BALTIC JOURNAL OF LAW & POLITICS
A Journal of Vytautas Magnus University
VOLUME 15, NUMBER 4 (2022)
ISSN 2029-0454

sciendo

Cite: *Baltic Journal of Law & Politics* 15:4 (2022): 175-184
DOI: 10.2478/bjlp-2022-004018

Implementation of an Innovative Adaptive Logistic Regression Algorithm to Improve the Early Identification Rate of Seizure in Comparison with Naive Bayes Algorithm

V. Narendra Reddy
Research Scholar, Department of Electronics and Communication Engineering , Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu. India. Pincode: 602105.

R. Priyanka
Project Guide, Corresponding Author, Department of Electronics and Communication Engineering , Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu. India. Pincode: 602105.

Received: August 8, 2022; reviews: 2; accepted: November 29, 2022.

Abstract

Aim: The aim of the research is to improve the early identification rate of epileptic seizures using the Innovative Adaptive logistic regression algorithm in comparison with naive bayes algorithm. **Methods and Materials:** The total of 6426 samples are collected from the UCI repository. Group 1 represents the Innovative Adaptive logistic regression algorithm and group 2 represents the naive bayes algorithm. The G power calculation was done with 80% of power and alpha of 0.05. Each group with 20 samples were taken for SPSS analysis. **Results:** The Innovative Adaptive logistic regression algorithm has achieved the accuracy, Precision, Recall and Specificity of 97.0 %, 95 %, 90 %, and 89 % when compared to naive bayes algorithm with 78 %, 94 %, 78 % and 74 %. The logistic regression algorithm has achieved the significance of 0.043 ($p < 0.05$). **Conclusion:** In this study it is concluded that the logistic regression algorithm has significantly greater accuracy when compared with the naive bayes algorithm.

KEYWORDS
Seizure Detection, Naive Bayes, Innovative Adaptive Logistic Regression Algorithm, Machine Learning, Early Identification Rate, Classification Accuracy, Outbreak Prediction.

INTRODUCTION

Epilepsy Seizures are caused by a disease in which nerve cell activity in the brain is disrupted. Epilepsy can be caused by a genetic condition or an acquired brain injury such as a stroke or trauma is a disease that causes frequent seizures and it is one of the most prevalent severe neurological conditions. Prediction and detection at the early stage is mandatory to avoid serious damage to the human brain cells and accurate detection of genetic epilepsy also plays a major role in today's situation (Logesparan, Rodriguez-Villegas, and Casson 2015). Electroencephalography (EEG) is useful for identifying epilepsy since it detects variations in voltage fluctuations across electrodes around the subject's scalp and provides temporal and spatial detail about the brain (Satapathy et al.

2019) (Siuly, Li, and Zhang 2017). In this study the early identification rate of it's concluded that the novel logistic regression algorithm has significantly greater classifications accuracy when compared with the naive bayes algorithm. The major applications of the EEG is considered the main Neurophysiological study used in the Identifying (Paul and Bhatia 2020). Diseases and Diagnosis, Drug Discovery and Manufacturing, Smart Health Records, Medical Imaging Diagnosis, outbreak prediction and iris monitoring (Thrun, Saul, and Schölkopf 2004) (Varma, Deekshitha Varma, and Priyanka 2022).

In the past five years, numerous research papers on epilepsy diagnosis have been published. In IEEE xplore 19 articles were published and 101 articles were published in google scholar. (Ein Shoka et al. 2021) proposed an EEG classification method and performed a cognitive task and determined the seizure occurrence and achieved the better detection rate. (Hu and Zhang 2019) proposed KNN classification and achieved the detection rate of in terms of accuracy. (Paul, Bhattacharya, and Bit 2019) used an independent complement analysis method for feature selection, because of unrelated feature selection the percentage of epilepsy detection in terms of precision, sensitivity, and specificity. (Ghazali et al. 2019) proposed comparison between Naive Bayes and KNN, Naive Bayes Provides high identification rate. (Abraham et al. 2018) comparison classification algorithms between Decision tree, KNN, Bayesian, and concluded that the Decision Tree's algorithms are more accurate. (Chowdhury et al. 2021) proposed and compared LNB with KNN, KNNDW and LWNB it showed that ICLNB algorithm performs significantly more accuracy.

