International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in

Autism Spectrum Disorder Analysis and Detection Using Machine Learning Techniques

[1]Mavilla Kavya & [2]Danti Srinivasulu

[1]M.Tech Student, Department of CSE, GVIC, Madanapalle

[2]Assistant professor, Department of CSE, GVIC, Madanapalle

Abstract

Autism Spectrum Disorder is a neurological condition that affects a person's ability to connect and communicate with others for the rest of their life. As per the ASD, the issue arises during childhood and persists throughout adolescence and maturity. This is an attempt to investigate the potential for using Multinomial Naïve Bayes, Support Vector Machines, Logistic Regression, K-Nearest Neighbors, Gaussian Naïve Bayes, Random Forest, Linear Discriminant Analysis, and Quadratic Discriminant Analysis for the prediction and analysis of ASD problems in children, adolescents, and adults. This research is motivated by the growing use of machine learning techniques in the research dimensions of medical diagnosis. Three datasets are used to assess the methods. There are 292 occurrences and 21 attributes in the first dataset about ASD screening in youngsters. There are 21 attributes and a total of 704 instances in the second dataset linked to adult individuals used for ASD screening. There are twenty-one attributes and 104 cases in the third dataset about ASD screening in adolescent individuals. With a cross validation score of 93.2% and F-beta score of 91.4% in the adult dataset, 80.3% and F-beta score of 79.7% in the adolescent dataset, and 93.5% and F-beta score of 97.2% in the children dataset, the "LDA" algorithm performed well after applying various machine learning algorithms in the prior work.With a cross validation score of 94.8% and an F-beta score of 98.3% in the adolescent dataset, a cross validation score of 99.3% and an F-beta score of 99% in the adult dataset, and a cross validation score of 99.6% and an F-beta score of 96.4% in the children dataset, the "Random Forest" algorithm performed admirably in the proposed work.

Keyword: - LDA, ML algorithms, Performance Evaluation

1. INTRODUCTION

The issue of autism spectrum disorder (ASD) is becoming more and more prevalent in people of all ages these days. The maintenance of the patient's physical and mental health can be substantially aided by early diagnosis of this neurological condition. The increasing use of machine learning-based models to forecast a range of human illnesses has made it appear feasible to identify these conditions early on using a variety of physiological and health-related factors.One issue that is connected to the development of the human brain is autism spectrum disorder. Individuals suffering from the illness typically find it difficult to communicate and interact with others [1] or [3]. A person with this condition typically experiences lifetime effects, and it is possible that both hereditary and environmental variables will eventually be identified as the disease's primary causes.This

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

issue could manifest symptoms as early as age three and last the rest of a person's life. Although the patient's condition cannot be entirely cured, if the symptoms are identified early on, its consequences can be temporarily lessened. Certain social interactions and communication issues include the following: The healthcare industry is greatly impacted by autism spectrum disorder because of the rise in autism cases as well as the time and expense associated with patient diagnosis. By providing patients with the necessary therapy or medicine and so lowering the long-term costs associated with delayed diagnosis, early detection of ASD can benefit both the patients and the healthcare industry. The project's main goal is to apply several machine learning algorithms to identify autism spectrum disorder risk.

2. IMPLEMENTATION

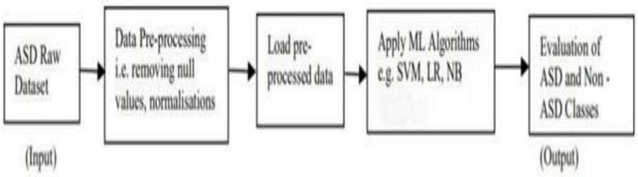

Figure 1 Architecture of Proposed System

- Import the datasets
- Clean the datasets i.e preprocessing
- Apply Machine learning techniques on the data for classification purpose
- Obtain the results by using different metrics
- Conclusion

Dataset

For identifying mental imbalance range jumble three datasets to be specific dataset containing the records of youngsters, dataset containing the records of youths and dataset containing the records of grown-ups were utilized. All the three datasets were gathered from the UCI AI Archive with 21 credits. The properties are of the kind Downright, nonstop and parallel. The dataset has been involved by numerous analysts in their works. In any case, the fundamental issue emerges with the construction of the dataset, that is the missing qualities that are not ascribed with appropriate mathematical qualities. Along these lines, we have Information pre-handling techniques and apparatuses to address this issue.

