0.6615 Mean accuracy, whereas the Modified Light Gradient Boost Model provided results with a 0.96708 accuracy. The obtained 2 tailed significance is 0.0 which is less than specified p=0.05 (p<0.05). Hence it is observed that, in terms of accuracy the drift detection systems based on KNN and modified light gradient boost are statistically significant. **Conclusion:** For the given dataset K-Nearest Neighbor Performs significantly less than the Modified Light Gradient Boost Model in finding the accuracy.

Keywords: Innovative Concept Drift, Data Stream Mining, Drift Detection, Classifier, Technology, Deep Learning, Modified Light Gradient Boost Model, K-Nearest Neighbor.

INTRODUCTION

The random word for a sudden change in the target variable's statistical characteristics over time is drift. The issue with this is that as time passes, the forecasts get less precise (Sun, Tang, and Qiao 2019) Lack of a suitable mechanism for handling the continuous flow of data is a major problem with data stream mining analysis. Lack of a suitable approach for handling the constant flow of data is a major issue with data stream mining analysis (Cavanillas, Curry, and Wahlster 2016). It is a challenging field for data mining since the method needs to identify changes quickly in order to extract useful data from it. It must maintain accurate statistics on this changing data. As input changes, the predictive model should be updated as well(Nagalakshmi et al. 2022) (Banerjee and Dutta 2013).

Over a thousand publications have appeared in Google Scholar, Springer, and IEEE Xplore in recent years. The majority of citations were found in 1–25 of 412 papers in IEEE Xplore and 18,600 in Google Scholar. In this citation they have done using classification-based stream mining in order to alert operators in the event of unintended system changes deep learning (Lughofer et al. 2015). Here this site uses a radical drift time expansion chamber (Thomas 1992). Here in this site they used collision detection on acceleration data (Becker and Ebner 2019). Here by doing surveys in hospitals they give digital radiography systems with wireless detectors (Shelmerdine et al. 2018). In this site by using online they use to update initially trained with offline labs (Chua, Jordan, and Muller 2020)

The term "ensemble" is used when many models of base approaches are applied to a single input to enhance overall performance. The similar idea is applied to data stream mining, where ensembles are created by combining Bagging, Boosting, and Stacking techniques (Alazzam, Alsmadi, and Akour 2017). Due to its impact on prediction model accuracy, the change in data stream mining idea must be managed carefully in deep learning. Such changes are collected by drift detection technology, which enables the prediction model to update at the innovative concept drift rate (Greco and Cerquitelli 2021).

The issue with previous work is that predictions become less accurate over time innovative concept drift. It's a term commonly used during machine learning and predictive analytics. The unexpected transformation over a period in the statistical properties of such variables that the concept is attempting to forecast is referred to as innovative concept drift.

MATERIALS AND METHODS

The experiment was conducted in a lab at the Saveetha Institute of Medical and Technical Sciences's Department of VLSI at the Saveetha School of Engineering. Two accuracy groups work on the project. The G power is 80%, with a sample size of 20 data points for each group. The K-Nearest Neighbor model is in Group 1, and the Modified Light Gradient Boost Model is in Group 2. Google Colab has been utilized to compare the accuracy of the results and the necessary algorithm for innovative concept drift.

The data set is implemented using Google Collab and code inspired by Collab. The data stream mining has been imported, and the data visualization is complete. Following the visualization, the data set goes through a data preparation stage in which the error numbers from the Google Collab drive are compared to the mounted code. In terms of accuracy, the current classifier, K-Nearest Neighbor, is fought against the Modified Light Gradient Boost Model.

K-Nearest Neighbor: The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised deep learning classifier that uses proximity to classify or predict the grouping of a single data point of technology. While it can be used for either regression or classification problems, it is most commonly used as a classification algorithm, based on the assumption that similar points can be found nearby. Modified Light Gradient Boost Model is a gradient boosting framework based on decision trees that improves model efficiency while reducing memory usage.

Modified Light Gradient Boost Model: The Train Using AutoML tool employs the Modified Light Gradient Boost Model, a gradient boosting ensemble technique that is based on K-Nearest Neighbor. Light Gradient Boost Model is a decision tree-based technique that may be applied to both classification and regression problems in deep learning. For excellent performance with dispersed systems, Modified Light Gradient Boost Model has been specially designed.

Test Procedure

Figure. 1. shows the workflow process, in which Google Collab-generated code is used to implement the innovative concept drift data set. The data has been imported, and the data visualization is finished. Following visualization, the data set is subjected to data preparation, in which the error numbers from the Google Collab drive are compared to the mounted code. The accuracy of the K-Nearest Neighbor was then compared to that of the current classifier, the Modified Light Gradient Boost Model.

STATISTICAL ANALYSIS

The SPSS software package is used to assess the accuracy of both the proposed study work and the preceding work method. A separate sample T-test is used for testing. K-Nearest Neighbor has been assigned to Group 1. The Modified Light Gradient Boost Model is used in Group 2. The independent variables are baseline information, properties of images. The dependent variable is accuracy (Fuadah, Setiawan, and Mengko 2015)**.**

RESULTS

Compare the accuracy percentage values of the K-Nearest Neighbor with Modified Light Gradient Boost Model both algorithms provide different Mean accuracy Modified Light Gradient Boost Model 0.96708 and K-Nearest Neighbor 0.6615.

In Fig. 2. shows the bar chart comparison of accuracy values (Modified Light Gradient Boost Model and K-Nearest Neighbor), in that K-Nearest Neighbor gives mean accuracy about 0.6615 where the

Modified Light Gradient Boost Model gives mean accuracy about 0.96708. Here the error bar is +/-1 Standard Error. Fig. 3. is the graphical representation with +/- 1 standard deviation.

Table 1. demonstrates the T-test tables. The K-Nearest Neighbor (0.6615) classifier has a higher mean accuracy value when compared to the Modified Light Gradient Boost Model mean accuracy 0.96708 classifier, which was evaluated with the N=20 for a group, in the table comparing the two classifiers. The means that were different for both of the classifiers present also revealed the standard deviation.

Table 2. independent sample test. This table shows the independent sample T-test is performed for the two groups and has found that accuracy as (t = -26.544) & Mean Difference = (-0.305575) and it has the same standard error difference is 0.006764. Between the two groups , there is a significant difference (means difference is -0.305575)(p<0.05).

DISCUSSION

The research project results revealed that the Modified Light Gradient Boost Model classifier outperformed the K-Nearest Neighbor algorithm with a 0.96708 accuracy rate in predicting the data with the most accurate values (p<0.05, Independent variable test, SPSS IBM tool), which is considered to be better work. Prior findings show that when determining the accuracy of the Modified Light Gradient Boost Model, the K-Nearest Neighbor support is not superior to it.

This site gives the accuracy percentage of K-Nearest Neighbor is 66%. Here the prediction accuracy/error rate was used to determine the relative importance of the selected variables in relation to the number of predictors in the models (Barrett et al. 2001). Here they got an accuracy percentage for K-Nearest Neighbor is 70%. As a result, characteristics that are not as useful to the opponent have less influence. Feature weighting has been shown to improve K-Nearest Neighbor accuracy in deep learning (Syaliman, Labellapansa, and Yulianti 2019). Here the site gives 65% accuracy of the K-Nearest Neighbor algorithm. It detects gauss-Markov random fields with K-Nearest Neighbor (Anandkumar, Tong, and Swami 2010) . In this site after many failures the accuracy of K-Nearest Neighbor(Swaroop, Nagalakshmi, and Subash Sharma 2022) (Arellano 2020) .