Previously our team has a rich experience in working on various research projects across multiple disciplines (Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020).The existing system has a drawback in that EEG signals are sampled at different sampling frequencies to remove artifacts, which results in the loss of major attributes and reduces prediction speed and classification accuracy. Therefore, the proposed system focused on selecting data sampled at the same frequency and validating it using an innovative adaptive logistic regression algorithm rather than a naive bayes algorithm.

MATERIALS AND METHODS

The proposed research was conducted in the Transducer lab, Department of Electronics and Communication Engineering, Saveetha School of Engineering at Saveetha Institute of Medical and Technical Sciences. This analysis consists of two different groups, group 1 as a logistic regression algorithm and group 2 as a naive bayes algorithm. A total of 6426 data samples were taken from two groups. The group 1 with 20 samples and group 2 with 20 samples were used in this analysis. The sample size was estimated for each group using a G Power calculator with 80% of pretest power, alpha error of 0.95 (Bugeja, Garg, and Audu 2016a).

Logistic regression Algorithm
Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.

The steps involved in Logistic regression algorithm :

Step-1:Data Preprocessing step.
Step-2:Fitting Logistic Regression to the Training set.
Step-3:Predicting the test result.
Step-4:Test accuracy of the result (Creation of Confusion matrix)
Step-5:Visualising the test set result.

 ISSN 2029-0454

Naive Bayes Algorithm

The Naive Bayes Classifier is a supervised learning technique that uses Bayes theorem with intermediate naïve assumptions to perform probabilistic classification represented in Equation (1). Here's how the Bayes rules work:

$$P(Bj\,|A) = P(Bj)P(A|Bj)\,P(A) \qquad (1)$$

Step 1: Pre-process/prepare the data with a naïve bayes classifier so that we can use it effectively in our code.
Step 2: Use the Naive Bayes Classifier model to fit the data to the Training set.
Step 3: Use the predict function to generate predictions about the test set outcomes.
Step 4: Using the Confusion matrix, test the accuracy of the Naive Bayes classifier.
Step 5: Using the Naive Bayes Classifier, visualise the training set results.

Python is used to set up the testing setup for the logistic regression and Naive Bayes algorithms (2021). The PYTHON (Subasi 2020) programme requires an 8th generation Intel i5 processor with 4GB of RAM as a minimum system requirement. It offers a user-friendly environment for algorithm development, data visualisation, data analysis, and numerical calculation. Equations (2) to (5) are useful for determining the suggested algorithm's accuracy, sensitivity, specificity, and recall.

Accuracy = (TP + TN) /(TN + FP +TP +FN) (2)

Sensitivity = TP/(TP + FN) (3)

Specificity = TN/(TN + FP) (4)

Recall = TP/(TP + FN) (5)

Statistical Analysis

For the statistical analysis of this proposed work, IBM SPSS (Bhattacharya and Chakraborty 2020) was employed. Independent sample t-tests are used to examine the dependent and independent variables. For classification outbreak prediction of accuracy, specificity, precision, recall, and Mean Deviation, independent sample t-tests were used. Mean, Variance, Standard Deviation are considered as independent variables, whereas the Accuracy, Precision, Specificity and Recall are considered dependent variables (Hu and Zhang 2019).

