International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 04 | April - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Autism Spectrum Disorder (ASD) Prediction

Miss. Senjuti Bhattacharya[1], Miss. Rutuja More[2], Mr. Mohit Shirke[3], Prof. Amol Patil[4]

[1]*Student, Dept. Information and Technology, A.C. Patil College of Engineering, Maharashtra, India*
[2]*Student, Dept. Information and Technology, A.C. Patil College of Engineering, Maharashtra, India*
[3]*Student, Dept. Information and Technology, A.C. Patil College of Engineering, Maharashtra, India*
[4]*Professor, Dept. Information and Technology, A.C. Patil College of Engineering, Maharashtra, India*

Abstract - Autism spectrum disorder (ASD) is a complex developmental condition which involves persistent challenges in social interaction, speech along with nonverbal communication, and restricted/repetitive behaviors. Effects of ASD and the severity of symptoms are different from person to person. Early diagnosis and treatment are important in managing the symptoms of autism and improving quality of life for people with autism and their families. There is no medical test for autism. It is diagnosed based on observing how the child talks and acts in comparison to their peers (children of the same age). We will be building a model for predicting ASD like traits on the basis of simple curated questions asked to the child or their parents/caregivers with the of Machine Learning's Supervise Learning Algorithm – KNN Classifier Model and will be deployed by Flask. Flask also Jinja 2 is a template engine written in pure Python. This provides with a Django-inspired non-XML syntax but supports inline expressions and an optional sandboxed environment. Flask is a Python-based micro web framework which allows you to write web applications quickly and efficiently. Flask uses templates to expand the functionality of its web applications while maintaining a simple and organized file structure. Templates are enabled using Jinja2 template engine and allow data to be shared and processed before being turned into content and sent back to the client.[6].

Key Words: ASD, KNN, K Nearest Neighbors Classifier, Flask, Jinja 2

1.INTRODUCTION

Autism Spectrum Disorder (ASD) also known as Autism is a developmental disability that can cause significant social, communication and behavioral challenges. Frequently it is nothing about how people with ASD look that sets them apart from other people, but people with ASD may communicate, interact, behave, and learn in ways that are different from most of the population. The learning, thinking, along with problem-solving abilities of people with Autism Spectrum Disorder can range from gifted to severely challenged. Some people with ASD need a lot of help in their daily lives; whereas others need less help.[1]. Autism differs from person to person in severity along with combinations of symptoms. There is a great range of abilities and characteristics of children with ASD — no two children appear or behave in the same way. Symptoms can range from mild to severe and can often change over time. The wide range of symptoms can in the future bring more traits on which classification can be done and thus improve performance of the model. Making the availability of better predictive models and wider datasets can prove revolutionary in understanding ASD and bringing upon ease in caregiving facilities. Understanding ASD will remove the stigma and pre notions attached to it and will bring awareness. As of to date there is no specific test to test for ASD but more of a complex bunch of logical, cognitive and psychometric assessment tests which need to be done on definitive intervals of a child's growth intervals in their early years, like at 9 months, 18 months and 30 months. Problem with these tests is that they at times are ambiguous, time consuming and very lengthy which can delay the help needed for the child if they actually have ASD because time, is a key constraint. These tests also have an accuracy of 70% only leading to the need of creating a better, faster, concise and more efficient system.

2. LITERATURE SURVEY

1. A report by Fadi Thabtah and David Peebles, proposed a new machine learning model based on induction of rules for autism detection. This paper focuses of building a system which incorporates these qualities into their system. The paper proposes to use datasets of ASD traits responses in children as well as adults and the working on various machine learning algorithms mainly supervised learning algorithms and predicting ASD. To further more classify this paper aims to build a new covering approach known as the Rules Machine Learning or RML. Alongside other observations, this study had also aimed to limit the role of manually derived rules embedded within current assessment tools using Machine Learning to increase classification accuracy, sensitivity, and specificity. This is particularly necessary for cases that are difficult to classify (e.g., cases unclearly associated with an ASD type).[2]

2. A report by Ayelet Ben-Sasson, ScD; Diana L Robins, PhD; Elad Yom-Tov, PhD which is Risk Assessment for Parents Who Suspect Their Child Has Autism Spectrum Disorder: Machine Learning Approach. The study has aimed to evaluate the ASQ and M-CHAT-R risk status of children with parents who have concerns in their social-communication domain, and whether this status correlates with a family history of Autism and expert assessment of ASD risk. This study aimed to test the possibility of automatically estimating a child's risk with ASD based on their parent's description of concerns. Thus, relying on the parent's text and a particular M-CHAT-R question that complemented that unique text. However, as this approach identified about two children who would have been missed if the M-CHAT-R/F had been administered only to those who screened positive on the initial M-CHAT-R screening. This procedure may have had led to elevated risk rates on the M-CHAT-R as the risk rates were higher than those reported in another study of infant siblings of the children with ASD.[3]

