International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

Predicting Autism Spectrum Disorder using Supervised Learning Algorithms

P. Kavya Sree[1], A. Phani Sai Kumar[2], R.Nandini[3], Tanay Lodh[4], N Sai Susmitha Naidu[5]

[1234] *UG student, Department Of Computer Science And Engineering, GITAM School of Technology, Visakhapatnam campus, India.*
[5]*Assistant Professor, Department Of Computer Science And Engineering, GITAM School of Technology, Visakhapatnam campus, Andhra Pradesh, India.*

Abstract - *Autism Spectrum Disorder (ASD) is one of the most rapidly rising diagnoses in the field of developmental disabilities. Behavioral characteristics, including social, sensory, and motor elements, are used to diagnose autism spectrum disorder. Families and individuals concerned about ASD traits in themselves or their relatives should consult their family physician. So, using ten enquiries, we looked at the behavioural pattern to see if the person had been diagnosed with autism or not. We have a total of 704 records, of which 189 were found to be autistic, accounting for around 26.85% of the total. From the total 20% of the records were utilised to train the algorithm, with the rest being used to test it. The algorithms used were random forests (RF), AdaBoost (Ada), and the Support Vector Machine (SVM) algorithm. These algorithms were used to find the accuracy of the dataset.*

Key Words: **Autism Spectrum Disorder, ASD, Machine Learning, Data Analysis, Flask Application**.

1. INTRODUCTION

Autism spectrum disorder (ASD) is a lifelong neurodevelopmental disorder marked by social communication and interaction difficulties. ASD is a rapidly growing problem in today's world, affecting people of all ages. Early identification of this neurological condition can help keep the subject's mental and physical health in good shape. Detection of ASD is challenging due to the fact that there are various different mental diseases with few symptoms that are quite similar to those with ASD symptoms, making this a tough endeavour.

Autism is on the rise all across the world, and it's spreading at an enormous speed. Some people with this disorder can live on their own, while others may require permanent care and help. Autism diagnosis takes too long and is very expensive. The symptoms of this condition can begin as early as three years old and can last a lifetime. Although it is impossible to completely heal a patient suffering from this condition, the consequences can be mitigated for a period of time if the signs are discovered early.

In order to minimize the symptoms of ASD and improve the quality of life of those affected, early detection and treatment are essential. However, no medical test that can accurately diagnose autism. The goal of this project is to develop a machine learning model that can predict autism. The model can be used to predict autism traits in any age group.

The remaining part of the paper is as follows. The second section of the paper goes over previous research in this field. The methodology is presented in Section III. The proposed system's detailed implementation is discussed in Section IV, and the feature built is analyzed in Section V. The proposed algorithm was briefly discussed in Section VI and how it was integrated into a flask application. Finally, Section VII summarises the paper's research contributions, research limitations, and future plans to develop on this research.

2. LITERATURE REVIEW

The course of early screening isn't extremely reliable in light of the fact that it depends on a questionnaire that guardians reply about their ward's way of behaving. Furthermore, these surveys regularly produce misleading false positives. Truth be told, of the kids whose guardians report early indications of ASD on the poll, just 50% have that diagnosis affirmed by an authorised ASD clinician. Also, in light of the fact that there are scarcely any authorised ASD clinicians qualified to circle back to the many guardians who report through ASD through the poll, the wait time that child will get a diagnosis could be well after the child's third birthday postponing treatment past the critical window of time to further develop results for child's with ASD possibly.

I. M.S. Mythili, A.R. Mohammed Shanavas made use of classification techniques in their study on ASD. The objective of their research was to study the autism disorder and the various levels it exists in. In this study the subject's behavioural and social interactions were analysed with the use of SVM, Neural network and Fuzzy techniques with WEKA tools.

II. Wall made use of Alternating Decision Trees (ADTree) to reduce the time takes to screen and detect ASD traits in a subject in an effort to make the diagnosis procedure significantly faster. They observed higher levels of accuracy with a data of 891 individuals by using a revised ADI-R method. Although the test failed to accurately predict ASD in varying age groups since the

study was limited within the age bracket of 5 years to 17 years.

III. A searching method was proposed by A. Kosimchi, V.Suchat, M.Duda and D.P. Wall over the set of traits in Autism prediction. The authors used a ML approach in this study for the evaluation of the clinical diagnosis of ASD. This study employs 8 different ML algorithms to perform step-by-step backward recognition on 4540 sheets.

IV. Bone used a Support Vector Machine in their study, which included 1264 people with ASD traits and 462 people without ASD traits. In this study they obtained 89.2% percent sensitivity and 59% specificity.