RESULTS

The comparison features extraction techniques used with innovative adaptive logistic regression algorithms. In this Adaptive logistic regression algorithm obtains significantly better accuracy in comparison with the naive bayes algorithm for Early Identification Rate. Figure 1 represnts the categorical dependent variable's outcome is outbreak prediction using logistic regression. Instead of fitting a regression line, a "S '' shaped logistic function can be fitted, which predicts two minimum values of 0.3 and a maximum value of 0.8. Fig. 2 shows the y-axis indicates the mean percentage of classification accuracy, precision, specificity, and recall for the group of naive bayes and logistic regression, with the Error Bars being 95% CI and the Error Bars being +/-2SD.
Table 1 shows the selected classifiers using the logistic regression algorithm, the practical values of accuracy, precision, recall, and specificity are 97%, 95%, 90%, 89%, and for the naive bayes algorithm, 78 %, 94 %, 78 %, 74 %. The logistic algorithm outperforms the naïve bayes algorithm in terms of accuracy and specificity. Table 2 denotes the statistical analysis of logistic regression algorithm and naive bayes algorithm with the

mean accuracy, standard deviation and standard error mean where logistic regression algorithms has 95.60 of mean accuracy, 7.450 of standard deviation and 3.332 standard error. Table 3 represents the t-test using independent samples for determining significance and standard error. P values of less than (P<0.05) are deemed statistically significant, and 95% confidence intervals are used to calculate the results.

DISCUSSION

In this analysis, performance of Adaptive Logistic regression Algorithm and Naive Bayes Algorithm is analysed in this work, with an accuracy of 97 %, precision of 95 %, specificity of 90 %, and recall of 89 % (P<0.05). In comparison to the Naive Bayes Algorithm, the suggested work shows that the Logistic regression Algorithm performs better classification.

For the datasets considered in this study, the naive bayes algorithm is able to classify Epileptic seizures with moderate classification of outbreak prediction accuracy of 94% (Bugeja, Garg, and Audu 2016b). This study is concerned with a comparative analysis between the algorithms SVM and KNN employed for EEG based epileptic seizure identification. Furthermore, this study will focus on patient-independent classifiers as they are more complicated due to the EEG variability and achieved classification accuracy of 93.2% (D'Elios and Rizzi 2018). This work aims to find a classifier for the purpose of personalised seizure detection using K-Nearest Neighbour and Support Vector Machines as classifiers. Both methods achieved accuracies over 80 percent, with KNN performing a bit better than SVM, and an onset sensitivity of 100 percent when tested on ten patients (Esposito et al. 2019; Siddiqui et al. 2020). (Lutsenko, Pechurina, and Sergeev 2018) The Reveal Algorithm uses the same data from UCI repository, with accuracy considerably higher to 76,7%. Proposed the patient specific classifiers with the accuracy significantly lower, than the expected outbreak prediction classification accuracy of 61% (Esposito et al. 2019; S et al. 2019).

The analysis made by (Islam et al. 2019; Malik and Amin 2017) they have used the raw EEG data of 20 epileptic patients who were under treatment. The goal of their work was to create a system for automatically searching for the optimum DWT settings in order to increase classification accuracy of outbreak prediction (accuracy > 90%) and reduce the computational expense of seizure detection (Chen et al. 2017). The main idea behind our study is to analyse and depict limitations of existing machine learning approaches that are being used in solving the data imbalance problem and to propose a highly efficient solution to resolve this issue (Yuan et al. 2017).
The Logistics Regression algorithm is not suitable for data sets with more number of attributes and also the artifacts present in the dataset make it not perform well. In future preprocessing of data should be done to improve the accuracy of the machine learning classifier.

CONCLUSION

The detection and prediction of epilepsy seizure at an early identification rate is done using adaptive logistic regression. The machine learning classifier logistic regression algorithm gives (18% higher) significantly better accuracy than the naive bayes algorithm.

DECLARATIONS

Conflict of interests
No conflict of interests in this manuscript.
Authors Contributions
Author NR was involved in data collection, data analysis, manuscript writing. Author RP was involved in conceptualization, data validation, and critical review of manuscript.

Acknowledgements
The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding: We thank the following organisations for providing financial support that enabled us to complete the study.
1.VK Technologies, Chennai.
2.Saveetha University
3.Saveetha Institute of Medical and Technical Sciences.
4.Saveetha School of Engineering.

REFERENCES

Abraham, Ajith, Paramartha Dutta, Jyotsna Kumar Mandal, Abhishek Bhattacharya, and Soumi Dutta. 2018. *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2018, Volume 2*. Springer.

