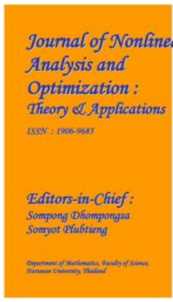
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1 : 2024
ISSN :**1906-9685**



A Comprehensive Study of Machine Learning Approaches for ASD Prediction

Lokesh Sirela
AssistantProfessor
Raghu Engineering College(A),Dakamarri, Visakhapatnam,India
lokesh.sirela@raghuenggcollege.in

Venkata Sai Sampath Kumar Gudivada
Student
Raghu Engineering College(A)Dakamarri, Visakhapatnam, India
sampathgudivada21@gmail.com

Veluturi Sai Prashanth
Student
Raghu Engineering College(A)Dakamarri, Visakhapatnam,India
saiprashanth506@gmail.com

Vanapalli Mounika Varalakshmi
Student
Raghu Engineering College(A)Dakamarri, Visakhapatnam,Indiamounikavanapalli22@gmail.com

Pudi Omkar
Student
Raghu Engineering College(A)Dakamarri, Visakhapatnam,Indiaomkarpudi88@gmail.com

DepartmentofComputer Science(Data Science)

***Abstract*-**Autism Spectrum Disorder (ASD) is a complex neurological condition that affects individuals of all ages, impacting their mental, social, and physical well-being. Traditional diagnostic methods for ASD are time-consuming and expensive. Machine learning techniques offer a promising approach to automate and improve the efficiency of ASD diagnosis. This study aims to identify key characteristics for automating the diagnostic process and evaluates several machine learning algorithms, including Logistic Regression, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Naïve Bayes, for predicting ASD occurrence. Experimental results show that the Naïve Bayes algorithm achieves the highest accuracy, reaching 99.6% compared to other algorithms.

Keywords: Machine Learning, Autism Spectrum Disorder, KNN, Logistic Regression, Naïve Bayes, SVM.

Introduction

The human brain, serving as the central organ in the body, coordinates various functions. Autism arises from neural disconnection and disturbances in brain maturation. Although it can appear at any age, autism typically emerges in childhood as a developmental condition. People with Autism Spectrum Disorder (ASD) face challenges in social interaction and communication, often showing restricted interests and repetitive behaviors that affect their daily lives. According to the WHO, autism affects around one in every hundred and sixty children. Some individuals with autism can live independently, while others may need lifelong care and support..

Detecting autism can be time-consuming, but early detection offers significant benefits. It enables timely intervention and appropriatetreatment, preventing further health deterioration and reducing the need for unnecessary treatments, thus lowering costs

This study aims to use classification methods to determine if an individual has Autism Spectrum Disorder (ASD). It also aims to identify the optimal classification algorithm for predicting ASD, using performance criteria such as errorrate and accuracy. The dataset on autism is preprocessed, converting

textual data into numerical format using techniques like One Hot Encoder and Label Encoder. Data purification is then performed using the mean method. Features in the dataset are identified using training data, while testing data is used for result validation. Significant attributes are selected through feature selection techniques.

The efficiency of the decision-making algorithm is evaluated based on criteria such as correctness and inaccuracy frequency. The classifier demonstrating superior effectiveness is considered the most suitable for ASD prediction. After implementing various classification models for autism prediction, it is observed that Naïve Bayes consistently outperforms other algorithms, providing significantly better results.

The paper continues with the following structure: Section II provides a concise overview of existing research on autism spectrum disorder. Section III outlines the proposed methodology. Section IV presents the experimental results and corresponding analysis. Section V concludes the study and discusses potential future research directions.

I. LITERATUREREVIEW

Within the research by Sunsirikul et al., they detail theirendeavorstocreateadataanalyticsinstrumentaimedataiding future medical practitioners in diagnostic procedures.Theirinvestigationentailedextractingbehavioral trendsfromdataandestablishingabehavioralguidelineforpatients,withtheobjectiveofidentifyingpotentialassociations between specific behaviors and symptoms ofautism, relying on an ample amount of patient records. Thepaper delves into various data mining techniques, aiming tofurnish physicians with a comprehensive set of tools for theintelligentevaluationofpatientdata.OnenotableoutcomeofthisstudyisitsrevelationoftheassociationbetweenbehavioralpatternsinautisticchildrenandPDDNOS,potentiallyleading toimprovementsinself-esteem,mitigating disabilities, and refining disorder classifications.Moreover, the research highlights the absence of medicalinformationregardingtypicalchildrenthroughoutthe training period.