In this site the accuracy percentage of K-Nearest Neighbor is 94.5% (Gao et al. 2022). They mentioned the system's high accuracy for cataract detection and used feature extraction evaluation as well as K-Nearest Neighbor (Bharathi, Shyamala Bharathi, and Khiran Sai 2022). In this site the accuracy they got was 90% (Fuadah, Setiawan, and Mengko 2015). They included some famous prototype implementations. They achieved 93% accuracy on this website. And rechecked and chosen from among all submissions in deep learning (Gama and Gaber 2007).

The disadvantage of K-Nearest Neighbor is that it can be too slow and inefficient for real-time forecasting when there are many trees. These algorithms are typically quick to train but slow to predict after training.

CONCLUSION

The Modified Light Gradient Boost Model produced drift detection technology results with 0.6615 and 0.96708 accuracy for K-Nearest Neighbor. The K-Nearest Neighbor algorithm is less accurate than the Modified Light Gradient Boost Model algorithm.

DECLARATIONS

Conflict of Interest

No conflict of interest in this manuscript.

Author Contributions

RL played a role in the planning, simulation, acquisition, evaluation, and drafting of the manuscript. TJNL contributed to the conceptualization, validation of data, and critical evaluation of the manuscript.

Acknowledgment

The authors express their gratitude to the Saveetha School of Engineering and the Saveetha Institute of Medical and Technical Sciences (formerly known as Saveetha University) for supplying the necessary resources that allowed for the successful completion of this research.

Funding

We would like to express our gratitude to the following organizations for giving the funding necessary for us to finish this research project.

1. Haneesh Sai Enterprises, Nellore.

2. Saveetha University.

3. Saveetha Institute of Medical and Technical Sciences.

4. Saveetha School of Engineering.

REFERENCES

Alazzam, Iyad, Izzat Alsmadi, and Mohammed Akour. 2017. “Software Fault Proneness Prediction: A Comparative Study between Bagging, Boosting, and Stacking Ensemble and Base Learner Methods.” International Journal of Data Analysis Techniques and Strategies. https://doi.org/10.1504/ijdats.2017.10003991.

Anandkumar, Animashree, Lang Tong, and Ananthram Swami. 2010. “Detection of Gauss-Markov Random Fields with Nearest-Neighbor Dependency.” https://doi.org/10.21236/ada536158.

Arellano, Giovanna. 2020. “Review for ‘Early Failure Detection of Paper Manufacturing Machinery Using Nearest Neighbor-based Feature Extraction.’” https://doi.org/10.1002/eng2.12291/v1/review2.

Banerjee, Bonny, and Jayanta K. Dutta. 2013. “A Predictive Coding Framework for Learning to Predict Changes in Streaming Data.” 2013 IEEE 13th International Conference on Data Mining Workshops. https://doi.org/10.1109/icdmw.2013.134.

Barrett, J., M. G. Cox, M. P. Dainton, and P. M. Harris. 2001. “A METHODOLOGY FOR TESTING THE NUMERICAL ACCURACY OF SCIENTIFIC SOFTWARE USED IN METROLOGY.” Advanced Mathematical and Computational Tools in Metrology V. https://doi.org/10.1142/9789812811684_0004.

Becker, Felix, and Marc Ebner. 2019. “Collision Detection for a Mobile Robot Using Logistic Regression.” Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics. https://doi.org/10.5220/0007768601670173.

Bharathi, P. Shyamala, P. Shyamala Bharathi, and G. Khiran Sai. 2022. “Artificial Neural Networks Are Used in a Dark Having to Learn Biometric Smart Card in Contrast to the Integrated Method.” 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES). https://doi.org/10.1109/icses55317.2022.9914367.

Cavanillas, José María, Edward Curry, and Wolfgang Wahlster. 2016. New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe. Springer.

Chua, Adelson, Michael I. Jordan, and Rikky Muller. 2020. “Unsupervised Online Learning for Long-Term High Sensitivity Seizure Detection.” Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference 2020 (July): 528–31.

Fuadah, Y. N., A. W. Setiawan, and T. L. R. Mengko. 2015. “Performing High Accuracy of the System for Cataract Detection Using Statistical Texture Analysis and K-Nearest Neighbor.” 2015 International Seminar on Intelligent Technology and Its Applications (ISITIA). https://doi.org/10.1109/isitia.2015.7219958.

Gama, João, and Mohamed Medhat Gaber. 2007. Learning from Data Streams: Processing Techniques in Sensor Networks. Springer Science & Business Media.

Gao, Long, Donghui Li, Lele Yao, and Yanan Gao. 2022. “Sensor Drift Fault Diagnosis for Chiller System Using Deep Recurrent Canonical Correlation Analysis and K-Nearest Neighbor Classifier.” ISA Transactions 122 (March): 232–46.

Greco, Salvatore, and Tania Cerquitelli. 2021. “Drift Lens: Real-Time Unsupervised Concept Drift Detection by Evaluating per-Label Embedding Distributions.” 2021 International Conference on Data Mining Workshops (ICDMW). https://doi.org/10.1109/icdmw53433.2021.00049.

Lughofer, Edwin, Eva Weigl, Wolfgang Heidl, Christian Eitzinger, and Thomas Radauer. 2015. “Drift Detection in Data Stream Classification without Fully Labelled Instances.” 2015 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS). https://doi.org/10.1109/eais.2015.7368802.

Nagalakshmi, T. J., N. Nalini, P. Jagadeesh, P. Shyamala Bharathi, V. Amudha, and G. Ramkumar. 2022. “Detection of Cervical Cancer with Texture Analysis Using Machine Learning Models.” In 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), 1–6.

Shelmerdine, Susan C., Dean Langan, John C. Hutchinson, Melissa Hickson, Kerry Pawley, Joseph Suich, Liina Palm, et al. 2018. “Chest Radiographs versus CT for the Detection of Rib Fractures in Children

(DRIFT): A Diagnostic Accuracy Observational Study." The Lancet. Child & Adolescent Health 2 (11): 802–11.

Sun, Zijian, Jian Tang, and Junfei Qiao. 2019. "Double Window Concept Drift Detection Method Based on Sample Distribution Statistical Test." 2019 Chinese Automation Congress (CAC). https://doi.org/10.1109/cac48633.2019.8996456.

Swaroop, Jyothi, T. J. Nagalakshmi, and S. Subash Sharma. 2022. "Girl Child Security System Based on IOT Technology with GPS Tracker Comparing with Fuzzy Classifier Based Safety Device." In 2022 International Conference on Cyber Resilience (ICCR), 1–6.

Syaliman, K. U., Ause Labellapansa, and Ana Yulianti. 2019. "Improving the Accuracy of Features Weighted K-Nearest Neighbor Using Distance Weight." Proceedings of the Second International Conference on Science, Engineering and Technology. https://doi.org/10.5220/0009390903260330.

Thomas, J. H. 1992. "Notes on Radial Drift for Dalitz Detection." https://doi.org/10.2172/6789418.

