International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 08 | Aug 2021 www.irjet.net p-ISSN: 2395-0072

Autism Spectrum Disorder Detection Using Machine Learning Techniques

Amrutha S M [1], K R Sumana[2]

[1]PG Student, Department of MCA, The National Institute of Engineering, Mysuru, Karnataka, India
[2]Assistant Professor, The National Institute of Engineering, Mysuru, Karnataka, India

---

Abstract - *Autism spectrum disorder is a neurological developmental disorder that affects person's interaction, communication and learning skills. The effects of Autism Spectrum Disorder and the severity of symptoms are different in each person, mainly this disorder is recognized between the age of 1-5 years and above and the symptoms they have are unusual behaviors, interests and also they might have social challenges. If it was not resolved at starting stage it becomes severe in upcoming days, so in order to reduce the number of cases increasing due to autism. Autism can be predicted at quite early stage using different machine learning techniques. In our proposed work, we are going to predict outcomes of autism diagnosed in children between age group of 1-5 years and above, in addition of assess and implementation of various models of machine learning. We have used a step-by-step approach to analyse the data of past decade. The predicted data of patient with autism and non-autism will be observed as new data and will be used for observing results for forthcoming patients. We are using different Machine Learning Algorithms such as Logistic Regression(LR) Algorithm, Naïve Bayes(NB) Algorithm, Decision Tree(DT) Algorithm, K-Nearest Neighbour(KNN) Algorithms to predict the Autism Spectrum Disorder. Once the former has been implemented, we will extend our project to showcase some advanced features so that accuracy will be maintained.*

Key Words: Naïve Bayes(NB), K-Nearest Neighbour(KNN) Decision Tree(DT), Logistic Regression(LR), Natural Language Processing(NLP).

1.INTRODUCTION

Autism Spectrum Disorder is a neurodevelopmental disorder. These disorders are causes hyperactivity, social communication deficits and learning and language deficiencies. Asperger syndrome can also be caused by hereditary (genetic), chemicals found in food also and among others. With a ratio of 1:68, the number of children with ASD is steadily increasing. ASD can be detected by a diagnosis, which may take 2 or 3 days depending on the severity of the ASD. Multiple streams of doctors observe a child's behavior to make a diagnosis. ASD can be detected manually at the age of 3 or older. And it can be seen at young age also. Using machine learning, ASD can be detected more quickly. When past data is used to predict autism, machine learning takes less time. Decision Tree, Naive Bayes, KNN, and Logistic Regression algorithms are used to detect Autism Spectrum Disorder, and the datasets are collected from online kaggle.com. In order to obtain accurate results that can be used to predict autism spectrum disorder (ASD) at an early stage, these data are analyzed. Using machine learning techniques to detect autism early will help so many people to lead a better life by improving their quality of life.

2 Literature Review

Analysis performed over these accurate results which will be useful to make right decisions in predicting autism spectrum disorder (ASD) at early stages. Thus, early autism intervention using machine learning techniques opens up a new way for autistic individuals to develop the potential to lead a better life by improving their behavioral and emotional skills [1]. In this paper, we present a survey of various supervised classification techniques. The goal of this survey is to provide an inclusive review of different supervised classification techniques such as decision tree, Support Vector Machine, Naive Bayes, K-Nearest Neighbor, Neural Network. Here we surveyed various supervised classification techniques. We have taken review from the different classification techniques. And we implemented Naïve Bayes and K-Nearest Neighbor algorithms in our project [2]. The key idea is to represent eye-tracking records as textual strings describing the sequences of fixations and saccades. As such, the study could borrow Natural Language Processing (NLP) methods for transforming the raw eye-tracking data. The NLP-based transformation could yield interesting features for training classification models. The experimental results demonstrated that such representation could be beneficial in this regard. With standard Convent models, our approach could realize a promising accuracy of classification [3]. The proposed model was evaluated with AQ-10 dataset and 250 real dataset collected from people with and without autistic traits. The evaluation results showed that the proposed prediction model provide better results in terms of accuracy, specificity, sensitivity, precision and false positive rate (FPR) for both kinds of datasets [4]. The paper distinguishes four phases by discussing different levels of NLP and components of Natural Language Generation (NLG) followed by presenting the history and evolution of NLP, state of the art presenting the various applications of NLP and current trends and challenges [5].

