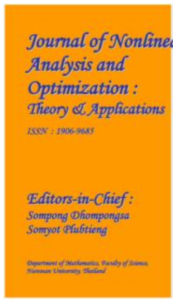
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1 : 2024
ISSN : **1906-9685**



COMPARATIVE ANALYSIS OF MACHINE LEARNING TECHNIQUES FOR PREDICTING AUTISM SPECTRUM DISORDER(ASD)

Mrs. P. PAVITHRA, Assistant Professor in Department of Information Technology,Vignan's Institute of Information technology(A), Visakhapatnam.
G. Uday Kumar, Student, Department of Information Technology, Vignan's Institute of Information technology(A), Visakhapatnam.
CH. Shyam Sampath, Student, Department of Information Technology, Vignan's Institute of Information technology(A), Visakhapatnam.
N. Noshith Chowdary, Student, Department of Information Technology, Vignan's Institute of Information technology(A), Visakhapatnam.
S. Ganesh Kumar, Student, Department of Information Technology, Vignan's Institute of Information technology(A), Visakhapatnam.
G. Lakshmi Sahithi, Student, Department of Information Technology, Vignan's Institute of Information technology(A), Visakhapatnam.

ABSTRACT:
The Autism Spectrum Disorder (ASD) is a neurological condition impacting an individual's mental, social, and physical well-being, affecting individuals of any age group. Detecting this disorder poses significant challenges due to its complexity and the time-consuming and costly nature of traditional diagnostic methods. Conversely, machine learning techniques offer a promising avenue for more efficient and intelligent diagnostics. This study focuses on identifying specific characteristics to automate the diagnostic process and conducts a relative evaluation of various machine learning algorithms, including Logistic Regression, K-Nearest Neighbour, SVM, and Naïve Bayes, for predicting ASD occurrence. Experimental analysis indicate that the Naïve Bayes algorithm achieves superior accuracy, reaching 99.6% compared to alternative algorithms.

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

1. INTRODUCTION

The brain serves as the primary organ in the human body, responsible for coordinating various bodily functions. Autism stems from neural disconnection and disturbances in brain maturation. While autism can manifest at any age, it typically begins in childhood and is recognized as a developmental condition. Individuals with Autism Spectrum Disorder (ASD) encounter challenges in social interaction and communication, often exhibiting restricted interests and repetitive behaviors, impacting their daily lives. According to the WHO, autism affects approximately one in every hundred and sixty children. While some individuals with autism can live independently, others may require ongoing care and support throughout their lives.

Detecting autism is often a time-consuming endeavor. However, early detection offers significant benefits by facilitating timely intervention and appropriate treatment for patients. Early prediction of the condition prevents further deterioration of health and minimizes the need for unnecessary treatments, ultimately reducing costs.

The objective of this study is to determine if an individual has ASD through the application of classification methods. Furthermore, the study seeks to pinpoint the optimal classification algorithm for forecasting ASD, relying on performance criteria such as error rate and accuracy. The dataset on

Autism undergoes preprocessing procedures, involving the transformation of textual data into numerical format utilizing One Hot Encoder and Label Encoder, followed by data purification using the mean method. Features within the dataset are identified through utilization of training data, while testing data is employed for result validation. Following this, significant attributes are chosen through feature selection techniques.
Efficiency of Decision-making algorithm is determined utilizing criteria such as correctness and inaccuracy frequency. Ultimately, the classifier demonstrating superior effectiveness is considered the most suitable for ASD prediction. Upon implementing various classification models for autism prediction, it is evident that Naïve Bayes consistently outperforms other algorithms, yielding significantly better results.
The paper continues with the following structure: Section II provides a concise overview of existing research on autism disorder. Section III outlines the proposed methodology. Section IV presents the experimental results and corresponding analysis. Section V concludes the proposed study and explores potential avenues for future research.