	id	A1_Score	A2_Score	A3_Score	A4_Score	A5_Score	A6_Score	A7_Score	A8_Score	A9_Score	...	gender	ethnicity	jundice	austim
0	1	1	1	1	1	0	0	1	1	0	...	f	White-European	no	no
1	2	1	1	0	1	0	0	0	1	0	...	m	Latino	no	yes
2	3	1	1	0	1	1	0	1	1	1	...	m	Latino	yes	yes
3	4	1	1	0	1	0	0	1	1	0	...	f	White-European	no	yes
4	5	1	0	0	0	0	0	0	1	0	...	f	NaN	no	no
5	6	1	1	1	1	1	0	1	1	1	...	m	Others	yes	no
6	7	0	1	0	0	0	0	0	1	0	...	f	Black	no	no

Table 1 List of ASD datasets

Data Pre-processing

Information preprocessing is a method where change the information into a significant organization. Certifiable information is ordinarily deficient and conflicting on the grounds that it contains loads of mistakes and invalid qualities. A decent pre-handled information generally yields a decent outcome. Different information pre-handling techniques are utilized to deal with fragmented and conflicting information like taking care of missing qualities, exception recognition, information discretization, information decrease (aspect and numerosity decrease) and so forth. The issue of missing qualities in these datasets has been taken care of by utilizing mode measure. Highlight Choice: Variable selection, which reduces algorithm complexity and avoids overfitting, also speeds up algorithm training. It is performed on the dataset to know the significant sections or elements from the given information. Highlight determination is the method involved with lessening the quantity of info factors while fostering a prescient model. Presently the information is parted into two examples. They are preparing information and test informational indexes. Preparing information and Test information are parted in the proportion 80:20.

3. EXPERIMENTAL RESULTS

Confusion Matrix:

For a given set of test data, the Confusion matrix is a square matrix that is used to assess how well the classification models perform. Only after the actual values for the test data are known can it be ascertained. It is sometimes referred to as the error matrix since it displays the errors in the model's performance as a matrix.

Cross Validation Score:

In applied machine learning, cross validation is mostly utilized to gauge a given machine learning model's proficiency with the available data. That is, to assess the model's projected overall performance using a small sample size when it comes to making predictions on data that was not used for model training.

F-Beta Score:

The weighted harmonic mean of recall and precision, or F-Beta score, has a maximum value of 1 and a minimum value of 0. The weight of recall in the total score is set by the beta parameter.

F-Beta Score = (2*(Precision*Recall))/(Precision+Recall)

Precision:

The precision of a model can be determined by counting the number of right outputs it produces or by determining the proportion of all positively predicted classes that the model correctly identified and found to be true.

Precision = TP/(TP+FP)

Recall:

Recall is the percentage of all positive classes that our model accurately predicted. There needs to be a maximum recall.

Recall = TP/(TP+FN)