3. Proposed System

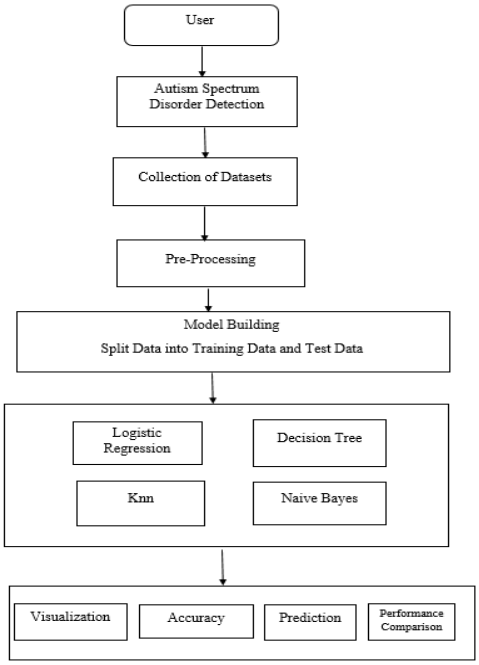

Fig-1 Work Flow of the Proposed System

The work flow of the project heart attack prediction as shown in below fig-1. The dataset is collected to build a model for prediction. Firstly, the dataset is pre-processed using the pre-processing techniques. Find the mean, median and mode of the overall data of each attributes then, check the null values, missing values andirrelevant data in the dataset. Apply the pre-processing technique, after the pre-processing technique, the datasetis saved in .csv file format. 75 percent of the data should be training dataset and then 25 percent should be testing dataset when building the model. To predict autism spectrum disorder, here machine learning algorithms are applied to the data. Later, user can enter the data in web browser, it displays the result and predict the autism spectrum with accuracy of the algorithm.

4. IMPLEMENTATION

4.1 Data Pre-Processing

The data in a dataset can contain a large number of irrelevant and missing components. Various data preprocessing techniques are used to handle this part of the process. These techniques include data cleaning, data integration, data reduction, and data transformation to deal with missing values, data cleaning techniques are used to fix the problem. Missing data, noisy data, etc. are all dealt with. This occurs when there are gaps in the data set. So we have various ways to deal with it they are Data cleaning, Missed Data, Remove Punctuation, Tokenization, Stop words, Vectorization. The dataset is saved in .csv format. CSV files, are a file type that allows us to save tabular data, such as spreadsheets.

4.2 Classification

Data will be classified using a variety of algorithms, including the Logistic Regression Algorithm, the Naive Bayes Algorithm, the Decision Tree Algorithm, and the K-Nearest Neighbor Algorithm, after it has been split into smaller segments. We get different types of results when we use these algorithms.

4.3 Web Interface

The patient's risk factor is obtained in web interface. To access the classifier and check the risk factor, an interactive web interface was created. A Python-based CGI script serves as the backend. It accepts medical form data as input to a trained classification model, which forecasts the individual's risk factor. The below fig-2 shows the user interface.

Fig-2: Web Interface of the project

5. RESULT AND DISCUSSION

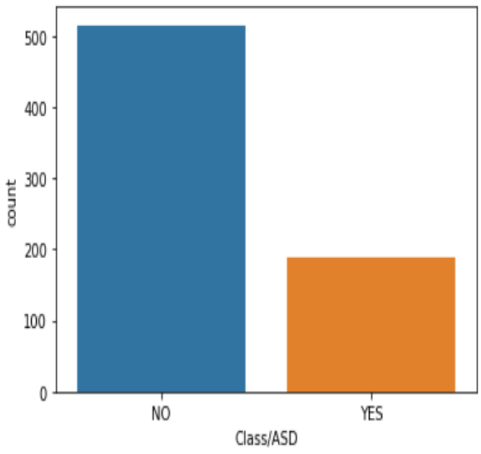

Fig-3: Autism Patient and Normal Patients Calculation

The fig-3 represents the total number of autism disorder patient and normal patient in dataset. There are 189 of normal patient records and 515 autism patient record.