Figure and Tables

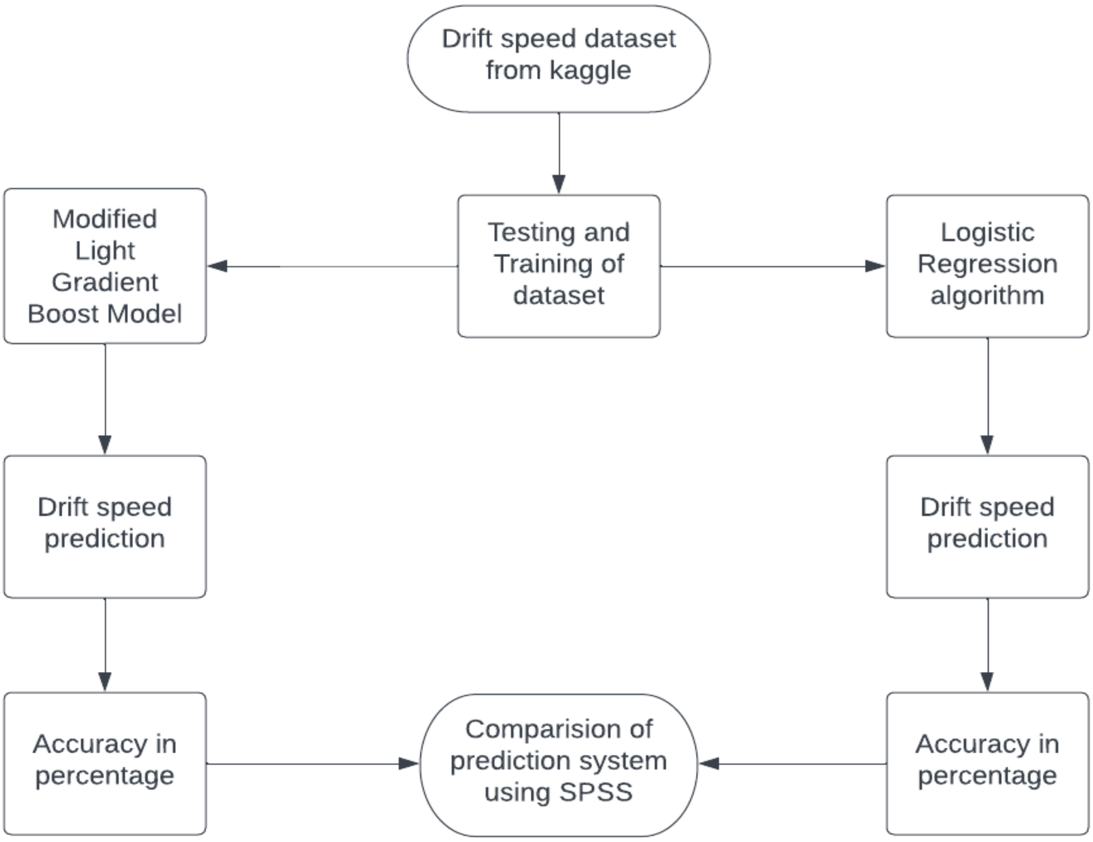

Drift speed dataset from kaggle
Testing and Training of dataset
Modified Light Gradient Boost Model
Logistic Regression algorithm
Drift speed prediction
Drift speed prediction
Accuracy in percentage
Comparision of prediction system using SPSS
Accuracy in percentage

Fig. 1. Process flow the accuracy finding using modified K-Nearest Neighbor which is starting from the basic process of importing the data to the program to giving results of the accuracy.

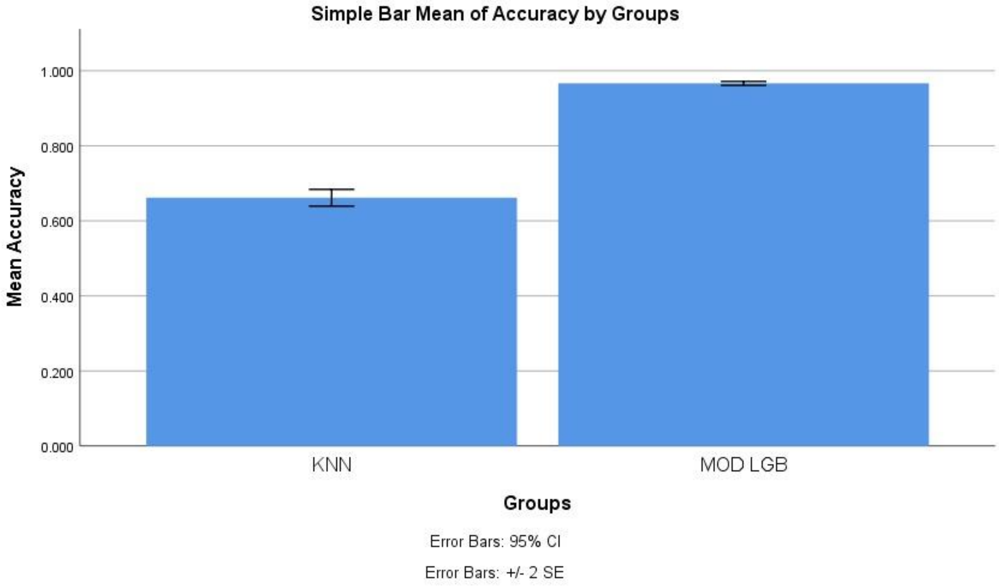


Fig. 2. The K-Nearest Neighbor and Modified Light Gradient Boost Model algorithms were designed and analyzed by comparing mean accuracy X axis: K-Nearest Neighbor (Group1), Modified Light Gradient Boost Model (Group2) and Y axis: Mean Accuracy, with +/-1SE

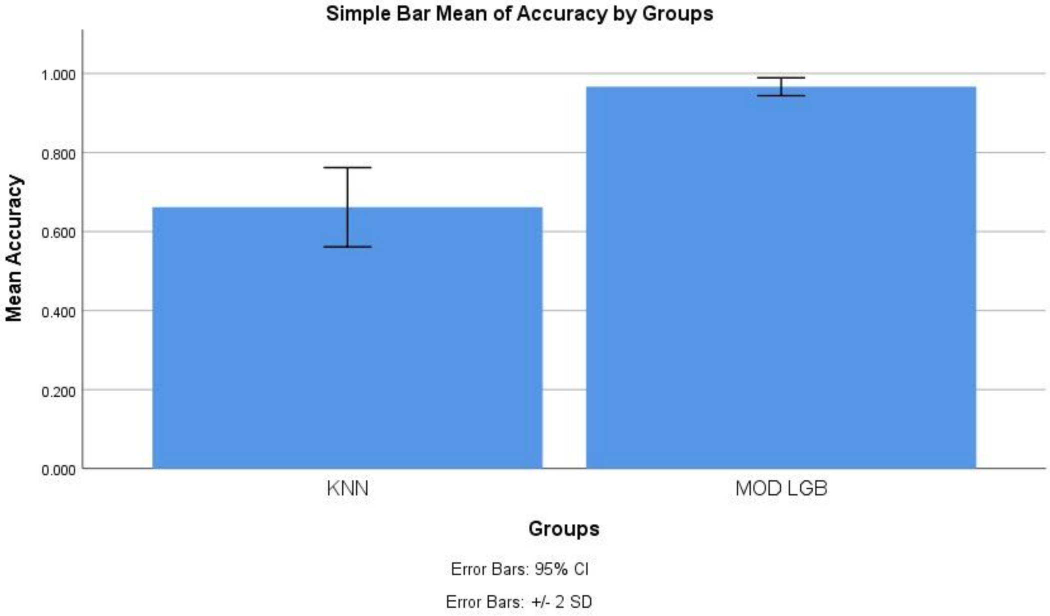


Fig. 3. The K-Nearest Neighbor and Modified Light Gradient Boost Model algorithms were designed and analyzed by comparing mean accuracy X axis: K-Nearest Neighbor (Group1), Modified Light Gradient Boost Model (Group2) and Y axis: Mean Accuracy, with +/- 1SD