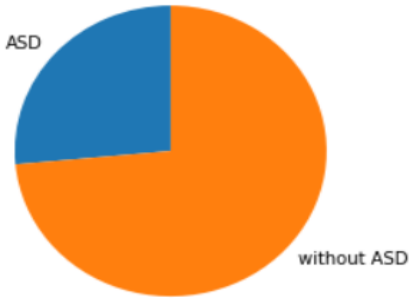
From the preceding section, it is evident that deep learning-based models have a great potential to detect ASD in human populations. The majority of works cited so far have relied on traditional machine learning algorithms, which are limited in terms of performance. In this study, multiple machine learning models were compared to see which was the most efficient.

3. RESEARCH METHODOLOGY

In this dissertation, five phases are involved: data collection, data pre-processing, developing the prediction model, assessing the prediction model, and developing an application in Flask. The following subsections briefly describe each phase:

3.1 Data Collection:

The dataset for this study was gathered from the UCI Repository, which is open to the public. Dataset Name-ASD Screening Data for Adult, Attribute Type-continuous and binary, Number of Instances-704. There are sixteen characteristics in the dataset, which are a combination of categorical and numerical data of which include: Question 1-10, Age, Gender, if the person born with jaundice, Any Family member with ASD, who is completing the test, and Class. The AQ-10 screening questions cover a variety of domains, including attention switching, imagination, communication, and social interaction. The questions are graded using a one-point scoring system for each of the ten questions. On each question, the user might earn 0 or 1 point depending on their response.



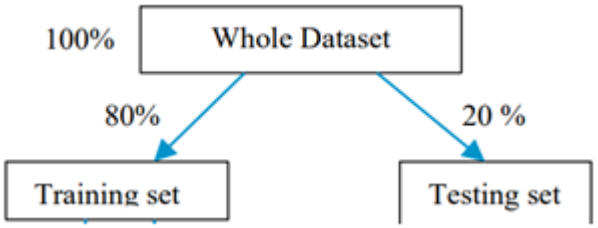
Number of Data set

3.2 Data Pre-processing

Data Pre-processing is for removing outliers and noise in the raw data and makes it available for training the model. In simple way, Data Pre-processing is the major step in the project evaluation to obtain the best accuracy. Thus, raw data is converted into something usable and understandable. Real-world data is frequently partial and inaccurate because it contains many errors and outliers. Several methods are to handle data, such as handling incomplete data, outlier analysis, data reduction, and discretization. The missing values in these datasets were solved using the imputation method.

3.3 Developing the prediction model

Considering an 80:20 ratio, the complete dataset is categorized into training set and testing set respectively as per requirements. In the wake of implementing several types of supervised learning systems such as random forests (RF), SVM, and AdaBoost, a system to predict ASD could be put into place.



3.4 Evaluating the prediction model

In addition to assessing accuracy, sensitivity, specificity and precision, the proposed model was also tested using the leave-one-out strategy on the AQ-10 dataset. As part of the validation process, field observations were conducted at various places using forms to collect over 189 ASD cases and 515 cases without ASD from a special education institute for people with special needs.

3.5 Developing a flask application

In the end, a flask application has been developed specifically for the general public. The user answers closed-ended questions to receive a result with regard to autism or not.

4. DEVELOPING THE PREDICTION MODEL

Data was split into two portions, one to train on and the other to test on: 80:20. For training, 80 % of the data is used, and for testing, 20 % is used. Algorithms were created, and their accuracy was tested to generate autism trait predictions. Random forest is used for implementing the ASD predictive system after completing several types of supervised learning, including random forests (RF), SVM, AdaBoost.

Initially, the SVM algorithm was used to predict an individual's autism traits. Random Forest was implemented for further improvement. Finally, the Random Forest classifier was tweaked to improve results, and the best results were obtained using the AdaBoost algorithm. The following are the 3 algorithms used:

4.1 Prediction model based on SVM

A SVM model (Support vector) is machine learning method used for regression and classification that study and examines data. SVM is a supervised learning method which divides the data into two groups based on the hyperplane N dimension. N represents the total of different features used to classify the data points.

4.2 Prediction model based on Random Forest

The best predictors from a subset of predictors chosen at random are used to split each node in a random forest. This seemingly paradoxical method outperforms other classifiers such as SVM, and neural networks. Random Forest classifier was used to improve the model's accuracy. The algorithm can be divided into two steps in this case as well: producing random forest and classifying test data.