2. REVIEW OF LITERATURE

This part outlines relevant studies concerning autism disorder, focusing on approaches utilizing machine learning techniques.
Within the research by Sunsirikul et al., they detail their endeavors to create a data analytics instrument aimed at aiding future medical practitioners in diagnostic procedures. Their investigation entailed extracting behavioral trends from data and establishing a behavioral guideline for patients, with the objective of identifying potential associations between specific behaviors and symptoms of autism, relying on an ample amount of patient records. The paper delves into various data mining techniques, aiming to furnish physicians with a comprehensive set of tools for the intelligent evaluation of patient data. One notable outcome of this study is its revelation of the association between behavioral patterns in autistic children and PDD NOS, potentially leading to improvements in self-esteem, mitigating disabilities, and refining disorder classifications. Moreover, the research highlights the absence of medical information regarding typical children throughout the training period.
Osman et al introduced. a methodology employing 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, recurring actions, and hobbies or profound interests. This study focuses on employing classification algorithms to identify ASD in children. The results of the classification process determine whether a child is identified as having ASD or not. The accuracy of the LDA algorithm reached 90.8%, whereas the KNN algorithm attained 88.5% accuracy.
Cincy Raju et al. proposed that heart disease is among the deadliest illnesses that can result in mortality. It endures a profound long-term detriment. This ailment strikes with significant intensity. The aim of this study is to utilize data analysis techniques to offer an efficient solution for therapeutic scenarios. Various classification algorithms are utilized. Support Vector Machine (SVM) stands out as the most superior among these methods.
In a model for ASD screening, incorporating machine learning adaptation and DSM-5 criteria, is utilized. Erik et al. introduced a framework for predicting the autism. The aim of this research is to introduce a comprehensive healthcare system designed to gather, assess, and oversee data concerning the evaluation and therapy of child with ASD.AMP, functioning as an intelligent web interface and statistical platform, facilitates the real-time collection and extraction of patient data by healthcare professionals and specialists. Furthermore, it offers automated feedback to adjust data filtering preferences. Comparable endeavors in predictive analysis have also been undertaken.
Canon et al. have provided empirical evidence regarding the prognosis of autism 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 disorder include machine learning analysis and pathway analysis

The current methods for predicting autism 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.

3. PROPOSED METHODODLOGY

This section outlines the proposed scheme for analyzing autism disorder. The process involves the following steps:

(a) Data Gathering
(b) Data Preprocessing
(c) Model Development
(d) Training and Validation

Figure 1 illustrates the flowchart outlining the method to be pursued for the complete process.

(a) Data Gathering

We've selected a dataset comprising 1054 unique cases for predicting autism spectrum disorder. Essential attributes from this dataset are utilized to train our model. The assorted fields in our dataset embody a range of indications, age groups, familial backgrounds, living locations, and more.

(b) Data Preprocessing

Data often arrives with missing components, discrepancies, inaccuracies, and non-numeric entries. Data preprocessing offers a robust solution to address these challenges. In this investigation, we preprocess the data employing One Hot Encoding and Label Encoding methodologies. This strategy adeptly converts qualitative data into numerical formats.

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

(c) Model Development

Various methodologies are utilized in constructing the model, aiding in the selection of the most suitable model for achieving the desired outcome. The algorithms utilized include LDM, SVM and Naïve Bayes.

(i) LDM

LDM aids in the prediction of a qualitative outcome measures using predictor variables. By maximizing the separability between classes while minimizing the variance within each class, LDA provides a powerful tool for pattern recognition and classification tasks.

(ii) SVM

Supervised learning becomes impractical for handling unlabeled data, necessitating the use of unsupervised techniques. In this approach, data is naturally grouped, and new data is subsequently mapped to these clusters. SVM assists in categorizing unlabeled data by employing statistics derived from the support vector.

The efficacy of SVM relies on the choice of kernel, its configurations, and the soft margin setting. A commonly preferred option is the Gaussian kernel, which entails just one parameter. However, limitations of SVM encompass the subsequent facets:

- Complete labeling of data inputs is essential.
- SVM's calibration of probabilities for classes may be inaccurate as it's grounded in Vapnik's theory, eliminating the necessity for probability estimation from finite data.
- SVM is primarily suitable for two-class tasks, necessitating techniques to transform multi-class tasks into a series of binary problems.
- The interpretation of parameters in a solved model can be challenging.