Table 1. Group statistics of K-Nearest Neighbor and Modified Light Gradient Boost Model obtained for 20 samples each. The mean of K-Nearest Neighbor is 0.6615 and for Modified Light Gradient Boost Model is 0.96708

	Group	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	1	20	0.6615	0.050186	0.011222
	2	20	0.96708	0.011483	0.002568

Table 2. Group of some independent sample tests for Equality of Variances. The significance taken for this research is 0. The mean difference obtained is determined as 0.305575

Levene's Test for Equality of Variances				t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Accuracy	Equal variances assumed	41.423	0	-26.544	38	0	-0.305575	0.011512	-0.32888	-0.28227
	Equal variances not assumed			-26.544	20.984	0	-0.305575	0.011512	-0.329517	-0.281633

CHAPTER-4

Title page:

Improving Prediction Accuracy in Drift Detection Using Light Gradient Boost Model in Comparing with Modified Light Gradient Boost Model

N.Raja Likitha[1], T.J.Nagalakshmi[2]

N.Raja Likitha[1]
Research Scholar,
Department of Electronics and Communication Engineering,
Saveetha School of Engineering,
Saveetha University of Medical and Technical Sciences,
Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.
natharajalikitha19@saveetha.com

T. J. Nagalakshmi[2]
Project Guide, Corresponding Author,
Department of VLSI MicroElectronics,
Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences,
Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.
nagalakshmitj.sse@saveetha.com

Keywords: Innovative Concept Drift, Data Stream Mining, Drift Detection, Classifier, Technology, Deep Learning, Modified Light Gradient Boost Model, Light Gradient Boost Model

ABSTRACT

Aim: The proposed work aims to improve prediction accuracy in drift detection systems using the Light Gradient Boost Model when compared to a modified light gradient boost model. **Materials and Methods:** A total of 40 drift samples were collected using a variety of predictions. To evaluate the increasing prediction accuracy in drift detection, these samples were divided into two groups (Group 1- Light Gradient Boost Model, Group 2- Modified Light Gradient Boost Model) of each having 20 samples and the accuracy was calculated using Light Gradient Boost Model and a modified light gradient boost model. The G power is thought to be 80%. **Result:** The Light Gradient Boost Model produced 0.922 mean accuracy simulation results, while the Modified Light Gradient Boost Model Produced 0.96708 mean accuracy results. The obtained 2 tailed significance p-value is 0.0 which is less than standard significance p=0.05 (p<0.05), in terms of accuracy, indicating that there is a statistical significance between the two models. **Conclusion:** In terms of accuracy, the Light Gradient Boost Model performs significantly worse than the Modified Light Gradient Boost Model for the given dataset.

Keywords: Innovative Concept Drift, Data Stream Mining, Drift Detection, Classifier, Deep Learning, Technology, Modified Light Gradient Boost Model, Light Gradient Boost Model

INTRODUCTION

Innovative Concept Drift is the statistical term for a sudden change in the statistical characteristics of the target variable over time. The problem is that as time passes, the forecasts become less precise (Robin and Vrac, n.d.). A major issue with data stream analysis is the lack of a suitable mechanism for dealing with the continuous flow of data. A major issue with data stream analysis is the lack of a suitable approach for dealing with the constant flow of data with Deep Learning (Weißbach, Hilbert, and Springer 2020). It is a difficult field for data mining because the method must detect changes quickly in order to extract useful data from it (Murray, Agard, and Barajas 2017). It must keep up to date statistics on this changing data. The predictive model should be updated as input changes (Huang and Zhong 2018).

In the last decade, more than one hundred papers have emerged in Google Scholar, Springer, and IEEE Xplore. The majority of citations were found in 1–25 of 412 papers in IEEE Xplore and 18,600 in Google Scholar. In this citation they have done using classification-based stream mining in order to alert operators in the event of unintended system changes (Lughofer et al. 2015). Here this site uses a radical Innovative Concept Drift time expansion chamber (Thomas 1992). Here in this site they used collision detection on acceleration data (P, Nagalakshmi, and Jaanaa Rubavathy 2022) (Becker and Ebner 2019). Here by doing surveys in hospitals they give digital radiography systems with wireless detectors (Shelmerdine et al. 2018). In this site by using online they update initially with offline labs (Chua, Jordan, and Muller 2020).

In machine learning, the way data must be processed depends on specific characteristics determined by how data are accessed and their availability (Hartanto 2019; Thomas 1992). In the case of data streams, they differ from other forms of data in the sense that instances arrive continuously and sequentially. Over time, the underlying data distribution of the Data Stream mining may change dynamically (Korycki and Krawczyk 2018). When many models of base approaches are applied to a single input to improve overall performance, the term "ensemble" is used (Hartanto 2019). Similarly, ensembles are created in data stream mining by combining Bagging, Boosting, and Stacking techniques (Rosset 2005) . The change in Data Stream mining idea must be managed carefully due to its impact on prediction model accuracy. Innovative Concept Drift detection techniques collect such changes, allowing the prediction model to update at the Innovative Concept Drift rate (P. Wang, Jin, and Fehringer 2020).

The problem is that as time passes, the forecasts become less precise. A major issue with data stream analysis is the lack of a suitable mechanism for dealing with the continuous flow of data. So the proposed method to improve prediction accuracy in drift detection technology using Light Gradient Boost Model when compared to a modified light gradient boost model.

MATERIALS AND METHODS

The experiment was carried out in a computer lab at the Saveetha Institute of Medical and Technical Sciences' Department of VLSI, which is part of the Saveetha School of Engineering. The project is being worked on by two accuracy groups. With a sample size of 20 data points for each group, the G power is 80%. The Light Gradient Boost Model model belongs to Group 1, while the Modified Light Gradient Boost Model model belongs to Group 2. Google Collab was used to compare the accuracy of the results as well as the necessary algorithm.

Google Collab and code inspired by Collab are used to implement the data set. The Data Stream mining has been imported, and the visualization is finished. Following the visualization, the data set goes through a data preparation stage in which the error numbers from the Google Collab drive are compared to the code that has been mounted. The current classifier, Light Gradient Boost Model, is pitted against the Modified Light Gradient Boost Model in terms of accuracy.

Light Gradient Boost Model: Light Gradient Boost Model is a gradient boosting framework based on decision trees that improves model efficiency while reducing memory usage. It employs two novel techniques: Gradient-based One Side Sampling and Exclusive Feature Bundling (EFB), which overcome the limitations of the histogram-based algorithm used in all GBDT (Gradient Boosting Decision Tree) frameworks. The characteristics of the Light Gradient Boost Model Algorithm are formed by the two techniques of Deep Learning GOSS and EFB described below. They work together to make the model work efficiently and to give it a competitive advantage over other GBDT frameworks.