Benin, S. R., S. Kannan, Renjin J. Bright, and A. Jacob Moses. 2020. "A Review on Mechanical Characterization of Polymer Matrix Composites & Its Effects Reinforced with Various Natural Fibres." *Materials Today: Proceedings* 33 (January): 798–805.

Bhattacharya, Sanjukta, and Chinmay Chakraborty. 2020. "Machine Learning for Biomedical and Health Informatics." *Big Data, IoT, and Machine Learning*. https://doi.org/10.1201/9780429322990-5.

Bugeja, Sylvia, Lalit Garg, and Eliazar E. Audu. 2016a. "A Novel Method of EEG Data Acquisition, Feature Extraction and Feature Space Creation for Early Detection of Epileptic Seizures." *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* 2016 (August): 837–40.

———. 2016b. "A Novel Method of EEG Data Acquisition, Feature Extraction and Feature Space Creation for Early Detection of Epileptic Seizures." *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* 2016 (August): 837–40.

Chen, Duo, Suiren Wan, Jing Xiang, and Forrest Sheng Bao. 2017. "A High-Performance Seizure Detection Algorithm Based on Discrete Wavelet Transform (DWT) and EEG." *PloS One* 12 (3): e0173138.

Chowdhury, Nakib Hayat, Mamun Bin Ibne Reaz, Fahmida Haque, Shamim Ahmad, Sawal Hamid Md Ali, Ahmad Ashrif A Bakar, and Mohammad Arif Sobhan Bhuiyan. 2021. "Performance Analysis of Conventional Machine Learning Algorithms for Identification of Chronic Kidney Disease in Type 1 Diabetes Mellitus Patients." *Diagnostics (Basel, Switzerland)* 11 (12). https://doi.org/10.3390/diagnostics11122267.

D'Elios, Mario Milco, and Marta Rizzi. 2018. *Humoral Primary Immunodeficiencies*. Springer.

Ein Shoka, Athar A., Monagi H. Alkinani, A. S. El-Sherbeny, Ayman El-Sayed, and Mohamed M. Dessouky. 2021. "Automated Seizure Diagnosis System Based on Feature Extraction and Channel Selection Using EEG Signals." *Brain Informatics* 8 (1): 1.

Esposito, Anna, Marcos Faundez-Zanuy, Francesco Carlo Morabito, and Eros Pasero. 2019. *Neural Approaches to Dynamics of Signal Exchanges*. Springer Nature.

Ghazali, Rozaida, Nazri Mohd Nawi, Mustafa Mat Deris, and Jemal H. Abawajy. 2019. *Recent Advances on Soft Computing and Data Mining: Proceedings of the Fourth International Conference on Soft Computing and Data Mining (SCDM 2020), Melaka, Malaysia, January 22–□23, 2020*. Springer Nature.

Gudipaneni, Ravi Kumar, Mohammad Khursheed Alam, Santosh R. Patil, and Mohmed

Isaqali Karobari. 2020. "Measurement of the Maximum Occlusal Bite Force and Its Relation to the Caries Spectrum of First Permanent Molars in Early Permanent Dentition." *The Journal of Clinical Pediatric Dentistry* 44 (6): 423–28.

Hu, Li, and Zhiguo Zhang. 2019. *EEG Signal Processing and Feature Extraction*. Springer Nature.

Islam, Rafiqul, Yun Sing Koh, Yanchang Zhao, Graco Warwick, David Stirling, Chang-Tsun Li, and Zahidul Islam. 2019. *Data Mining: 16th Australasian Conference, AusDM 2018, Bahrurst, NSW, Australia, November 28–30, 2018, Revised Selected Papers*. Springer.

Logesparan, Lojini, Esther Rodriguez-Villegas, and Alexander J. Casson. 2015. "The Impact of Signal Normalization on Seizure Detection Using Line Length Features." *Medical & Biological Engineering & Computing* 53 (10): 929–42.

