Journal of Engineering Sciences
Vol 11, Issue 5,May/2020
ISSN NO:0377-9254

Autism Spectrum Disorder Using Bernoulli's Naive Bayes

Syeda Roshni Ahmed[1], Suvarna A B[2],Vasantha I M[3], Tejaswini V[4]
[1]*Atria Institute OfTechnology,ISE, VTU,syeda.roshni@atria.edu,India*
[2]*Atria Institute Of Technology,ISE,VTU,suvarna8861@gmail.com,India*
[3]*Atria Institute Of Technology,ISE, VTU,ishwarvasantha17@gmail.com, India*
[4]*Atria Institute Of Technology, ISE, VTU,tejashwinikumbi@gmail.com, India*

***Abstract*— Detection of autism spectrum disorder through screening tests is expensive and truly time consuming. With the advancement of artificial intelligence and machine learning (ML), autism areoften predicted at quite early stage. Therefore this particular paper is aimed toward proposing an efficient prediction model supported ML technique and also to develop a mobile application for predicting ASD for people of any age. As outcomes of this research, an autism prediction model was developed by using Bernoulli's Naive Bayes algorithm and to tell why we have used Bernoulli's Naive Bayes algorithm why not any other algorithms in ML? We have taken Support vector machine algorithm for comparison. The evaluation results showed that the proposed prediction model provide better results in terms of all these real world entities such as accuracy, specificity , sensitivity, precision and false positive rate.**

***Keywords*— Machine Learning, AQ-10 dataset, random forest CART, ID3, ASD**

1. INTRODUCTION

Autism spectrum disorder (ASD) it is a developmental disorder, which begins in early childhood and continues throughout life. Autism spectrum disorder affects every aspect of life along the way.Thinking, language, social skills are typically delayed compared to their peers without the disorder. According to WHO [2], about 1 out of each this disorder can live independently, while others require thought life .Diagnosing ASD are often difficult since there's no medical test, sort of a biopsy , to diagnose the disorders. Doctors inspect the child's behavior and development to make a diagnosis's ASD can sometimes be detected at 18 months or younger. By age 2, a diagnosis by an experienced professional are often considered very reliable.

Diagnosis of autism requires significant amount of time and cost. Earlier detection of autism can come to a great help by prescribing patients with proper medication at an early stage. It can prevent the patient's condition from deteriorating further and would help to reduce long term costs associated with delayed diagnosis. Thus a time efficient, accurate and straight forward screening test tool is extremely much required which might predict autism traits in a private and identify whether or not they require comprehensive autism assessment.

The objective of this work is to propose an autism prediction model using ML techniques that could effectively predict autism traits of an individual of any age. In other words, this work focuses on developing an autism screening application for predicting the ASD traits among people aged groups 4-11 years, 12-17 years and for people of age 18 and more.

2. LITERATURE REVIEW

In this section we are briefly discussingabout the works related to the prediction techniques of ASD. The efficiency of ML is quite appreciable in predicting different types of diseases based on syndrome. For example, in[1] Cruztried todiagnosecancerusingML whilein[2] Khan used ML to predict if a person has diabetes or not. [3] Wall usedAlternatingDecision Tree (AD Tree) for reducing the screeningtime and faster detection ofASD traits.

They used Autism Diagnostic Interview, Revised (ADI-R) method and achieved high level of accuracy with adata of 891 individuals and above. But the test was done for limited ages, that's within 5to17 and did not predict ASD for various age groups (children, adolescent andadults).

Which we are implementing during this paper Bone [4] applied ML for an equivalent purpose and have used support vector machine (SVM) to get 89.2% sensitivity and59% specificity. Their research included 1000 individuals with ASD and 462 individuals with NON-ASD traits. Thanks to wide2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019range aged (4-55 years) their research wasn't accepted for people of all age bracket

as screening approach. Then Allison [5] used 'Red Flags'tool.