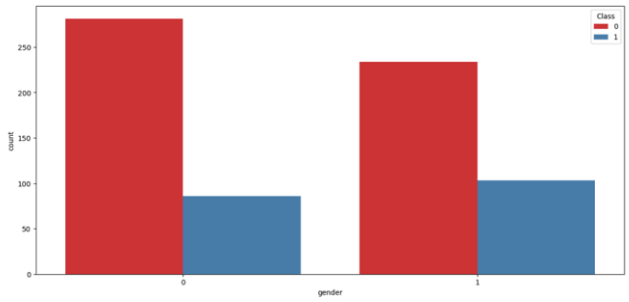
4.3 Prediction model based on AdaBoost

One of several boosting algorithms AdaBoost is first boosting algorithm in solving practices. It is used to merge several "weak classifiers" into one "strong classifier."

Weak classifiers are introduced one by one and trained on a weighted dataset. After creating a predetermined number of weak learners (a user parameter), the process is repeated until the training dataset cannot be improved anymore. AdaBoost is frequently referred to as the best out-of-the-box classifier (with decision trees as the weak learners).

AdaBoost algorithm offers the advantage of predicting more accurate results and has better performance than the current framework.

5. EVALUATING THE PREDICTION MODEL



The above bar graph represents the whether the person is autistic or not with respect to the gender.

	precision	recall	f1-score	support
0	0.75	1.00	0.86	106
1	0.00	0.00	0.00	35
accuracy			0.75	141
macro avg	0.38	0.50	0.43	141
weighted avg	0.57	0.75	0.65	141

The results of the SVM Algorithm

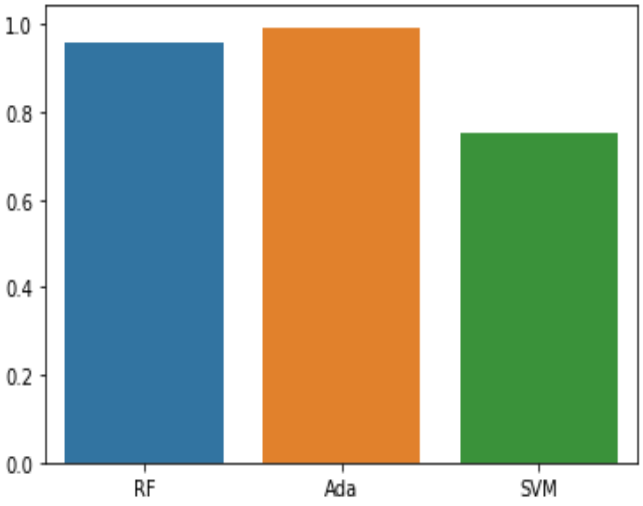
	precision	recall	f1-score	support
0	1.00	0.97	0.99	106
1	0.92	1.00	0.96	35
accuracy			0.98	141
macro avg	0.96	0.99	0.97	141
weighted avg	0.98	0.98	0.98	141

The results of the AdaBoost Algorithm

	precision	recall	f1-score	support
0	1.00	1.00	1.00	106
1	1.00	1.00	1.00	35
accuracy			1.00	141
macro avg	1.00	1.00	1.00	141
weighted avg	1.00	1.00	1.00	141

The results of the Random Forest Algorithm

The precision, recall, support and f1-score of the random forests algorithm are shown in the above figure.