Modified Light Gradient Boost Model: Modified Light Gradient Boost Model is a gradient boosting framework based on decision trees that improves model efficiency while reducing memory usage in Deep Learning.

Testing Procedure

Figure. 1. depicts the workflow process, which includes the use of Google Collab-generated code to implement the data set. The data has been imported, and the visualization of the data is complete. Following visualization, the Data Stream mining set goes through data preparation, which compares the error numbers from the Google Collab drive to the mounted code. The Light Gradient Boost Model's accuracy was then compared to the current classifier, the Modified Light Gradient Boost Model.

STATISTICAL ANALYSIS

The proposed study work and the preceding work method's accuracy are both evaluated using the SPSS software package. Testing is carried out using a separate sample T-test. Group 1 has been taken as the Light Gradient Boost Model. Group 2 is taken as a Modified Light Gradient Boost Model. The independent variables are baseline information, properties of images. The dependent variable is accuracy (Kumar et al. 2021).

RESULTS

Compare the accuracy percentage values of the Light Gradient Boost Model with Modified Light Gradient Boost Model both algorithms provide different mean accuracy Modified Light Gradient Boost Model 0.96708 and Light Gradient Boost Model 0.922.

In Fig. 2. shows the bar chart compares accuracy values (Light Gradient Boost Model and Modified Light Gradient Boost Model), in that Light Gradient Boost Model gives accuracy about 0.922 where the Modified Light Gradient Boost Model gives about 0.96708. Here the error bar is +/-1 Standard Error. Figure 3: is the graphical representation with +/- 1 standard deviation.

Table. 1. shows the T-test tables. In the table comparing the two classifiers, the Light Gradient Boost Model (0.922) classifier has a higher mean value than the Modified Light Gradient Boost Model mean accuracy 0.96708 classifier, which was evaluated with the N=20 for a group. The standard deviation was revealed by the means that were different for both of the classifiers present.

Table 2. independent sample test. This table shows the independent sample T-test is performed for the two groups and has found that accuracy as(t =-7.266) & Mean Difference = (-0.045075) and it has the same standard error difference is 0.006764. Between the two groups , there would be a difference (means difference is -0.045075)(p<0.05).

DISCUSSION

According to the research project results, the Modified Light Gradient Boost Model classifier outperformed the Light Gradient Boost Model Algorithm with a 0.96708 mean accuracy rate in predicting the data with the most accurate values (p<0.05, independent variable test, SPSS IBM tool), which is considered better work. Previous research has shown that when it comes to determining the accuracy of the Modified Light Gradient Boost Model, Light Gradient Boost Model support is not superior to it.

This site gives the accuracy percentage for Light Gradient Boost Model is 91.73% (Harini and Sashi Rekha 2022).In this article based on machine learning the accuracy percentage of Light Gradient Boost Model is nearly 94% (Sharma et al., n.d.). This site gives accuracy for the Light Gradient Boost Model is 92%. By constructing a model from Deep Learning of the most recent access records, their goal is to develop a method for effectively detecting new unidentified attacks (Terado and Hayashida 2020). In this site the percentage of Light Gradient Boost Model algorithm accuracy is 59% is very less (“Music Genre Classification Using Linear Regression Compared with Extreme Gradient Boost Algorithm with Improved Accuracy” 2022). This gives the accuracy of 90% for the Light Gradient Boost Model (K. Wang et al. 2021).

This site gives an accuracy percentage of 95% nearly using statistical significance and sample tests with Deep Learning (Felix, Yovan Felix, and Sasipraba 2019; Subburaj et al. 2022) (Felix, Yovan Felix, and Sasipraba 2019). In this site they have many failures but at last they got the accuracy rate for the Light Gradient Boost Model is 78% (Sheng, Chen, and Tian 2018). In this site they have achieved an accuracy percentage of 98%for the Light Gradient Boost Model algorithm (Batchu and Seetha 2022).

There few problems experienced in this research work are Integration with Datasets Because Light Gradient Boost Model is susceptible to overfitting, it is simple to overfit small Data Stream mining and Negative gradients are taken into account in Gradient Boosting to improve the loss function, but Taylor's expansion is taken into account here.

CONCLUSION

The Innovative Concept Drift Detection technology for the Light Gradient Boost Model achieved an mean accuracy of 0.922, while the Modified Light Gradient Boost Model achieved an mean accuracy of 0.96708. When compared to the Light Gradient Boost Model algorithm, the accuracy of the Modified Light Gradient Boost Model algorithm is higher.

DECLARATIONS

Conflict of Interest

No conflict of interest in this manuscript.

Author Contributions

RL played a role in the planning, simulation, acquisition, evaluation, and drafting of the manuscript. TJNL contributed to the conceptualization, validation of data, and critical evaluation of the manuscript.

Acknowledgment

The authors express their gratitude to the Saveetha School of Engineering and the Saveetha Institute of Medical and Technical Sciences (formerly known as Saveetha University) for supplying the necessary resources that allowed for the successful completion of this research.

Funding

We would like to express our gratitude to the following organizations for giving the funding necessary for us to finish this research project.

1. Haneesh Sai Enterprises, Nellore.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

REFERENCES

Batchu, Raj Kumar, and Hari Seetha. 2022. “A Hybrid Detection System for DDoS Attacks Based on Deep Sparse Autoencoder and Light Gradient Boost Machine.” Journal of Information & Knowledge Management. https://doi.org/10.1142/s021964922250071x.

Becker, Felix, and Marc Ebner. 2019. “Collision Detection for a Mobile Robot Using Logistic Regression.” Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics. https://doi.org/10.5220/0007768601670173.

Chua, Adelson, Michael I. Jordan, and Rikky Muller. 2020. “Unsupervised Online Learning for Long-Term High Sensitivity Seizure Detection.” Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference 2020 (July): 528–31. https://doi.org/10.1109/EMBC44109.2020.9176122.

Felix, A. Yovan, A. Yovan Felix, and T. Sasipraba. 2019. “Flood Detection Using Gradient Boost Machine Learning Approach.” 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE). https://doi.org/10.1109/iccike47802.2019.9004419.

Harini, K., and K. K. Sashi Rekha. 2022. “Evaluating Performance Of Identifying At - Risk Students And Learning Achievement Model Using Accuracy And F-Measure by Comparing Logistic Regression, Generalized Linear Model And Gradient Boost Machine Algorithm.” 2022 International Conference for Advancement in Technology (ICONAT). https://doi.org/10.1109/iconat53423.2022.9725848.

Hartanto, Isnaeni Murdi. 2019. Integrating Multiple Sources of Information for Improving Hydrological Modelling: An Ensemble Approach. CRC Press. https://books.google.com/books/about/Integrating_Multiple_Sources_of_Informat.html?hl=&id=QceWDwAAQBAJ.

Huang, Shi-Ting, and Jessie Zhong. 2018. “ITMIG 2017—Dr. Giuseppe Giaccone: Focus on Data and Keep up-to-Date.” Mediastinum. https://doi.org/10.21037/med.2017.12.02.

Korycki, Lukasz, and Bartosz Krawczyk. 2018. “Clustering-Driven and Dynamically Diversified Ensemble for Drifting Data Streams.” 2018 IEEE International Conference on Big Data (Big Data). https://doi.org/10.1109/bigdata.2018.8622038.