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in
IJARST

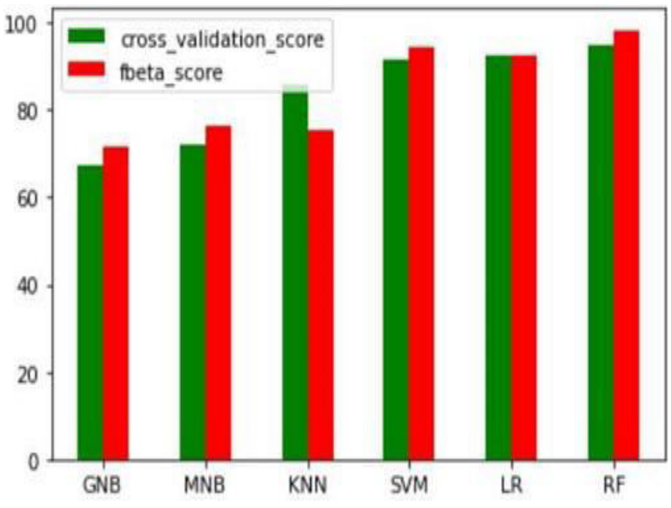

Figure 2 Comparison of algorithms in adolescent dataset

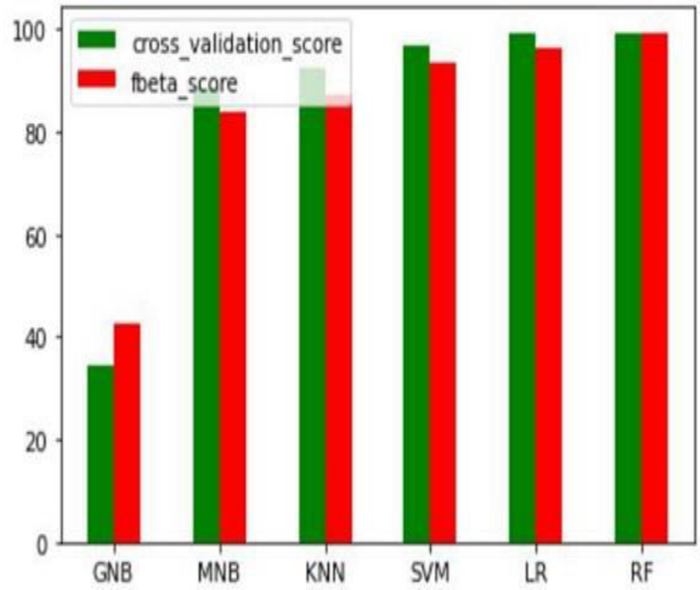

Figure 3 Comparison of algorithms in adult dataset

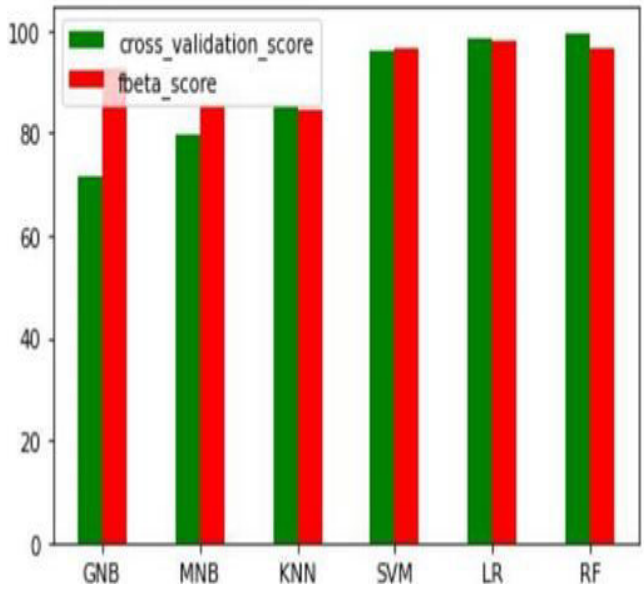

Figure 4 Comparison of algorithms in children dataset

4. CONCLUSION

Early diagnosis and prompt treatment of autism spectrum disorder are critical because the condition can have serious health consequences. The major objective is to compare several machine learning methods, such as Random Forest, Support Vector Machines, Multinomial Naïve Bayes, Logistic Regression, K-Nearest Neighbors, and Gaussian Naïve Bayes, in order to predict the probability of autism spectrum disorder at an early stage. As a result, the algorithms were compared by examining the F-beta and cross-validation scores. With a cross validation score of 94.8% and F-beta score of 98.3% in the adolescent dataset, 99.3% and F-beta score of 99% in the adult dataset, and 99.6% and F-beta score of 96.4% in the children dataset, it was discovered that the "Random Forest" algorithm outperformed all other algorithms in all three datasets.

REFERENCE

1. Thabtah, Fadi. "Machine learning in autistic spectrum disorder behavioural research: A review and ways forward. (2018) " Informatics for Health andSocial Care : 1-20.
2. Thabtah, Fadi, Firuz Kamalov, and Khairan Rajab. (2018) "A new computational intelligence approach to detect autistic features for autism screening." International journal of medical informatics 117: 112-124.
3. Vaishali, R., and R. Sasikala. "A machine learning based approach to classify Autism with optimum behaviour sets. (2018) " International Journal of Engineering & Technology 7(4): 18.
4. Constantino, John N., Patricia D. Lavesser, Y. I. Zhang, Anna M. Abbacchi, Teddi Gray, and Richard D. Todd. (2007) "Rapid quantitative assessment of autistic social impairment by classroom teachers." Journal of the American Academy of Child & Adolescent Psychiatry 46(12): 1668-1676.
5. Daniel Bone, Matthew S. Goodwin, Matthew P. Black, Chi-Chun Lee, Kartik Audhkhasi, and Shrikanth Narayanan. (2015) "Applying machine learning to facilitate autism diagnostics: pitfalls and promises." Journal of autism and developmental disorders 45(5): 1121-1136.
6. Dennis Paul Wall, J. Kosmicki, T. F. Deluca, E. Harstad, and Vincent Alfred Fusaro. (2012) “Use of machine learning to shorten observationbased screening and diagnosis of autism.” Translational psychiatry, 2(4): e100.
7. Dennis P. Wall, Rebecca Dally, Rhiannon Luyster, Jae-Yoon Jung, and Todd F. DeLuca. (2012) “Use of artificial intelligence to shorten the behavioral diagnosis of autism.” PloS one, 7(8): e43855.
8. Fadi Thabtah. (2017). “Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment.” In Proceedings of the 1st International Conference on Medical and Health Informatics, pp. 1-6. ACM.