(iii) Naïve Bayes

The Naïve Bayes algorithm, a renowned method for constructing predictive models, employs an "event-based model" founded on hypotheses regarding feature distributions, frequently utilizing Multinomial and Bernoulli distributions for categorical attributes encountered in document sorting, resulting in the development of two distinct yet often misinterpreted models.

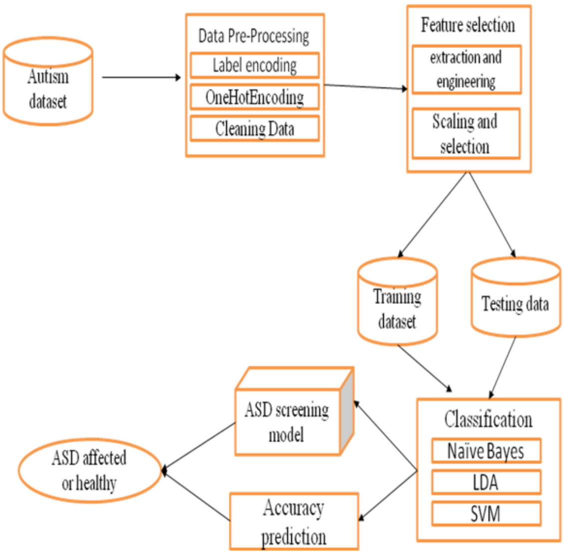


Figure-1Proposed Process Flow

(d)Training and Validation

After opting for diverse models, the dataset undergoes training using various construction models. Afterward, the data is tested across all models, and the one exhibiting superior accuracy is selected for further utilization.

4. TESTED RESULTS AND ANALYSIS

In this part, test outcomes are showcased and deliberated. The format of the dataset is visually illustrated, and the efficiency of the algorithms in forecasting autism disorder is meticulously examined.