Kumar, Sanjeev, M. J. P. Rohilkhand University, Bareilly, India., Ravendra Singh, M. J. P. Rohilkhand University, Bareilly, and India. 2021. “Comparative Analysis of Drift Detection Based Adaptive Ensemble Model with Different Drift Detection Techniques.” Journal of University of Shanghai for Science and Technology. https://doi.org/10.51201/jusst/21/06492.

Lughofer, Edwin, Eva Weigl, Wolfgang Heidl, Christian Eitzinger, and Thomas Radauer. 2015. “Drift Detection in Data Stream Classification without Fully Labelled Instances.” 2015 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS). https://doi.org/10.1109/eais.2015.7368802.

Murray, Paul W., Bruno Agard, and Marco A. Barajas. 2017. “Market Segmentation through Data Mining: A Method to Extract Behaviors from a Noisy Data Set.” Computers & Industrial Engineering. https://doi.org/10.1016/j.cie.2017.04.017.

“Music Genre Classification Using Linear Regression Compared with Extreme Gradient Boost Algorithm with Improved Accuracy.” 2022. Journal of Pharmaceutical Negative Results. https://doi.org/10.47750/pnr.2022.13.s04.198.

P, Jyothi Swaroop, T. J. Nagalakshmi, and S. Jaanaa Rubavathy. 2022. “Girl Child Security System Based on IOT Technology with Temperature Sensor Comparing with Fuzzy Classifier Based Safety Device.” In 2022 14th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS), 1–7. https://doi.org/10.1109/MACS56771.2022.10023417.

Robin, Yoann, and Mathieu Vrac. n.d. “Is Time a Variable like the Others in Multivariate Statistical Downscaling and Bias Correction?” https://doi.org/10.5194/esd-2021-12.

Rosset, Saharon. 2005. “Robust Boosting and Its Relation to Bagging.” Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining. https://doi.org/10.1145/1081870.1081900.

Sharma, Richa, G. K. Sharma, Manisha Pattanaik, and V. S. S. Prashant. n.d. “Structural and SCOAP Features Based Approach for Hardware Trojan Detection Using SHAP and Light Gradient Boosting Model.” https://doi.org/10.36227/techrxiv.21806022.

Shelmerdine, Susan C., Dean Langan, John C. Hutchinson, Melissa Hickson, Kerry Pawley, Joseph Suich, Liina Palm, et al. 2018. “Chest Radiographs versus CT for the Detection of Rib Fractures in Children (DRIFT): A Diagnostic Accuracy Observational Study.” The Lancet. Child & Adolescent Health 2 (11): 802–11. https://doi.org/10.1016/S2352-4642(18)30274-8.

Sheng, Peng, Li Chen, and Jing Tian. 2018. “Learning-Based Road Crack Detection Using Gradient Boost Decision Tree.” 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA). https://doi.org/10.1109/iciea.2018.8397897.

Subburaj, T., T. J. Nagalakshmi, N. Krishnamoorthy, J. Uthayakumar, R. Thiyagarajan, and S. Arun. 2022. “Descriptive Analytics Solution for Attack Detection by Utilizing DL Strategies.” In 2022 Smart Technologies, Communication and Robotics (STCR), 1–5. https://doi.org/10.1109/STCR55312.2022.10009596.

Terado, Ryosuke, and Morihiro Hayashida. 2020. “Improving Accuracy and Speed of Network-Based Intrusion Detection Using Gradient Boosting Trees.” Proceedings of the 6th International Conference on Information Systems Security and Privacy. https://doi.org/10.5220/0008963504900497.

Thomas, J. H. 1992. “Notes on Radial Drift for Dalitz Detection.” https://doi.org/10.2172/6789418.

Wang, Kun, Jie Lu, Anjin Liu, Guangquan Zhang, and Li Xiong. 2021. “Evolving Gradient Boost: A Pruning Scheme Based on Loss Improvement Ratio for Learning Under Concept Drift.” IEEE Transactions on Cybernetics. https://doi.org/10.1109/tcyb.2021.3109796.

Wang, Pingfan, Nanlin Jin, and Gerhard Fehringer. 2020. “Concept Drift Detection with False Positive Rate for Multi-Label Classification in IoT Data Stream.” 2020 International Conference on UK-China Emerging Technologies (UCET). https://doi.org/10.1109/ucet51115.2020.9205421.

Weißbach, Manuel, Hannes Hilbert, and Thomas Springer. 2020. “Performance Analysis of Continuous Binary Data Processing Using Distributed Databases within Stream Processing Environments.” Proceedings of the 10th International Conference on Cloud Computing and Services Science. https://doi.org/10.5220/0009413301380149.

Hartanto, Isnaeni Murdi. 2019. Integrating Multiple Sources of Information for Improving Hydrological Modelling: An Ensemble Approach. CRC Press.

Huang, Shi-Ting, and Jessie Zhong. 2018. “ITMIG 2017—Dr. Giuseppe Giaccone: Focus on Data and Keep up-to-Date.” Mediastinum. https://doi.org/10.21037/med.2017.12.02.

Murray, Paul W., Bruno Agard, and Marco A. Barajas. 2017. “Market Segmentation through Data Mining: A Method to Extract Behaviors from a Noisy Data Set.” Computers & Industrial Engineering. https://doi.org/10.1016/j.cie.2017.04.017.

Weißbach, Manuel, Hannes Hilbert, and Thomas Springer. 2020. “Performance Analysis of Continuous Binary Data Processing Using Distributed Databases within Stream Processing Environments.” Proceedings of the 10th International Conference on Cloud Computing and Services Science. https://doi.org/10.5220/0009413301380149.