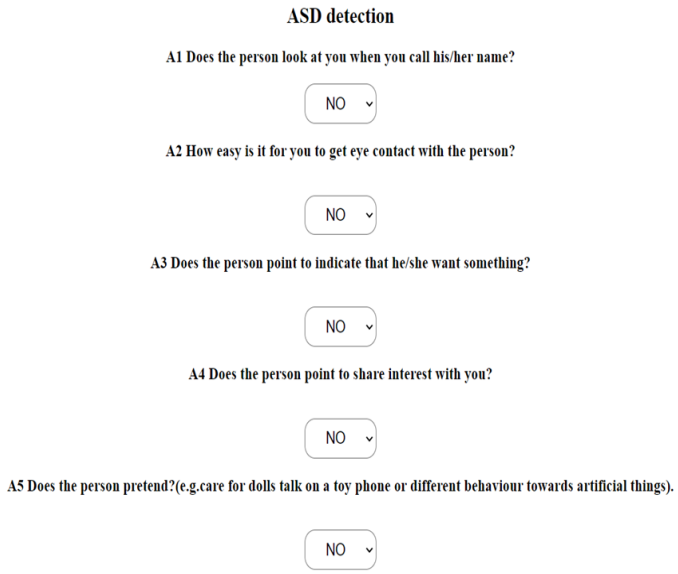


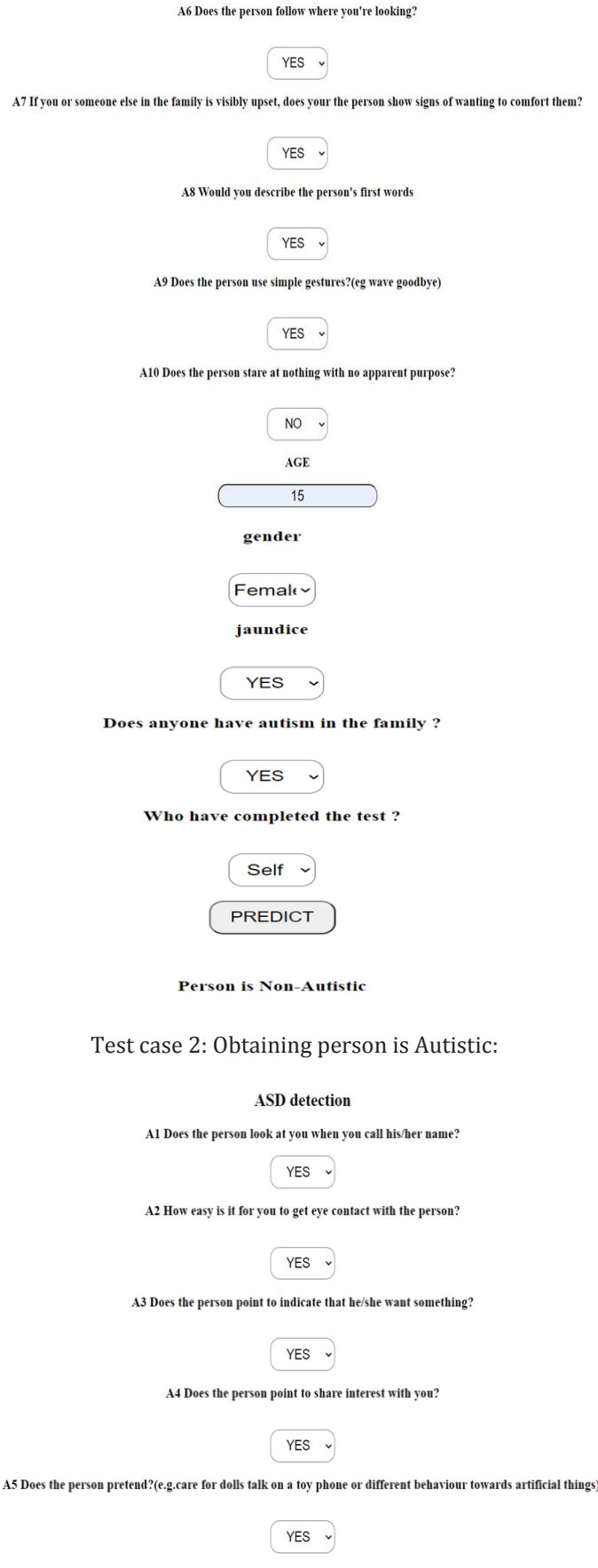
Above graph represents the comparison among the different models used. Random forest and AdaBoost models have the higher accuracies when compared to SVM.