Here is some examples of what's meant by Red flag: Limited use of gestures like giving, showing, waving, clapping, pointing, ornodding their head, late speech or no social babbling/chatting, Makes weird sounds or has an unusual tone of voice. For diagnosing ASD with Autism Spectrum Quotient for children and adult, then they need shortlisted them to AQ-10 dataset with quite 90% accuracy. Then comesThabtah [6] he compared the previous results of autism traits which was done using ML Algorithm, while Hauck and Kliewer [7] tried to spot relatively more important screening questions for ADOS (Autism Diagnostic Observation Schedule) And ADI-R (Autism Diagnostic Interview Revised) screening methods and located that ADI-R and ADOS screening test can work better once they both are combined together. Backroom [8] used several ML techniques including naive Bayes, SVM and random forest algorithm to work out ASD traits in children like developmental delay, obesity, less physical activity and compared those results. Wall [9] did a study on classifying autism with short screening test andvalidation percentage of sensitivity, specificity and accuracy.

Heinsfeld [10] applied deep learning algorithm and neural network algorithm to identify ASD patients using large brain imaging dataset from the Autism Imaging Data Exchange(ABIDE I) and achieved a mean classification accuracy of 70%with an accuracy range from 66% to 71%. The SVM classifier achieved mean accuracy of 65% while the Random Forest classifier achieved mean accuracy of 63%. Liu [11] did area search on whether the face scanning patterns could be put in use to identify children with ASD by adopting ML algorithm to analyze an eye movement dataset for theclassification purpose. This study showed an accuracy of 88.51%; specificity86.21%; sensitivity 93.10%; AUC 89.63%. Bone [12] analyzed the previous works of Wall et al and Kosmickietal [13] to seek outthe problems in conceptual problem formation, methodological implementation and interpretation then reproduced the result using Ml approach

From the literature review it is seen that, though a number of researches have been carried out in this field but the researchers did not come to adecisive conclusion on using the ML approach to generalize autism screening test tool in terms of the age groups. Different tools and techniques are adapted before for autism screening tests, but none within the sort of app based solution for different age groups. So this is what we are trying to overcome in this base paper.

3. METHODOLOGY

The project is carried out in four different steps, firstly the data collection, secondly the Data Synthetization, third the Developing the prediction model, and the forth Evaluating the prediction model.

A brief Description of the Steps is given below:

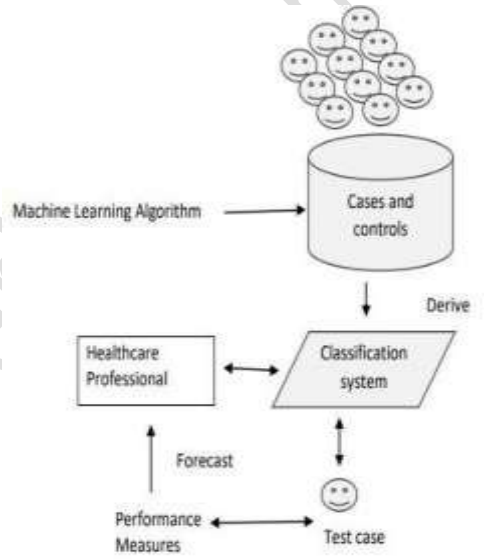

Fig. 1 Architecture diagram of the project

3.1 Data collection:

Data Collection may be a process of gathering and measuring information on targeted variables duringa systematic fashion. Formal data collection process is required as it ensures the data is defined and accurate so that the decisions based on the data arevalid.

3.2 Data Synthetization:

The data is collected from hospitals and different repository then this data is synthesized to remove irrelevant features and rows. For example, the ID column was irreverent to develop a prediction model, thus it isremoved.

3.3 Naive Bayes Classification using Bernoulli's

The Naive Bayes classification algorithm may be a probabilistic classifier. It is based on

probability models that incorporate strong independence assumptions by which the diseased genes can be classified. Naïve Bayes is a conditional probability model.

