## Information Visualization (18CSE301J)

ABHIPSA SAHOO (RA2011003010346)

January 2023

#### 1 What is Autism Spectrum Disorder?

Autism Spectrum Disorder (ASD) is a neuro-disorder in which a person has a lifelong effect on interaction and communication with others. Autism can be diagnosed at any stage in once life is said to be a behavioral disease" because in the first two years of life symptoms usually appear. According to the ASD problem starts with childhood and continues to keep going on into adolescence and adulthood. Propelled with the rise in use of machine learning techniques in the research dimensions of medical diagnosis, there is an attempt to explore the possibility to use Naive Bayes, Support Vector Machine, Logistic Regression, KNN, Neural Network and Convolutional Neural Network for predicting and analysis of ASD problems in a child, adolescents and adults. There is a wide variance in the range as well as the severity of its symptoms. A few common symptoms the individual faces are difficulties in communication, especially in social settings, obsessive interests, and mannerisms, which take a repetitive form. To identify ASD, an extensive examination is required. This also includes an extensive evaluation and a variety of assessments by psychologists for children and various certified professionals.

#### 2 Why is it important to analyze this data?

Detecting and treating Autism Spectrum Disorder in its early stages are extremely crucial as this helps to decrease or alleviate the symptoms to a certain extent, thus improving the overall quality of life for the individual. However, owing to the gaps between initial concern and diagnosis, a lot of valuable time is lost as this disorder remains undetected. Machine Learning methods would not only help to assess the risk for ASD in a quick and accurate manner, but are also essential to streamline the hole diagnosis process and help families access the much needed therapies faster. The economic impact of autism and the increase in the number of ASD cases across the world reveals an urgent need for the development of easily implemented and effective screening methods. Therefore, a time-efficient and accessible ASD screening is imminent to help health the number of ASD cases worldwide necessitates datasets related to behavioral traits. However, such datasets are rare making it difficult to perform through analyses to improve the efficiency, sensitivity, specificity and predictive accuracy of ASD Screening Process. Presently, very limited autism datasets associated with clinical or screening are available and most of them are genetic in nature. However, this dataset is related to autism screening of adults that contained 20 features to be utilized for further analysis especially in determining features including ten individuals characteristics that have proved to be effective in detecting the ASD cases from controls in behavior science.



#### **3** Dataset Source

There are many things to be considered while working with external or internal data. For example, make sure it is always coming from the primary source. One such is the University of California, Irvine Machine Learning Repository. Some research papers have been written using this dataset. It has been used to classify ASD problem using machine learning, the input utilizing a training dataset of cases and controls that have already been diagnosed.

#### 4 Attributes

Presently, very limited autism datasets associated with clinical or screening are available .Since the dataset intends to find a correlation between behavioral traits and Autism diagnosis, there are many attributes that come into play. Basic information such as age, gender, ethnicity are presented. Other vital information such as whether the person going through screening was born with jaundice or not is also available. It is said that babies born with jaundice were between 56 percent and 88 percent more likely to develop any kind of psychological development disorder and they were 67 percent more likely to develop infantile autism specifically, a study has reported. Thus it is definitely an attribute that can exhibit a correlation with autism. Another important piece of information included in the dataset, which is a boolean value to determine whether a family member has PDD or not. Pervasive Devlopmental Disorders are also known as Autism Spectrum Disorder (ASD), and are characterized by delays in the development of social and communication skills. For an individual, the risk of autism is increased 10 fold if a sibling has the diagnosis and about 2 fold if a cousin has the diagnosis. These may help counseling families with affected children. Autism Screening is typically a questionnare which needs to be filled out by the child or parent. The details of who the form is being filled out by is also present in the dataset. The dataset includes 10 questions that question behavioral traits that have been proven to be effective in detecting the ASD cases from controls in behaviour science. These questions are going to focus on cognitive skills, social skills and learning habits. These questions will be framed by me in order to display useful correlations between behavioral traits and Autism Screening. Possible questions are: Were the parents older . The Attributes can be depicted through the table below:

Attribute	Туре	Description
Age	Number	years
Gender	String	Male or Female
Ethnicity	String	List of common ethnicities in text format
Born with jaundice	Boolean (yes or no)	Whether the case was born with jaundice
Family member with PDD	Boolean (yes or no)	Whether any immediate family member has a PDD
Who is completing the test	String	Parent, self, caregiver, medical staff, clinician ,etc.
Country of residence	String	List of countries in text format
Used the screening app before	Boolean (yes or no)	Whether the user has used a screening app
Screening Method Type	Integer (0,1,2,3)	The type of screening methods chosen based on age category (0=toddler, 1=child, 2= adolescent, 3= adult)
Question 1 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 2 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 3 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 4 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 5 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 6 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 7 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 8 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 9 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 10 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Screening Score	Integer	The final score obtained based on the scoring algorithm of the screening method used. This was computed in an automated manner

### 5 Possible 10 questions:

1. Do they respond to facial gestures? (ex: smiling back)

- 2. Does your child remember words or forget them easily?
- 3. Does your child respond to their name?
- 4. Do they avoid eye contact?
- 5. Do they find it difficult coordinating movements?
- 6. Do they get upset at food textures and bedding textures?

7. Do the develop unusual routines and get upset when these routines are not followed?

8. Do they consistently prefer to play alone?

- 9. Do they seem indifferent to pain and not show expected reactions to it?
- 10. Do they start conversations in an age appropriate way?

#### 6 Conclusion

To sum it up, this dataset could be vital for ASD Screening and help us find correlations between behavioral traits and Autism. This data could be further used in Machine Learning to classify ASD. Thus, this dataset is vital and could really help individuals who could be on the autism spectrum.

# 7 Links

:

https://www.overleaf.com/5583377672wnbsjvhsfbkj https://archive.ics.uci.edu/ml/machine-learning-databases/00419/