Osman et al. developed a methodology that utilizes data analysis methods to diagnose Autism Spectrum Disorder (ASD) in patients. ASD significantly impacts an individual's well-being, characterized by social and interpersonal challenges, repetitive actions, and intense interests or hobbies. This study focuses on using classification algorithms to identify ASD in children. The classification process results determine whether a child is diagnosed with ASD or not. The LDA algorithm achieved an accuracy of 90.8%, while the KNN algorithm reached 88.5% accuracy.

Cincy Raju et al. highlighted heart disease as one of the deadliest illnesses leading to mortality, causing profound long-term damage. This study aims to use data analysis techniques to provide an effective solution for therapeutic scenarios. Various classification algorithms were used, with the Support Vector Machine (SVM) proving to be the most superior among them.

Erik et al. introduced a model for ASD screening that incorporates machine learning adaptation and DSM-5 criteria. Their framework aims to predict autism and introduce a comprehensive healthcare system for gathering, assessing, and managing data related to the evaluation and therapy of children with ASD. The Autism Management Platform (AMP) serves as an intelligent web interface and statistical platform, facilitating real-time collection and extraction of patient data by healthcare professionals and specialists. It also provides automated feedback to adjust data filtering preferences. Similar efforts in predictive analysis have been undertaken

.

Canon et al. have provided empirical evidence regarding the prognosis of autism spectrum disorder. This manuscript identifies and presents supplementary evidence needed for further investigation. However, it underscores the importance of examining the functional impacts on individuals.

Karunakaran et al. have introduced a method that integrates an adaptive functioning classifier with early learning techniques. This method effectively addresses the challenge of handling less noisy data, a notable limitation. Additionally, alternative approaches for predicting autism spectrum disorder include machine learning analysis and pathway analysis.

The current methods for predicting autism spectrum disorder lack promise, primarily due to the omission of critical parameters in the analysis. Therefore, it is imperative to incorporate all essential parameters to enhance the effectiveness of the proposed algorithm.

II. PROPOSEDMETHODODLOGY

This section outlines the proposed scheme for analyzingautismdisorder.Theprocessinvolvesthefollowing steps:

(a) Data Gathering

(b) DataPreprocessing

(c) ModelDevelopment

(d) TrainingandValidation

Figure1illustratestheflowchartoutliningthemethodtobepursued for thecompleteprocess.

(a) Data Gathering We collected a dataset comprising 1054 unique cases for predicting autism spectrum disorder. Essential attributes from this dataset were used to train our model. The dataset includes various fields such as indications, age groups, familial backgrounds, living locations, and more.

(b) Data Preprocessing Data often contains missing components, discrepancies, inaccuracies, and non-numeric entries. Data preprocessing offers a robust solution to address these challenges. In this study, we preprocess the data using One Hot Encoding and Label Encoding methodologies. This strategy effectively converts qualitative data into numerical formats.

In the Autism Assessment dataset, data preprocessing involves data cleaning techniques followed by feature selection to filter out irrelevant or redundant features. Subsequently, classification is performed using algorithms such as LDM, SVM, and Naive Bayes. This is followed by forecasting outcomes and assessing for superior precision.

(c) Model Development Various methodologies are employed in constructing the model, aiding in the selection of the most suitable model for achieving the desired outcome. The algorithms used include LDM, SVM, and Naive Bayes.

(a) .