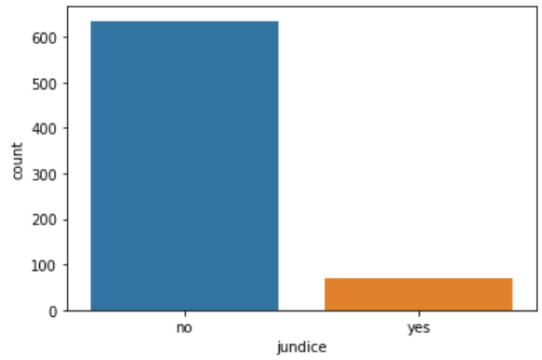

Fig-4: Jaundice Patient and Normal Patients Calculation

The fig-4 represents the total number of Jaundice patients and normal patients.

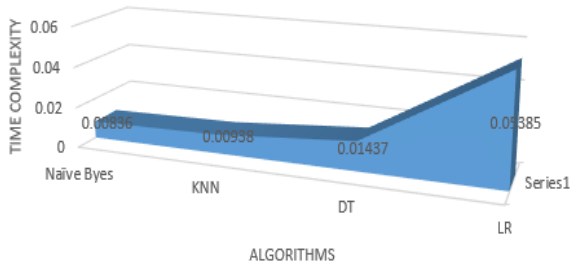

Fig-5: Time Complexity of Algorithms

The fig.5 represents the time taken by all the algorithms to predict the Autism Spectrum Disorder.

Algorithm Used	Accuracy Score	Time Taken (sec)
Naïve Bayes(NB)	96.59	0.00836
K- Nearest Neighbor(KNN)	97.73	0.00938
Decision Tree(DT)	100.00	0.01437
Logistic Regression(LR)	99.37	0.0538

Table-1: Accuracy Score and Time Complexity

From the above results shown in Table-1, we can conclude that Decision Tree algorithm gives the highest accuracy. DT take more time to complete the process, compared to all other algorithms. And least time taken by Naïve Bayes algorithm.

5. CONCLUSIONS

As a result of using machine learning techniques for predictive analysis, we are able to detect autism spectrum disorders earlier, saving lives through the anticipation of a treatment. To predict patients with health care data and healthy patients using different Machine Learning techniques: Decision Tree (100%), KNN (97.73%), Naive Bayes (96.59%), and Logistic Regression (99.37%) Natural Language Processing (NLP) are used. Simulation results showed that the Naive Bayes classifier, KNN, Logistic Regression, Decision tree gives the best results in terms of prediction accuracy and execution time.

REFERENCES

[1] Prediction of Autism Spectrum Disorder Using Supervised Machine Learning Algorithms T. Lakshmi Praveena1 and N. V. Muthu Lakshmi

[2] A survey on various supervised classification algorithms Uma Narayanan; Athira Unnikrishnan; Varghese Paul; Shelbi Joseph.

[3] NLP-Based Approach to Detect Autism Spectrum Disorder in Saccadic Eye Movement Mahmoud Elbattah; Jean-Luc Guérin; Romuald Carette; Federica Cilia; Gilles Dequen

[4] A Machine Learning Approach to Predict Autism Spectrum Disorder Kazi Shahrukh Omar; Prodipta Mondal; Nabila Shahnaz Khan; Md. Rezaul Karim Rizvi; Md Nazrul Islam

[5] Natural Language Processing: State of The Art, Current Trends and Challenges Kazi Shahrukh Omar, Prodipta Mondal

[6] Datasets -https://www.kaggle.com/faizunnabi/autism-screening