3. A report by Wenbo Liu, Zhiding Yu, Bhiksha Raj, Li Yi, Xiaobing Zou, Ming Li proposed an ASD prediction system based on machine learning techniques. The work features the novel development and application of machine learning methods over traditional ASD evaluation protocols. Specifically, interested in discovering the latent patterns that possibly indicate the symptom of ASD underneath the observations of eye movement. A group of subjects (either ASD or non-ASD) are shown with a set of aligned human face images, with eye gaze locations on each image recorded sequentially and then model is accordingly trained. The key focus of the paper lies in conducting eye movement analysis at (face) image level. Specifically, a single feature is extracted from the eye movement data recorded per face image per subject and then each feature is labeled either positive or negative based on the identity (ASD/non-ASD) of the subject. Once the features are extracted, a prediction model is trained and is tested at image-level for the new test subject.[4]

4. A report by Devika Varshini G and Chinnaiyan R proposed an optimized Machine Learning Classification Approaches for Prediction of Autism Spectrum Disorder. The paper uses datasets of ASD traits responses in children and their working on various machine learning algorithms such as Logistic Regression, Random Forest Classifier, Decision Tree Classifier and predicting ASD and calculating accuracy, precision, recall and F-1 Score. The proposed system has an Accuracy Score and the accuracy for KNN is 69.2%. With increasing the dataset, the F1-score remains almost the same creating consistent predictions. Some of the predicted responses were found contradictory to the actual medical data as known and had given a rise in ambiguity.[5]

3. PROPOSED SYSTEM

The proposed system is for testing ASD in a concise and accurate way by developing a Web based Application using fundamental concepts of Machine Learning with Python 3.x for Data Preprocessing, Feature Engineering, Graphical Representations and the use of Supervised Algorithm – KNN Classifier to train and test the dataset for the model using various other libraries of Python, pickle dumping the model onto Flask 2.0 and the with some help of HTML and CSS for a better UI/UX Experience.

3.1 Module 1: Signup

User can at first signup with the use of email wherein email acts as username, the password will be sent over to the user's email address. As the user enters details at signup, they get added into the system's database which includes their details as asked.

3.2 Module 2: Login

User can login with credentials of email as username and the password sent on the email address as password respectively. These details are cross checked with database to verify and let the user access the system if user has signed up. If user hasn't signed up yet they are redirected to signup page. If details are present in the system and username and password used by the user is correct then they are granted access to the system.

3.3 Module 3: Input Data

The system asks a set of 10 questions which needs to be answered in Yes or No format. Answers are taken according to the response gained from the child in question and can be filled by the child itself or by their caregiver/parent and/or a Professional Health Care personnel. Along with this user also has to enter child's age in months, i.e., between the values 12 to 36 only (Since toddlers only). A few questions regarding any medical history of child's family and the child themselves is also asked to complete the questionnaire.

3.4 Module 4: Model Prediction

Based on the responses as registered to the system, model will predict if the child has ASD traits else will predict the child has no ASD traits and won't need medical/special care.

3.5 Module 5: FAQ Section

For Frequently Asked Questions and doubts of parents and caregiver, FAQ section will be provided and for other queries to be answered, user can leave a query to be answered by Medical Professionals. Queries asked by the user is stored into a database on the backend.

3.6 Module 6: Forget Password

If user forgets their password or fails to find password, they can have their password resent from this page. User enters their registered email address which is verified if present in the database and if present, new password will be sent.

Following figures, Fig 1. and Fig 2. are Block Diagram and Activity Diagram respectively which depicts the design of the system and its various components.