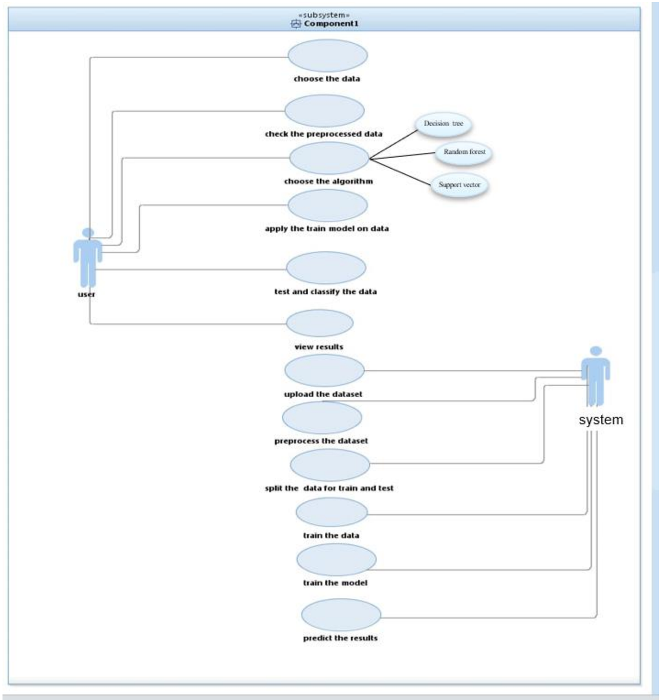


Figure-1ProposedProcessFlow

(i) LDM

Linear Discriminant Analysis (LDA) is a statistical method used for predicting qualitative outcome measures using predictor variables. LDA aims to maximize the separability between classes while minimizing the variance within each class. This makes LDA a powerful tool for pattern recognition and classification tasks.

(ii) SVM

When dealing with unlabeled data, supervised learning methods become impractical, leading to the necessity of using unsupervised techniques. In unsupervised learning, data is naturally grouped into clusters, and new data points are then mapped to these clusters. Support Vector Machines (SVM)

The effectiveness of Support Vector Machines (SVM) depends on the choice of kernel, its configurations, and the soft margin setting. The Gaussian kernel, which has just one parameter, is a commonly preferred option. However, SVM has some limitations:

- Complete labeling of data inputs is required.
- SVM's calibration of probabilities for classes may be inaccurate, as it is based on Vapnik's theory, which eliminates the need for probability estimation from finite data.
- SVM is primarily suitable for two-class tasks, so techniques are needed to transform multi-class tasks into a series of binary problems.
- The interpretation of parameters in a solved model can be challenging.

.

(iii) NaïveBayes

The Naive Bayes algorithm is a well-known method for building predictive models. It uses an "event-based model" based on hypotheses about feature distributions, often employing Multinomial and Bernoulli distributions for categorical attributes encountered in document sorting. This leads to the development of two distinct yet often misinterpreted models.

(iv) Training and Validation: After selecting different models, the dataset undergoes training using various construction models. Subsequently, the data is tested across all models, and the one demonstrating superior accuracy is chosen for further use.

.

III. TESTEDRESULTSANDANALYSIS

Inthispart,testoutcomesareshowcasedanddeliberated. The format of the dataset is visually illustrated,and the efficiency of the algorithms in forecasting autismdisorder ismeticulouslyexamined.

A. DataOverview

The data pool utilized for identifying autism spectrumconditioncomprises1054observationsandincludes19 distinctattributessuchassymptoms,gender,age,ethnicity,nationality,etc.Figure2illustratesthefeaturespresentindatapool.

```
Index(['Case_No', 'A1', 'A2', 'A3', 'A4', 'A
5', 'A6', 'A7', 'A8', 'A9', 'A10',
       'Age_Mons', 'Qchat-10-Score', 'Sex', 'E
thnicity', 'Jaundice',
       'Family_mem_with_ASD', 'Who completed t
he test', 'Class/ASD Traits '],
      dtype='object')
```

Figure-2Attributes

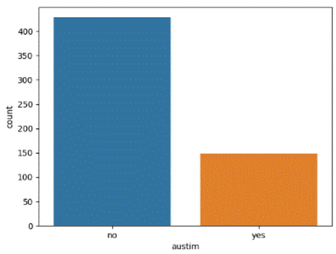


Figure-3CountofTraits

Thedatasetundergoespreprocessing,andappropriateattributes are chosen for establishing the framework. Prior toframework development and construction, the data's patternsare comprehended through visualization. Figure 3 displaysthe distribution of traits among individuals with and withoutautism,indicatingthecountofaffectedandunaffectedindividualsacrossdifferentcharacteristics.

This following figure illustrates the association amongthe features withindata pool.