A. Data Overview

The data pool utilized for identifying autism spectrum condition comprises 1054 observations and includes 19 distinct attributes such as symptoms, gender, age, ethnicity, nationality, etc. Figure 2 illustrates the features present in data pool.

```
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

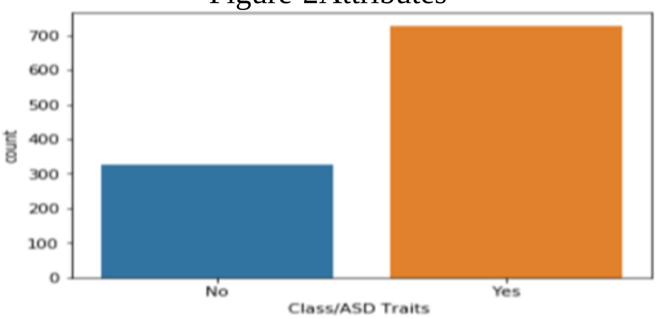


Figure-3Count of Traits

The dataset undergoes preprocessing, and appropriate attributes are chosen for establishing the framework. Prior to framework development and construction, the data's patterns are comprehended through visualization. Figure 3 displays the distribution of traits among individuals with and without autism, indicating the count of affected and unaffected individuals across different characteristics. This following figure illustrates the association among the features within data pool.

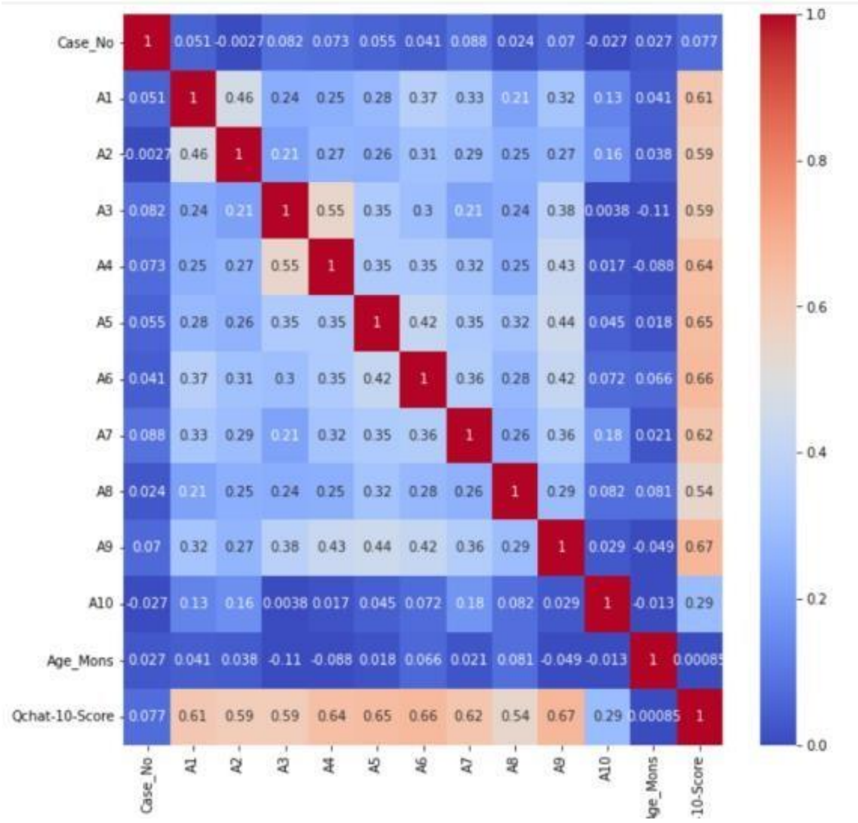


Figure-4HeatMap

B. Comparative Analysis

The proposed method for predicting autism is evaluated using multiple 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 are Support Vector Machine, K-Nearest Neighbour, Logistic Regression and Naïve Bayes. The figure clearly indicates that Naïve Bayes surpasses other models in forecasting autism, demonstrating superior accuracy.

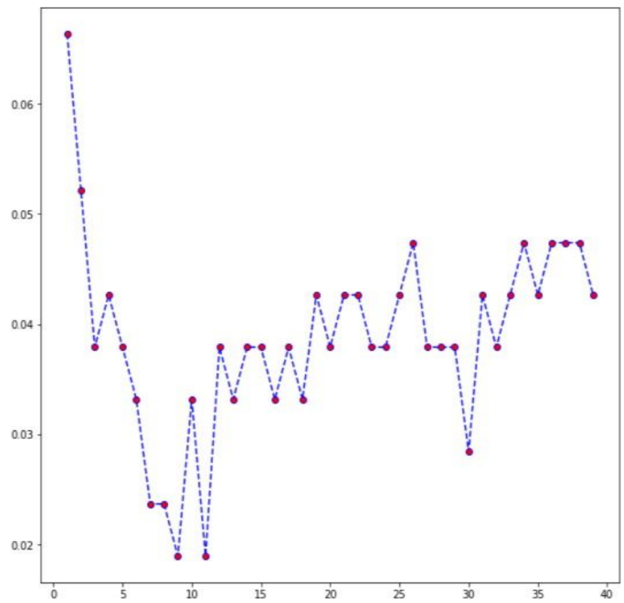
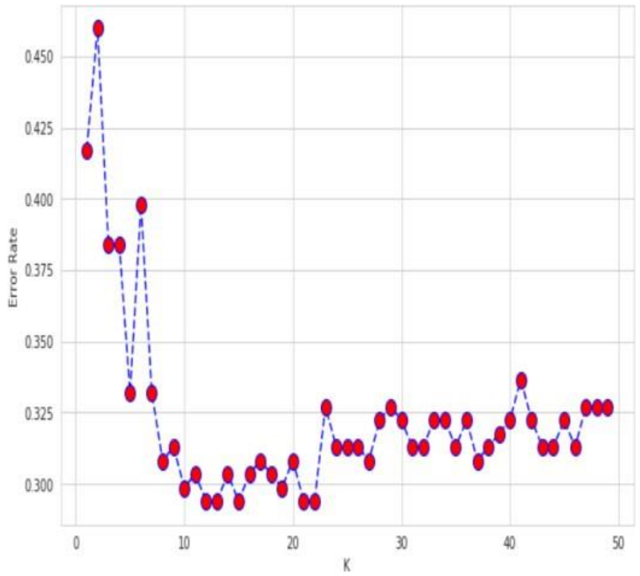


Figure-5Errorrate in Naïve Bayes

Figure-6Error Rate in SVM



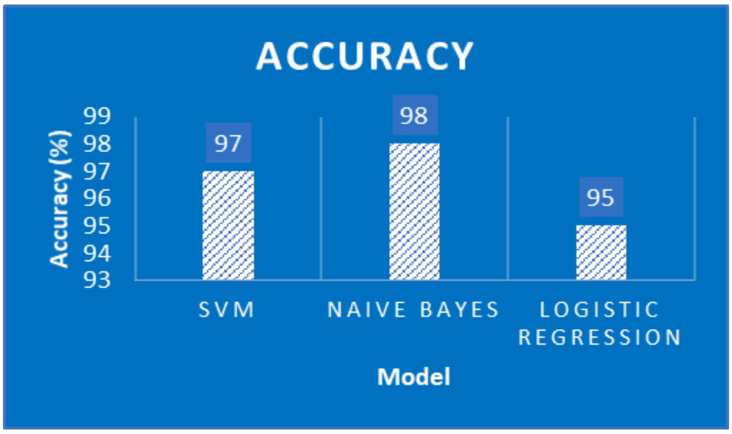


Figure-7Accuracy

Table 1 presents an overview of the contrast between existing and suggested methodologies.

Table1: Comparative Analysis

Approach	**Accuracy (%)**	**Error Rate(%)**
K-Nearest Neighbour	98.2	0.06
LogisticRegression	97.2	0.07
Support Vector Machine	98.4	0.05
NaïveBayes	99.6	0.02

5. CONCLUSION

In this manuscript, a meticulous comparative analysis of various algorithms, namely Support Vector Machine, Naive Bayes, and logistic regression, is conducted to predict the onset of autism in adults. By rigorously examining the experimental outcomes related to autism disorder prediction, it is evident that Naive Bayes demonstrates superior performance compared to its counterparts. This finding underscores the significance of employing