Lutsenko, Y. V., E. K. Pechurina, and A. E. Sergeev. 2018. "DEVELOPING A VETERINARY TEST FOR THE DIAGNOSIS OF GASTROINTESTINAL DISEASES IN HORSES BASED ON DATA FROM THE UCI REPOSITORY WITH THE USE OF ASC-ANALYSIS." *Polythematic Online Scientific Journal of Kuban State Agrarian University*. https://doi.org/10.21515/1990-4665-141-033.

Malik, Aamir Saeed, and Hafeez Ullah Amin. 2017. *Designing EEG Experiments for Studying the Brain: Design Code and Example Datasets*. Academic Press.

Nalini, Devarajan, Jayaraman Selvaraj, and Ganesan Senthil Kumar. 2020. "Herbal Nutraceuticals: Safe and Potent Therapeutics to Battle Tumor Hypoxia." *Journal of Cancer Research and Clinical Oncology* 146 (1): 1–18.

Paul, Sudip, and Dinesh Bhatia. 2020. *Smart Healthcare for Disease Diagnosis and Prevention*. Academic Press.

Paul, Sudip, Pallab Bhattacharya, and Arindam Bit. 2019. *Early Detection of Neurological Disorders Using Machine Learning Systems*. IGI Global.

Reddy, Poornima, Jogikalmat Krithikadatta, Valarmathi Srinivasan, Sandhya Raghu, and Natanasabapathy Velumurugan. 2020. "Dental Caries Profile and Associated Risk Factors Among Adolescent School Children in an Urban South-Indian City." *Oral Health & Preventive Dentistry* 18 (1): 379–86.

Satapathy, Sandeep Kumar, Satchidananda Dehuri, Alok Kumar Jagadev, and Shruti Mishra. 2019. *EEG Brain Signal Classification for Epileptic Seizure Disorder Detection*. Academic Press.

Sathish, T., and S. Karthick. 2020. "Gravity Die Casting Based Analysis of Aluminum Alloy with AC4B Nano-Composite." *Materials Today: Proceedings* 33 (January): 2555–58.

Sathish, T., D. Bala Subramanian, R. Saravanan, and V. Dhinakaran. 2020. "Experimental Investigation of Temperature Variation on Flat Plate Collector by Using Silicon Carbide as a Nanofluid." In *PROCEEDINGS OF INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MECHANICAL AND MATERIALS ENGINEERING: ICRTMME 2019*. AIP Publishing. https://doi.org/10.1063/5.0024965.

Siddiqui, Mohammad Khubeb, Ruben Morales-Menendez, Xiaodi Huang, and Nasir Hussain. 2020. "A Review of Epileptic Seizure Detection Using Machine Learning Classifiers." *Brain Informatics*. https://doi.org/10.1186/s40708-020-00105-1.

Siuly, Siuly, Yan Li, and Yanchun Zhang. 2017. *EEG Signal Analysis and Classification: Techniques and Applications*. Springer.

Sivasamy, Ramesh, Potu Venugopal, and Rodrigo Espinoza-González. 2020. "Structure, Electronic Structure, Optical and Magnetic Studies of Double Perovskite Gd2MnFeO6 Nanoparticles: First Principle and Experimental Studies." *Materials Today Communications* 25 (December): 101603.

S, Niveditha, S. Niveditha, S. V. Jagmohan, Abhishek K. Verma, and Minni Meka. 2019. "Diagnostic Accuracy of Xpert MTB Compared to Smear Microscopy in Pulmonary vs Extrapulmonary Tuberculosis." *International Journal of Contemporary Medical Research [IJCMR]*. https://doi.org/10.21276/ijcmr.2019.6.7.1.

Thrun, Sebastian, Lawrence K. Saul, and Bernhard Schölkopf. 2004. *Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference*. MIT Press.

Varma, D. R. Deekshitha, D. R. Deekshitha Varma, and R. Priyanka. 2022. "Performance

Analysis of Novel Iris Monitoring System Based on Canny Detection Algorithm in Comparison with Prewitt Algorithm." *2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)*. https://doi.org/10.1109/iciptm54933.2022.9754088.