The Bernoulli naive Bayes classifier assumes that all our features are binary such that they take only two values that IA true value is considered as 0 and negative values as 1 P(Ck)=P(Ck).P(x|Ck)P(x)(1)

3.4 Evaluating the Prediction Model:

With the classified dataset (training dataset) the test data can be predicted for autism [3]. And the corresponding positive and negative predictions with their probabilities are obtained. A comparison is made with another algorithm Support vector machine (SVM) to bring out the reason for doing with Naïve Bayes algorithm .The result of comparing the two algorithms proved that Bernoulli's Naive Bayes algorithm works better for Autism for all age groups with a 95%accuracy.

4. IMPLEMENTATION

4.1Collection

The data required for the autism predictionis the heterogeneous genomes which vary from each individual. Autism is a heterogeneous neurodevelopmentsyndrome. It involves complex genetics etiology, DNA and gene. It has a large dataset with complex genetic structureswhich has to be handled to remove the noisy and inconsistent data. The dataset will be in the form of AQ 10 dataset developed using Autism spectrum tool. The AQ 10 dataset is divided into three types based on the age. Child AQ dataset (4- 8 years), Adolescent AQ 10 dataset (teen12-18 years) Adult AQ 10 (Adults 40-65 years).

4.2 Pre Processing:

The Preprocessing of genetic data includes the following:

4.2.1 Data Transformation:

Normalization: scaling the values to a specific range. Aggregation: assigning probabilistic values to the genes. Construction: replacing or adding new genes inferred by the existing genes.

4.2.2 Data Reduction:

Searching for a lower dimensional space which will best represent the info. Removing the irrelevant data from the genome dataset. Sampling can be used to simplify the process of classification using smalldataset.

4.2.3 Applying Algorithm:

Below is the basic algorithm used for solving any Bernoulli problems?

TRAIN BERNOULII NB(C, D)
1 V ⟵EXTRACT VOCABULARY (D)
2 N ⟵COUNTDOCS (D)
3 for each c ϵ C
4 do N_C ⟵COUNT DOCS IN CLASS (D, c)
5 prior[c] ⟵N_C/N
6 for each t ϵ V
7 do Nct ⟵COUNT DOCS IN CLASS CONTANT TERM (D, c, t)
8 condprob[t] [c] ⟵(Nct +1)/ (N_c +2)
9 return V, prior, condprob

APPLY BERNOULII NB(C, V, prior, condprob, d)
1 V_d ⟵EXTRACT TERMS FROM DOC (V, d)
2 for each c ϵ C
3 do score [c] ⟵log prior[c]
4 for each t ϵ V
5 do if t ϵ V_d
6 then score[c] + = log condprob[t] [c]
7 else score[c] + = log (1 – condprob[t] [c])
8 return arg $\max_{c \epsilon C}$ score[c]

Given a genetic instance to be classified, represented by a vector x=(x1, x2…..xn) representing some n genes with the assigned probabilities.

P (Ck, x1…..xn) for every of k possible outcomes or classes Ck. Thus the diseased genes are classified using Bayes' Theorem. We have divided the set into training and testing dataset, where 80% of dataset is taken for training .same methodology used for the SVMalgorithm.

4.3 Creating Web Page:

We have used flask frame work to design front end. Syntax such app.route is used to connect html page.seperate html pages contain design layout of the webpage. Designingweb page is done using normal htmltag. The session kept active using secret key .after running the algorithm we have come to conclusion that Bernoulli's naive

base 95.66% accuracy where a svm gives 100%accuracy.