sophisticated machine learning techniques in addressing complex medical conditions such as autism. Moreover, the study suggests that further enhancements to the system could be made in the future to enhance accuracy levels and minimize prediction errors, thereby highlighting the potential for ongoing advancements in predictive modeling for autism diagnosis and intervention.

The results presented in this manuscript offer valuable insights into the efficacy of different algorithms for forecasting autism in adults, with Naive Bayes emerging as the most promising approach among the tested models. The implications of these findings extend beyond the realm of autism prediction, potentially informing the development of more accurate and reliable diagnostic tools for various medical conditions. Additionally, the study emphasizes the need for continuous refinement and optimization of predictive models to accommodate evolving datasets and improve overall performance. By leveraging the advancements in machine learning technology, researchers and practitioners can strive towards more effective strategies for early detection and intervention in autism spectrum disorder, ultimately contributing to better outcomes and quality of life for affected individuals.

REFERENCES:

1. Karunakaran, P., & Hamdan, Y. B. "Early prediction of autism spectrum disorder by computational approaches to fmri analysis with early learning technique." Journal of Artificial Intelligence, 2(04), 207-216, 2020.
2. Cannon, J., O'Brien, A. M., Bungert, L., & Sinha, P. "Prediction in autism spectrum disorder: a systematic review of empirical evidence." Autism Research, 14(4), 604-630, 2016.
3. Madhu keerthana, Y. BevishJinila, "A review on rough set theory in medical images", Research Journal of Pharmaceutical, Biological and Chemical 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 of autism spectrum disorder using complex network measures in a machine learning framework." Biomedical Signal Processing and Control, 62, 102099, 2020.
6. Kojovic, N., Natraj, S., Mohanty, S. P., Maillart, T., & Schaer, M. "Using 2D video-based pose estimation for automated prediction of autism spectrum disorders in young children." Scientific Reports, 11(1), 1-10, 2021.