Venu, Harish, and Prabhu Appavu. 2021. "Experimental Studies on the Influence of Zirconium Nanoparticle on Biodiesel–diesel Fuel Blend in CI Engine." *International Journal of Ambient Energy* 42 (14): 1588–94.

Yuan, Qi, Weidong Zhou, Liren Zhang, Fan Zhang, Fangzhou Xu, Yan Leng, Dongmei Wei, and Meina Chen. 2017. "Epileptic Seizure Detection Based on Imbalanced Classification and Wavelet Packet Transform." *Seizure*. https://doi.org/10.1016/j.seizure.2017.05.018.

TABLES AND FIGURES

Table 1. For selected classifiers using the logistic regression algorithm, the practical values of classifiactions accuracy, precision, recall, and specificity are 97%, 95%, 90%, 89%, and for the naive bayes algorithm, 78%, 94%, 78%, 74%. The logistic algorithm outperforms the naïve bayes algorithm in terms of accuracy and specificity.

Algorithm	**Accuracy(%)**	**precision(%)**	**Recall(%)**	**Specificity(%)**
Logistic regression	97%	95%	90%	89%
Naive Bayes	78%	94%	78%	74%

Table 2. The Group Statistical analysis of mean, std.deviation, std.error mean for logistic regression algorithm is 95.60, 7.450, 3.332 and for naive bayes algorithm is 77.00, .548, .245 It is observed that the Logistic regression algorithm performed significantly better than the naive bayes algorithm.

Group Statistics					
	Group	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	Logistic regression	20	95.60	7.450	3.332
	Naive Bayes	20	77.00	.548	.245
precision	Logistic regression	20	91.20	13.982	6.253
	Naive Bayes	20	93.00	.447	.200
specificity	Logistic regression	20	89.80	20.376	9.113

ISSN 2029-0454
2022

	Naive Bayes	20	12.20	1.924	.860
recall	Logistic regression	20	88.20	30.627	13.697
	Naive Bayes	20	77.00	2.775	1.241

Table 3. T-test using independent samples for determining significance and standard error. P values of less than (P<0.05) are deemed statistically significant, and 95% confidence intervals are used to calculate the results.

Independent Samples Test											
		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Sig. (one-side dp)	Sig. (two-side dp)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
										Lower	Upper
Accuracy	Equal variances assumed	5.130	.043	-5.568	8	.091	.182	-3.000	2.054	-7.737	1.737
	Equal variances not assumed			-5.568	4.315	.106	.213	-3.000	2.054	-8.543	2.543
Precision	Equal variances assumed	6.639	.033	-.671	8	.021	.043	4.200	1.744	.179	8.221

ISSN 2029-0454
2022

	Equal variances not assumed			-.671	4.008	.025	.051	4.200	1.744	-013	8.413
Specificity	Equal variances assumed	5.871	.042	-8.478	8	.026	.053	1.400	.616	-.022	2.822
	Equal variances not assumed			-8.4	4.0	.02	.05	1.400	.616	-.039	2.839
Recall	Equal variances assumed	7.483	.026	-.887	8	.048	.095	-81.400	9.726	-40.829	4.029
	Equal variances not assumed			-8.87	4.066	.065	.130	-18.400	9.726	-45.240	8.440