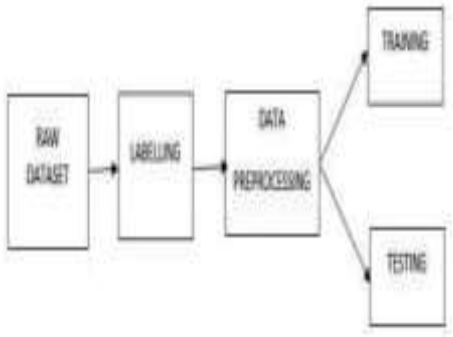

Fig. 2 Data Pre Processing Steps

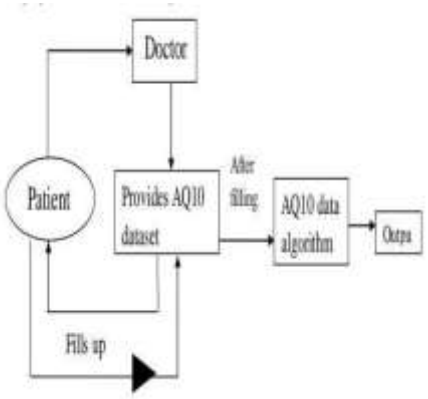

Fig. 3 Flow Chart for Implementation

5. ADVANTAGES AND DISADVANATAGES

5.1 Advantages:

1. Improves observation skills there is a listen, learn, look approach to learning.
2. Helps to obtain absorb and retain factsthe long term memory excellent with superiorrecall.
3. Visual skill-tend to be visuallearners and detailfocused.

5.2 Disadvantages:

1. Being bullied in school or other places where people are ignorant above the condition and what it really.
2. There is therapy known as applied behavioral analysis which is expensive.
3. Since all data entered in the form will be entered in a binary format it becomes difficult for patients to fill in without the medical team assistance.

6. RESULTS:

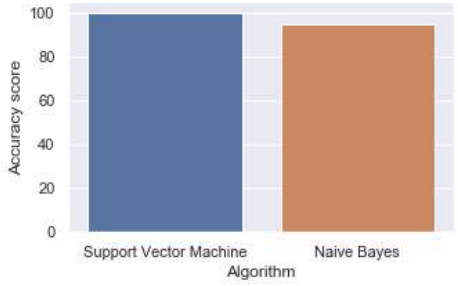

Fig. 4 Comparison Graph

Fig .5 Home Page

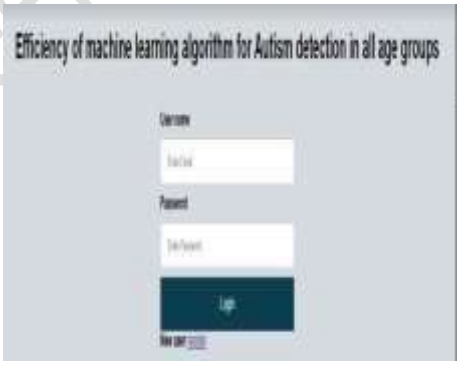

Fig .6 Login Page

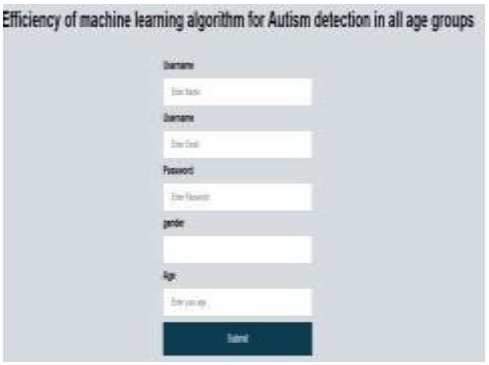

Fig .7 Register Page

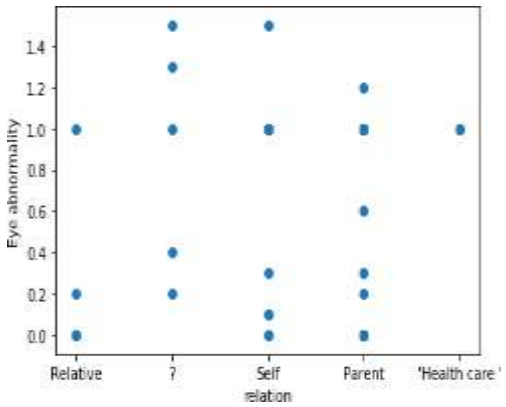

Fig .8 Different attribute Graph

7. CONCLUSIONS

A Research Contributions Previous researches provided outcomes in percentages using many machine learning algorithms.