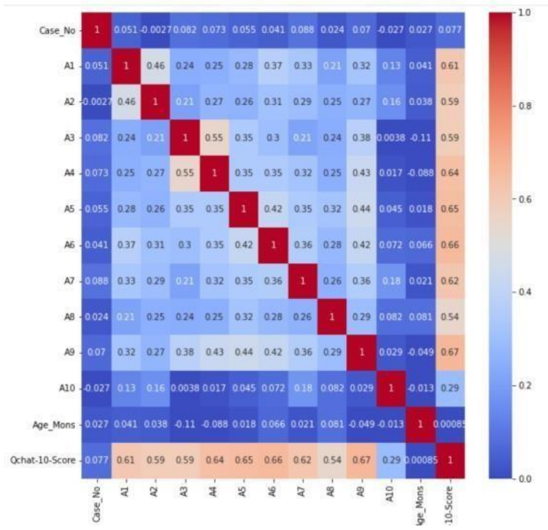


Fig.4HeatMap

B. ComparativeAnalysis

The proposed method for predicting autism is evaluated usingmultiple measurements such as precision and inaccuracy. Figure V illustrates the inaccuracy achieved with Naïve Bayes, while Figure VI depicts the inaccuracy with SVM. Additionally, Figure VII presents an analysis of accuracy among the constructed frameworks, which include Support Vector Machine, K-Nearest Neighbor, Logistic Regression, and Naïve Bayes. The figure clearly indicates that Naïve Bayes surpasses other models in forecasting autism, demonstrating superior accuracy

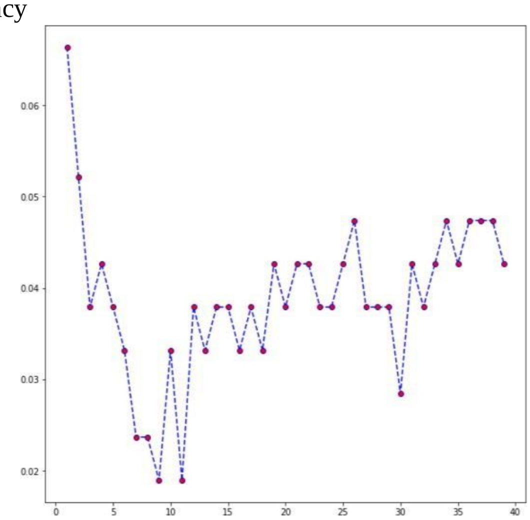


Figure-5ErrorrateinNaïveBayes

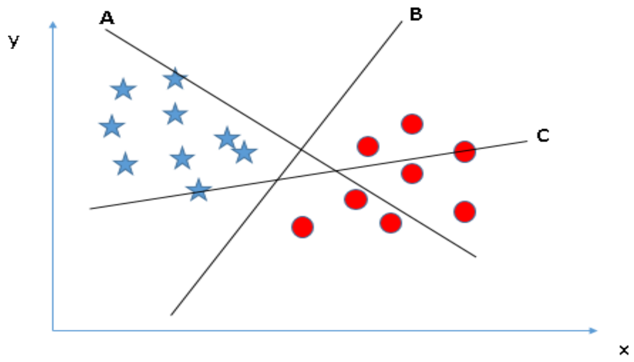


Figure-6ErrorRateinSVM

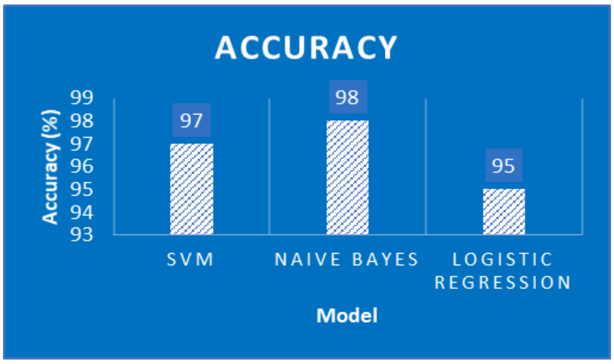


Figure-7Accuracy

Table1presentsanoverviewofthecontrastbetweenexistingand suggestedmethodologies.

Table1:ComparativeAnalysis

Approach	**Accuracy(%)**	**Error Rate(%)**
K-NearestNeighbour	98.2	0.06
LogisticRegression	97.2	0.07
SupportVector Machine	98.4	0.05
NaïveBayes	99.6	0.02

IV. CONCLUSION

This manuscript conducts a comparative analysis of various algorithms, including Support Vector Machine, Naive Bayes, and logistic regression, to predict the onset of autism in adults The experimental results regarding autism prediction indicate that Naive Bayes outperforms the other models. There potential to improve the system in the future to increase accuracy and reduce errors.