6. DEVELOPING THE FLASK APPLICATION

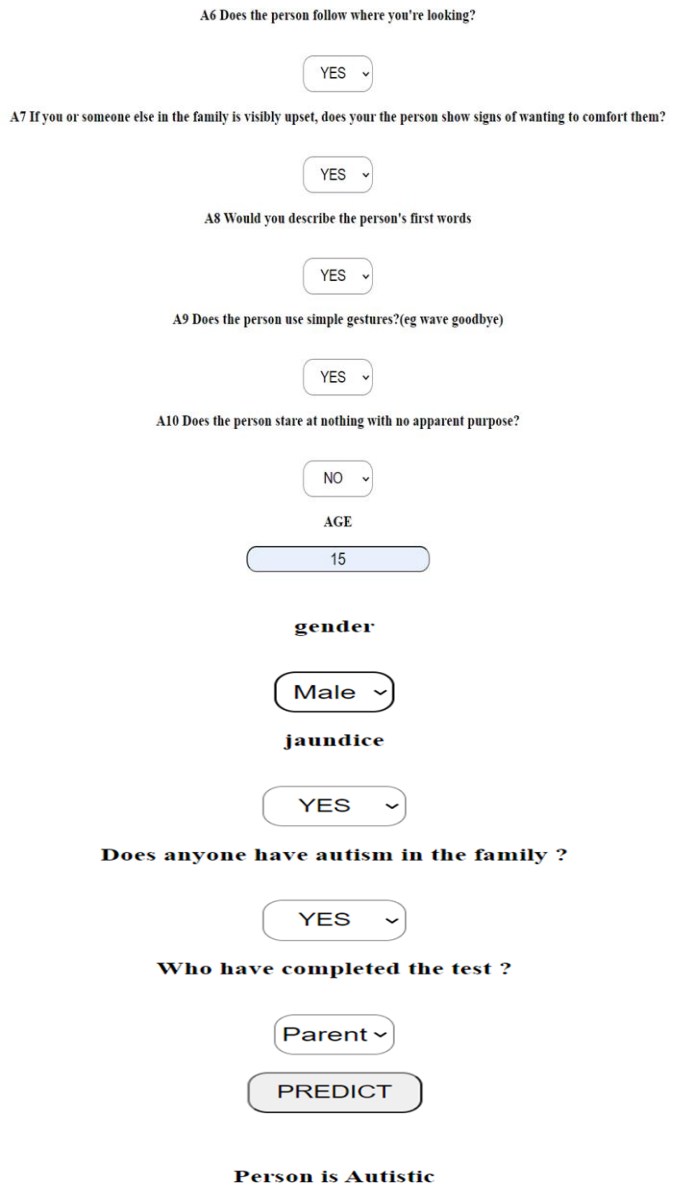
In order to make the application user-friendly, the Random Forest algorithm was integrated with flask into a screening web application. Based on the AQ-10 characteristics and some other inputs, various questions were asked for different age groups. According to the inputted answers by the user, the application indicated whether the person is autistic or not.

Test case 1: Obtaining person is Non-Autistic:





Test case 2: Obtaining person is Autistic:



7. DISCUSSIONS AND CONCLUSIONS

The dataset for this study was gathered from the UCI Repository, which is open to the public. Dataset Name-ASD Screening Data for Adult, Attribute Type-continuous and binary, Number of Instances-704. The dataset includes 16 parameters, the first 10 questions contain the behavioural characteristics of the person based on communication and social interaction. Next parameters are gender, age, person born with jaundice, any family member containing autism and who completed the dataset. The questions are graded using a one-point scoring system for each of the ten questions. On each question, the user might earn 0 or 1 point depending on their response. The entire dataset has been divided into two sections, one for training and the other for testing, using an 80:20 ratio. Algorithms were created and their accuracy was tested in order to generate autism trait predictions. With the adoption of many machine learning algorithms such as RF, SVM, and AdaBoost. The implementation of the ASD predictive system was proposed using random forest. Using the AQ10 dataset and real-world data, the proposed predictive model was assessed for efficiency, specificity, and precision. On the AQ-10 dataset, the leave one out technique is used to test the performance of the suggested model. Using the AQ10 dataset and real-world data, the proposed predictive model was assessed for accuracy, correctness, validity, efficiency. A total of 704 records were examined, with 189 being identified as autistic, accounting for around 26.85% of the total and 515 records are identified as non-autistic accounting for 73.15%.

8. Limitations and Future Work

The key constraint of the project is the absence of sufficient information to develop the prediction system. Our next steps will be to collect more data from a range of resources and improve the efficiency of the proposed machine learning classifier.

In our future endeavors, we would like to add audio as an output to our developed flask application to make it more reliable for users.

9. REFERENCES

[1] A Study on Autism Spectrum Disorders using Classification Techniques. M. S. Mythili, A. R. Mohamed Shanavas.

[2] M. Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. J A Kosmicki [1], V Sochat [2], M Duda [3], D P Wall [3]

[3] N. S. Khan, M. H. Muaz, A. Kabir, and M. 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.

[4] 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.

[5] 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.

[6] 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

[7] Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion. Daniel Bone,[1] Somer Bishop,[2] Matthew P. Black,[3] Matthew S. Goodwin,[4] Catherine Lord,[5] and Shrikanth S. Narayanan[1]

[8] 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.

[9] F. Thabtah, "Autism spectrum disorder screening: machine learning adaptation and dsm-5 fulfillment," in Proceedings of the 1st International Conference on Medical and Health Informatics 2017. ACM, 2017.

[10] F. Hauck and N. Kliewer, "Machine learning for autism diagnostics: Applying support vector classification."

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