Those research provided three fold outcomes: firstly, a prediction model was developed to predict autism traits. Using the AQ-10 dataset, the proposed model can predict autism with 92.26%, 93.78%, and 97.10% accuracy in case of child, adolescent and adult persons, individually using different methodologies, for each age group but not a single algorithm for all age's groups at atime.

Those results were considered much better than those screening tests performed manually using medical equipment'searlier. Our proposed model can predict autism traits for different age groups, while many other existing approaches (like [5]) missed this feature.

Secondly, it shows the comparative view of ML approach in terms of their performance. Earlier results showed that Random Forest-CART showed better performance than the Decision Tree-CART algorithm.Comparing to both the Random Forest- CART and Decision Tree-CART algorithm.

Finally, our model works for limited amount of dataset and is able is predict whether a person is having this autism trait or not at the earliest. This outcome indicated an extension of the many other existing work, since most of the prevailing works mainly specializein developing and comparing the performance of prediction model or techniques. As Autism Spectrum Disorder is not genetically inherited disorder it becomes quite difficult to predict at the earliest even using the best screening equipment? So now with the help of technology and machine learning algorithms it is has been easier for the doctors to take the correct step and provide necessary precautions.

REFERENCES

[1] J. A. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis," Cancer informatics, vol. 2,2006.

[2] N. S. Khan, M. H. Muaz, A. Kabir, andM.N. Islam, "Diabetes predicting mhealth application using machine learning," in 2017 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE). IEEE, 2017, pp. 237–240.

[3] D. P. Wall, R. Dally, R. Luyster, J.-Y. Jung, and T. F. DeLuca, "Use of artificial intelligence to shorten the behavioral diagnosis of autism," PloS one, vol. 7, no. 8, p. e43855,2012.

[4] D. Bone, S. L. Bishop, M. P. Black, M. S. Goodwin, C. Lord, and S. S. Narayanan, "Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion," Journal of Child Psychology and Psychiatry, vol. 57,2016.

[5] C. Allison, B. Auyeung, and S. Baron- Cohen, "Toward brief "red flags" for autism screening: the short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls," Journal of the American Academy of Child & Adolescent Psychiatry, vol. 51,2012.

[6] F. Thabtah, "Autism spectrum disorder screening: machinelearning adaptation and dsm-5fulfillment,"in Proceedings of the1st International Conference onMedical and Health Informatics 2017. ACM, 2017.

[7] F. Hauck and N. Kliewer, "Machine learning for autism diagnostics: Applying support vectorclassification."

[8] B. van den Bekerom, "Using machine learning for detection of autism spectrum disorder,"2017

[9] D. Wall, J. Kosmicki, T. Deluca, E. Harstad, and V. Fusaro, "Use of machine learning to shorten observation-based screening and diagnosis of

autism," Translational psychiatry, vol. 2, no. 4, p. e100, 2012.

[10] S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the abide dataset," NeuroImage: Clinical, vol. 17,2018.

[11] W. Liu, M. Li, and L. Yi, "Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework," Autism Research, vol. 9, no. 8, pp. 888–898,2016.

[12] D. Bone, M. S. Goodwin, M. P. Black, C.-C. Lee, K. Audhkhasi, and S. Narayanan, "Applying machine learning to facilitate autism diagnostics: pitfalls and promises," Journal of autism and developmental disorders, vol. 45, no. 5, pp. 1121–1136,2015.

[13] J. Kosmicki, V. Sochat, M. Duda, and D. Wall, "Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning," Translational psychiatry, vol. 5, no. 2, p. e514, 2015.

[14] F. Thabtah, "UCI machine learning repository," 2017. [Online]. Available: https://archive.ics.uci.edu/ml.

[15] T. Booth, A. L. Murray, K. McKenzie, R. Kuenssberg, M. O'Donnell, and Burnett.