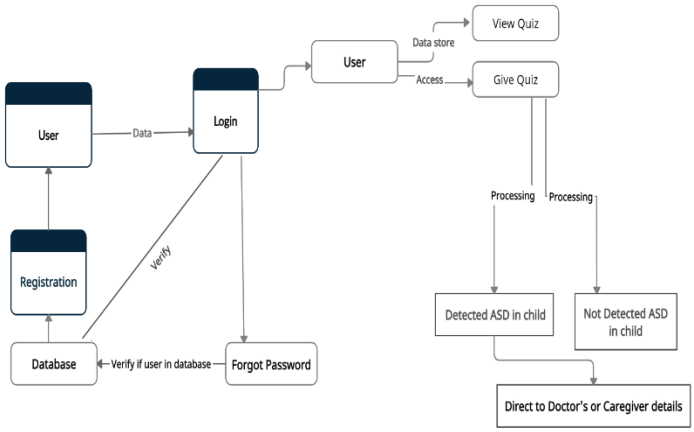

Fig 1: Block Diagram

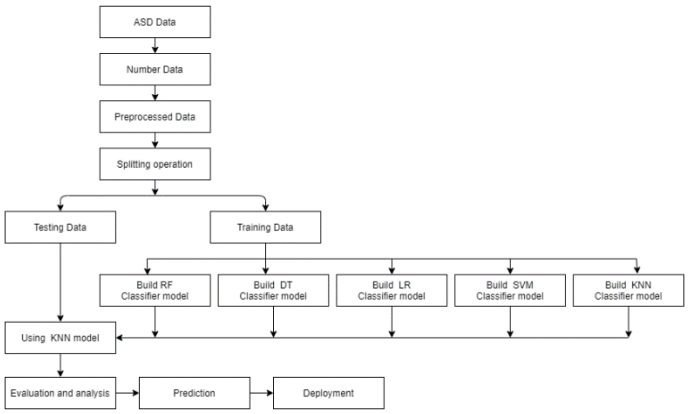

Fig 2. Activity Diagram

4. DESIGN AND IMPLEMENTATION

4.1 Technology

The main aim of this project is to develop a system for ASD Prediction which is able to help to characterize, search and predict for the likelihood of Autism affecting a toddler as young as 12 months old. The system developed hence will reduce the time taken to convey a proper diagnosis for the kid. Within the meantime, when the response of a positive test is calculated of the given user/ user's child, the user can drop any queries they need regarding the diagnosis or ASD normally can send their queries. Finally queries generated by user are stored within the database which can tend access to the medical professional only by consent of parents/caregivers.

The application uses various libraries of python – pandas, NumPy, scikit learn, pickle and using fundamental concepts of Machine Learning with Python 3.x for Data Preprocessing, Feature Engineering, Graphical Representations and also the use of Supervised Algorithm – KNN Classifier for training and testing of the dataset for the model. After the model file is made, we generate the Flask Application and webpages with the assistance of HTML5 and CSS, to produce with better Interface for user and User Experience. We also create 2 databases with SQLite3, one to store the main details of users while they signup and whereas the other to store queries of the users being generated at our FAQ page. The developed system is simple to run on any latest version of browsers like Google Chrome [7].

4.2 Algorithms

i) KNN (K Nearest Neighbors Classifier) –
KNN is a supervised machine learning algorithm. KNN or K-Nearest Neighbors is a Supervised Machine Learning Algorithm which stores all available cases and predicts the target supported on a similarity measure. KNN is often used for both Classification in addition to Regression problems. Classification algorithm used for predicting task when target variable is categorical and Regression algorithm used for predicting task when target variable is numeric.

Working:

Step 1: Select value of K-Neighbors. This will be done by taking square root of the overall number of rows of data or it can also be visualized by calculating error rate.

Step 2: Find K-Nearest Data points/ Neighbors supported by Euclidean Distance Let's consider points to be A1 and B2, then Euclidean Distance is often calculated as

$$\text{Euclidean Distance between } A_1 \text{ and } B_2 = \sqrt{(X_2-X_1)^2+(Y_2-Y_1)^2}$$

Fig. 3: Euclidean Distance Formula

Step 3: Combine the neighbors into a prediction for new data point Since our target class is categorical (yes, no), prediction will be the most occurrence(mode) value.

Step 4: Among the K neighbors, we have to count the number of the data points in each category.

Step 5: Next, we assign the new data points to that category for which the number of the neighbor is maximum.

Step 6: Our model is ready.

5. RESULTS AND DISCUSSION

Considering the dataset, we have used for model; we find the relationship amongst the various columns of the dataset trying to establish a relationship amongst them. On finding the most reliable data, we train and test the model on different algorithms to find precision, accuracy and recall score. Here we found that amongst Logistic Regression, KNN Classifier Model, Decision Tree Classifier, Random Forest Classifier and Support Vector Classifier – KNN Classifier gives the best results without any overfitting or underfitting issues.

We select KNN Classifier Model and find the error rate for best possible K-Neighbors to use and the classification report is also printed. They described in Fig. 4 and Fig. 5 respectively below. These metrics calculated are the decision factor as to why KNN Classifier Model was selected for this system has been described following.