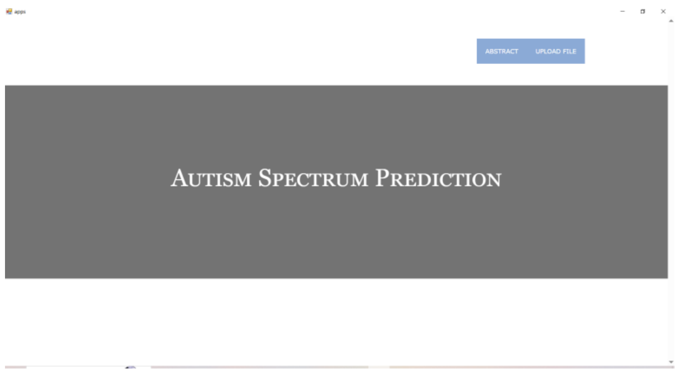


Fig-Home page in project

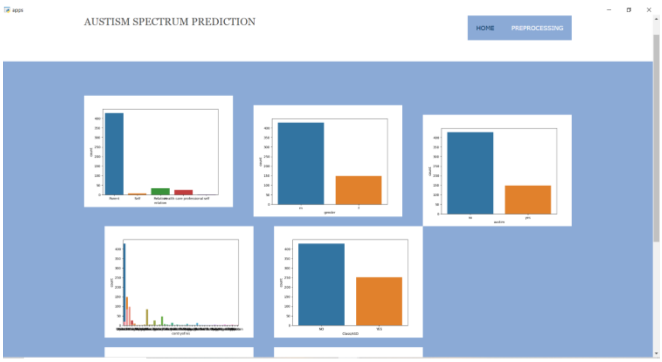


Fig-Analysis page

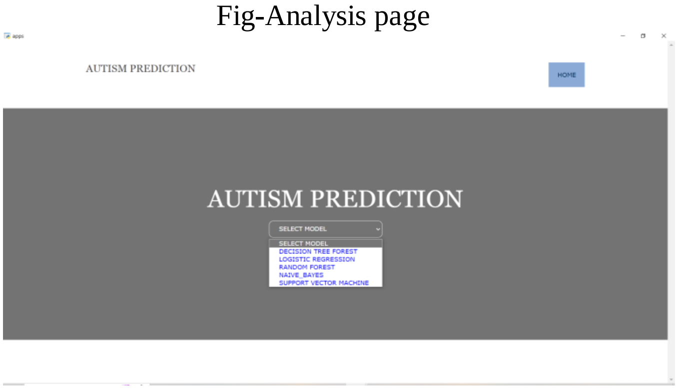


Fig-Model selection Page

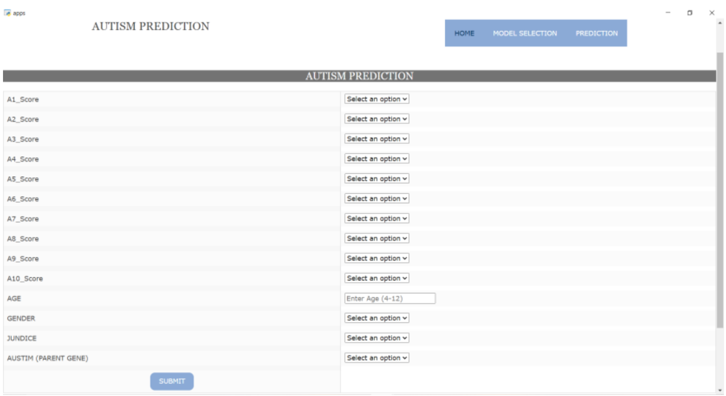


Fig-Prediction Page

REFERENCES

[1] Karunakaran,P.,&Hamdan,Y.B."Earlypredictionofautismspectrum disorder by computational approaches to fmri analysis withearlylearningtechnique."JournalofArtificialIntelligence, 2(04),207-216,2020.

[2] Cannon, J., O'Brien, A. M., Bungert, L., & Sinha, P. "Prediction inautism spectrum disorder: a systematic review of empirical evidence."AutismResearch,14(4),604-630,2016.

[3] Madhukeerthana, Y. BevishJinila, “A review on rough set theory inmedical images”, Research Journal of Pharmaceutical, Biological andChemical Sciences, Vol.7, Issue 1, pp. 815-822, ISSN : 0975-8585,2016.

[4] Lord,C.,Elsabbagh,M.,Baird,G.,&Veenstra-Vanderweele,J.(2018). Autism spectrum disorder. The Lancet, 392(10146), 508-520.

[5] Chaitra, N., Vijaya, P. A., & Deshpande, G. "Diagnostic prediction ofautismspectrumdisorderusingcomplexnetworkmeasuresinamachine learningframework."BiomedicalSignalProcessingandControl,62,102 099,2020.

[6] Kojovic, N., Natraj, S., Mohanty, S. P., Maillart, T., & Schaer, M."Using 2D video-based pose estimation for automated prediction ofautismspectrumdisordersinyoungchildren."ScientificReports,11(1), 1-10,2021.