ISSN 2029-0454
2022

GRAPH

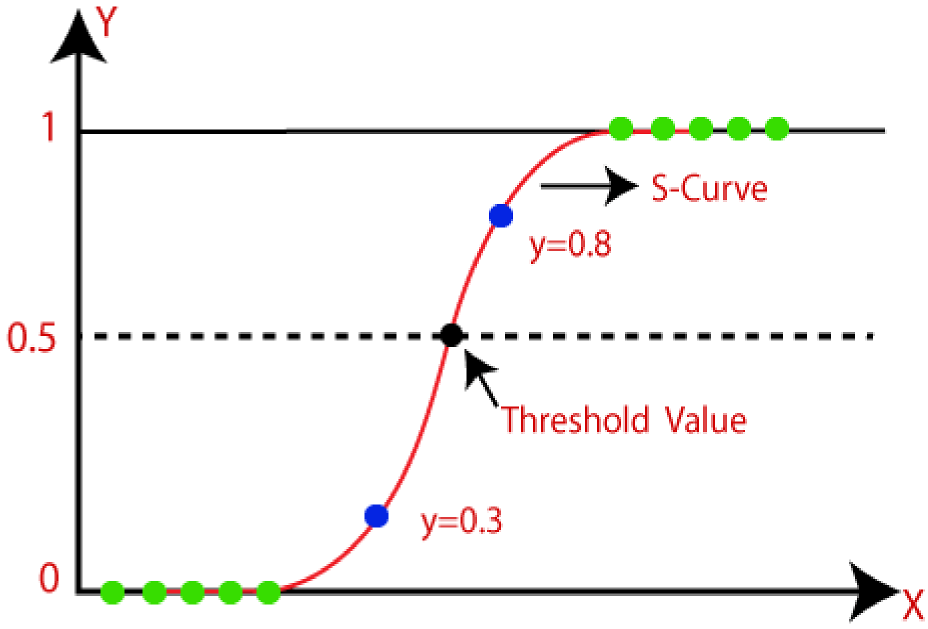

Fig. 1. Logistic regression predicts the output of a categorical dependent variable. It gives the probabilistic values which lie between 0 and 1, 0.5 is the midpoint of Logistic regression, instead of fitting a regression line, it can fit an "S" shaped logistic function, which predicts two minimum values is 0.3 and maximum value is 0.8.

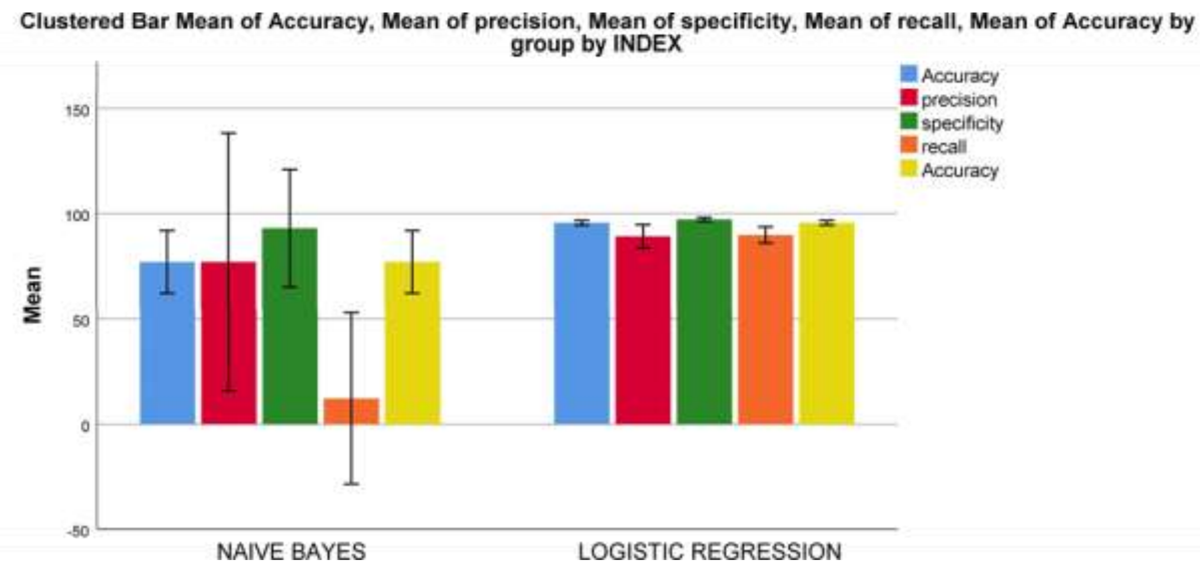

Fig. 2. The y-axis shows the group of naive bayes and logistic regression and the y-axis shows the mean percentage of accuracy, precision, specificity and recall with the Error Bars being 95% CI and the Error Bars being +/-2SD.