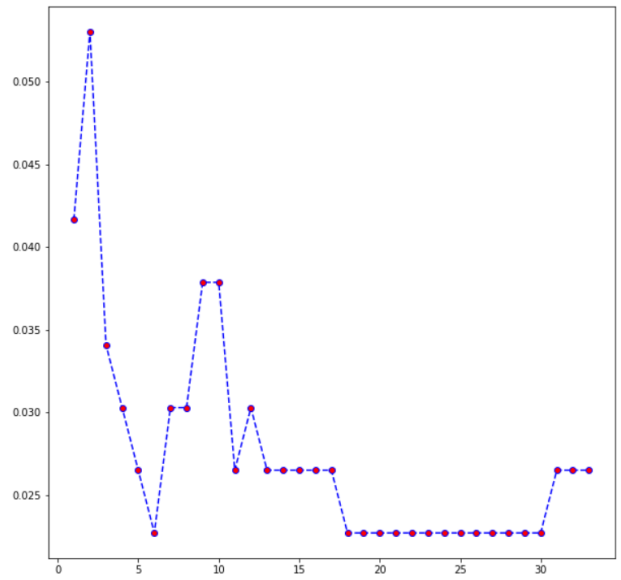

Fig. 4: Error Rate for Best Possible K-Neighbors in KNN

```
'              precision    recall  f1-score   support\n\n          No       0.94      0.99      0.96       76\n         Yes
0.99      0.97      0.98      188\n\n    accuracy                           0.98       264\n   macro avg       0.97      0.98
0.97       264\nweighted avg       0.98      0.98      0.98       264\n'
```

Fig. 5: Classification Report for KNN Classifier Model

As we can see in the Classification Report in Fig. 5, we have an accuracy of 99% in training data and an accuracy of 97% in testing data. These conditions negate both overfitting and underfitting issues and hence we get best possible and accurate outcome as target (output) from the features (input).

6. CONCLUSIONS

The proposed system helps in predicting ASD with the help of Machine Learning using Python 3.x and Machine Learning Supervised Learning Algorithm which is the KNN Classifier Model and furthermore pickle dumping the trained and tested model onto Flask 2.0, we have created a web application which is our system which can effectively predict the outcome of whether or not the given child – which can be between the ages of 12 months to 36 months is suspected of having Autism Spectrum Disorder and if or not will need medical and /or professional help. The Web Application is easily accessible and provided with user's comfort and security to handle sensitive data.

Considering the dataset, we have used for model; we find the relationship amongst the various columns of the dataset trying to establish a relationship amongst them. On finding the most reliable data, we train and test the model on different algorithms to find precision, accuracy and recall score. Here we found that amongst Logistic Regression, KNN Classifier Model, Decision Tree Classifier, Random Forest Classifier and Support Vector Classifier – KNN Classifier gives the best results without any overfitting or underfitting issues.

We select KNN Classifier Model and find the error rate for best possible K-Neighbors to use and the classification report is also printed. we have an accuracy of 99% in training data and an accuracy of 97% in testing data. These conditions negate both overfitting and underfitting issues and hence we get best possible and accurate outcome as target (output) from the features (input).

Adding more features to the web application, another page will be created wherein user gives input as to what their/ their child's score was for the questionnaire being asked previously. A set of extra questions will be asked to describe if they perceive any other symptoms to ones not listed in the home page or severity in any of the traits asked. This data after received will be stored into the database to be sent to a medical professional/practitioner and child's parents/caregivers will be given the contact details of the medical help. This web application will then be deployed onto the Internet using pythonanywhere.com site which allows hosting of Machine Learning with Flask models.

ACKNOWLEDGEMENT

We would like to acknowledge our college A. C. Patil College of Engineering, Principal, Head of Department and management for the guidance, encouragement and co-operation throughout the project.

REFERENCES

[1] Centers for Disease Control and Prevention - cdc.gov/ncbddd/autism/facts.html.

[2] Health Informatics Journal 2020, Vol. 26(1) 264–286 © The Author(s) 2019 Article reuse guidelines: sagepub.com/journals-permissions DOI: 10.1177/1460458218824711 journals.sagepub.com/home/jhi.

[3] Ben-Sasson A, Robins D, Yom-Tov E Risk Assessment for Parents Who Suspect Their Child Has Autism Spectrum Disorder: Machine Learning Approach J Med Internet Res 2018;20(4): e134 URL: https://www.jmir.org/2018/4/e134 DOI: 10.2196/jmir.9496.

[4] W. Liu, X. Yu, B. Raj, L. Yi, X. Zou and M. Li, "Efficient autism spectrum disorder prediction with eye movement: A machine learning framework," 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), 2015, pp. 649- 655, doi: 10.1109/ACII.2015.7344638.

[5] Devika Varshini, G., and R. Chinnaiyan. "Optimized Machine Learning Classification Approaches for Prediction of Autism Spectrum Disorder." Ann Autism Dev Disord. 2020; 1 (1) 1001 (2020).

[6] Templating with Jinja2 in Flask: Essentials – code.tutsplus.com

[7] https://en.wikipedia.org/wiki/Google_Chrome