Figure and Tables

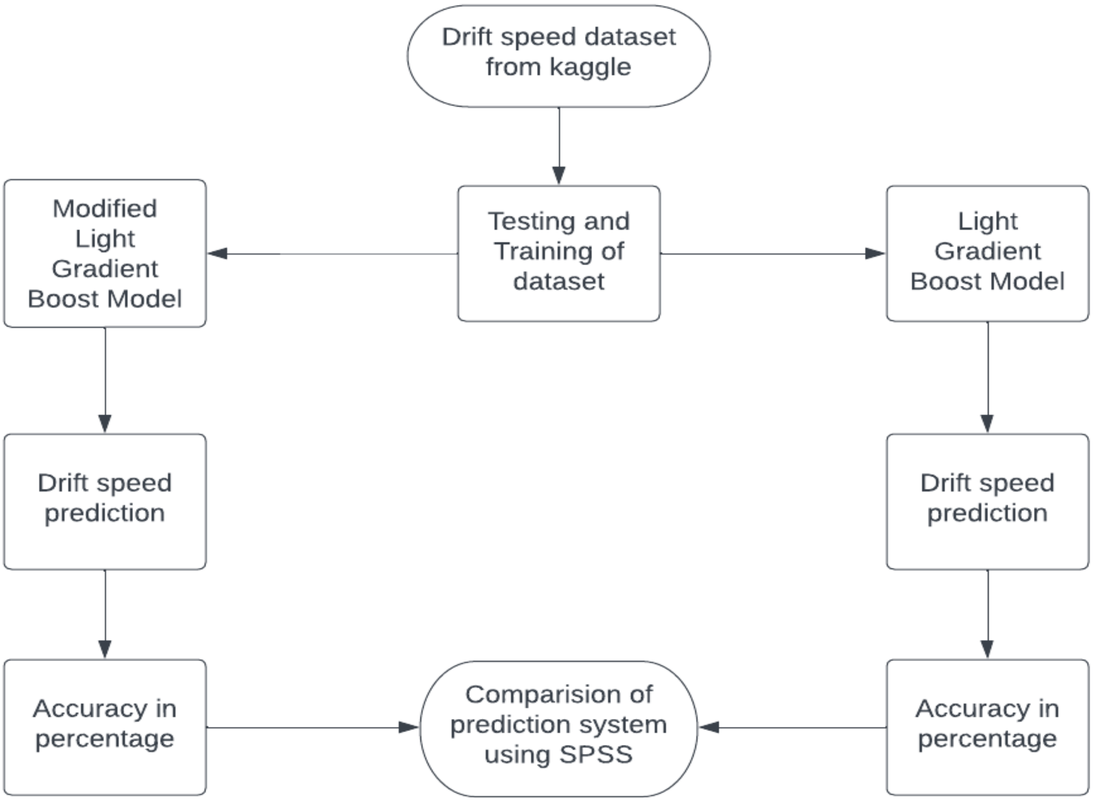

Drift speed dataset from kaggle
Testing and Training of dataset
Modified Light Gradient Boost Model
Light Gradient Boost Model
Drift speed prediction
Drift speed prediction
Accuracy in percentage
Comparision of prediction system using SPSS
Accuracy in percentage

Fig. 1. Process flow the accuracy finding using modified Light Gradient Boost Model which is starting from the basic process of importing the data to the program to giving results of the accuracy.

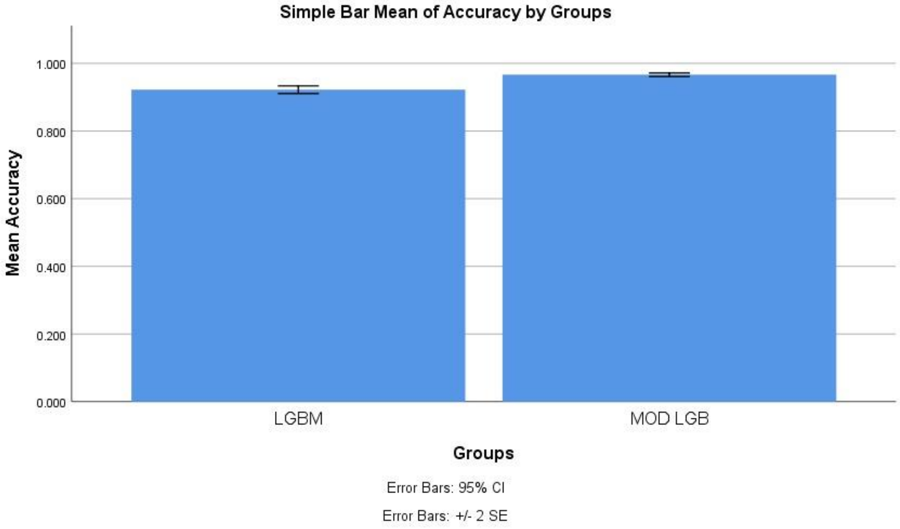


Fig. 2. The Light Gradient Boost Model and Modified Light Gradient Boost Model algorithms were designed and analyzed by comparing mean accuracy X axis: Light Gradient Boost Model (Group1), Modified Light Gradient Boost Model (Group2) and Y axis: Mean Accuracy, with +/-SE.

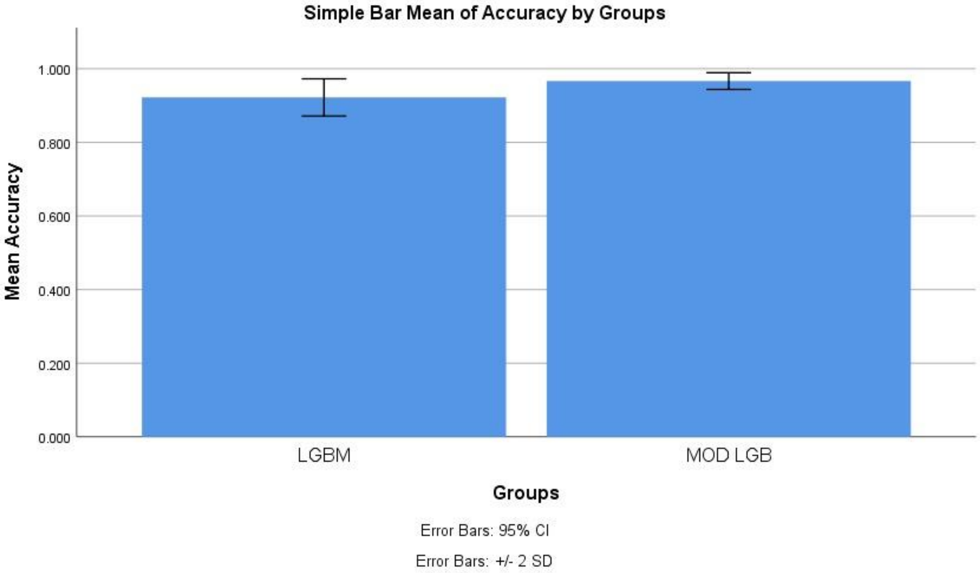


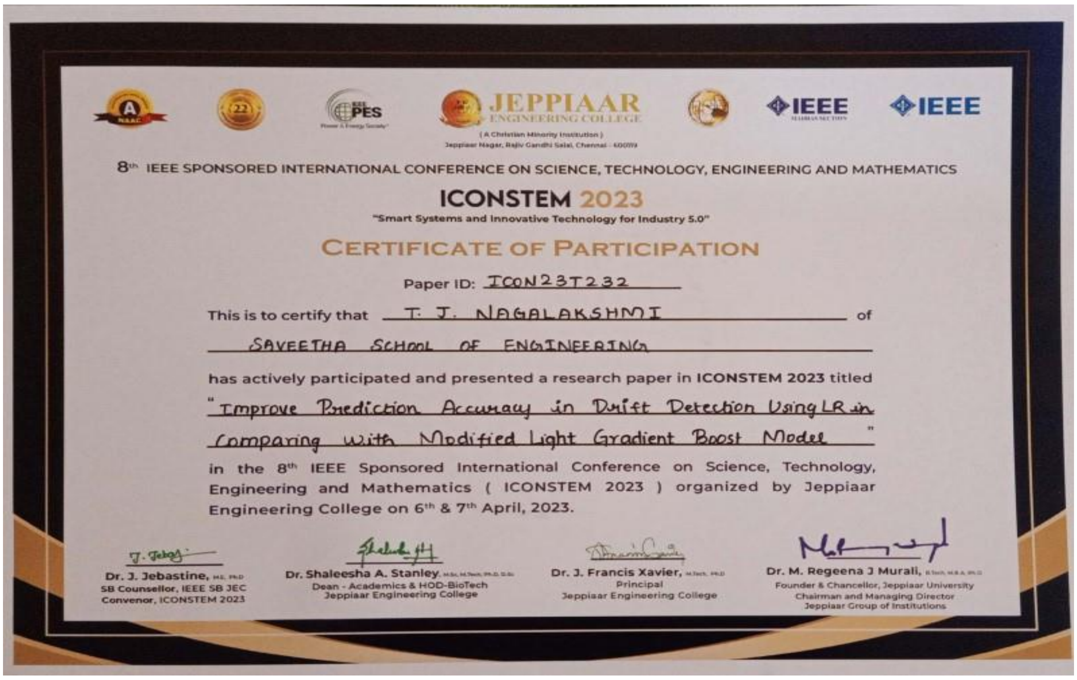
Fig. 3. The Light Gradient Boost Model and Modified Light Gradient Boost Model algorithms were designed and analyzed by comparing mean accuracy X axis: Light Gradient Boost Model (Group1), Modified Light Gradient Boost Model (Group2) and Y axis: Mean Accuracy, with +/- 1SD.

Table. 1. Group statistics of Light Gradient Boost Model and Modified Light Gradient Boost Model obtained for 20 samples each. The mean of the Light Gradient Boost Model is 0.922 and for the Modified Light Gradient Boost Model is 0.96708.

	Group	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	1	20	0.922	0.025257	0.005648
	2	20	0.96708	0.011483	0.002568

Table. 2. Group of some independent sample tests for Equality of Variances. The significance taken for this research is 0. The mean difference obtained is determined as 0.045075.

Levene's Test for Equality of Variances				t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Accuracy	Equal variances assumed	20.274	0	-7.266	38	0	-0.045075	0.006204	-0.057634	-0.032516
	Equal variances not assumed	-	-	-7.266	26.533	0	-0.045075	0.006204	-0.057815	-0.032335



IEEE PES Power & Energy Society

JEPPIAAR ENGINEERING COLLEGE
(A Christian Minority Institution)
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600119

8th IEEE SPONSORED INTERNATIONAL CONFERENCE ON SCIENCE, TECHNOLOGY, ENGINEERING AND MATHEMATICS

ICONSTEM 2023

"Smart Systems and Innovative Technology for Industry 5.0"

CERTIFICATE OF PARTICIPATION

Paper ID: ICON23T232

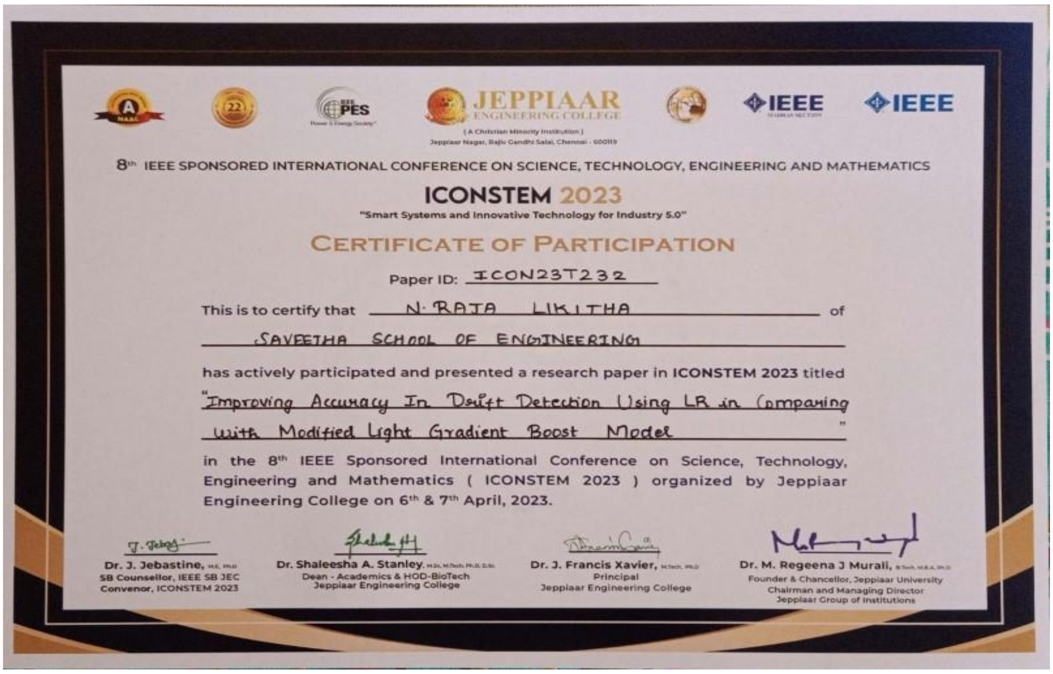
This is to certify that T. J. NAGALAKSHMI of SAVEETHA SCHOOL OF ENGINEERING has actively participated and presented a research paper in ICONSTEM 2023 titled "Improve Prediction Accuracy in Drift Detection Using LR in Comparing with Modified Light Gradient Boost Model" in the 8th IEEE Sponsored International Conference on Science, Technology, Engineering and Mathematics (ICONSTEM 2023) organized by Jeppiaar Engineering College on 6th & 7th April, 2023.

J. Jebas
Dr. J. Jebastine, M.E., Ph.D
SB Counsellor, IEEE SB JEC
Convenor, ICONSTEM 2023

Dr. Shaleesha A. Stanley, M.Sc, M.Tech, Ph.D, D.Sc
Dean - Academics & HOD-BioTech
Jeppiaar Engineering College

Dr. J. Francis Xavier, M.Tech, Ph.D
Principal
Jeppiaar Engineering College

Dr. M. Regeena J Murali, B.Tech, M.B.A, Ph.D
Founder & Chancellor, Jeppiaar University
Chairman and Managing Director
Jeppiaar Group of Institutions



IEEE PES Power & Energy Society

JEPPIAAR ENGINEERING COLLEGE
(A Christian Minority Institution)
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600119

8th IEEE SPONSORED INTERNATIONAL CONFERENCE ON SCIENCE, TECHNOLOGY, ENGINEERING AND MATHEMATICS

ICONSTEM 2023

"Smart Systems and Innovative Technology for Industry 5.0"

CERTIFICATE OF PARTICIPATION

Paper ID: ICON23T232

This is to certify that N. RAJA LIKITHA of SAVEETHA SCHOOL OF ENGINEERING has actively participated and presented a research paper in ICONSTEM 2023 titled "Improving Accuracy In Drift Detection Using LR in Comparing with Modified Light Gradient Boost Model" in the 8th IEEE Sponsored International Conference on Science, Technology, Engineering and Mathematics (ICONSTEM 2023) organized by Jeppiaar Engineering College on 6th & 7th April, 2023.

J. Jebas
Dr. J. Jebastine, M.E., Ph.D
SB Counsellor, IEEE SB JEC
Convenor, ICONSTEM 2023

Dr. Shaleesha A. Stanley, M.Sc, M.Tech, Ph.D, D.Sc
Dean - Academics & HOD-BioTech
Jeppiaar Engineering College

Dr. J. Francis Xavier, M.Tech, Ph.D
Principal
Jeppiaar Engineering College

Dr. M. Regeena J Murali, B.Tech, M.B.A, Ph.D
Founder & Chancellor, Jeppiaar University
Chairman and Managing Director
Jeppiaar Group of Institutions

https://drive.google.com/file/d/1eZwXo5kwqwb2qwTG-SiVCXGRTopYJt6t/view?usp=drive_link