Autism Spectrum Disorder Detection in Toddlers for Early Diagnosis Using Machine Learning

by

Shirajul Islam
16301112
Tahmina Akter
16101299
Sarah Zakir
16101310
Shareea Sabreen
16101092

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
October 2020

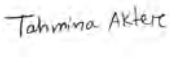
Declaration

It is hereby declared that

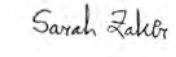
1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

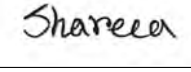
Shirajul Islam
16301112



Tahmina Akter
16101299



Sarah Zakir
16101310



Shareea Sabreen
16101092

Approval

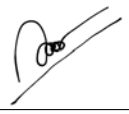
The thesis/project titled "Autism Spectrum Disorder Detection in Toddlers for Early Diagnosis Using Machine Learning" submitted by

1. Shirajul Islam (16301112)
2. Tahmina Akter (16101299)
3. Sarah Zakir (16101310)
4. Shareea Sabreen (16101092)

Of Summer, 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on October 9, 2020.

Examining Committee:

Supervisor:
(Member)



Dr. Mohammad Iqbal Hossain
Assistant Professor
Department of Computer Science and Engineering
BRAC University

Program Coordinator:
(Member)

Dr. Golam Rabiul Alam
Associate Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Dr. Mahbubul Alam Majumdar
Professor and Dean
Department of Computer Science and Engineering
Department of Mathematics and Natural Sciences
BRAC University

Abstract

Autism spectrum disorder (ASD) is a disorder where patients are unable to express and interact. Recently it is an issue to be concerned that one in 59 children has identified as an autism spectrum disorder patient. ASDs start from childhood but symptoms can be detected in adulthood. That is why these children are not being able to have proper treatment at an early age and that causes more complexity in their health. Research shows that a diagnosis of autism at an earlier age can be more reliable and stable. Therefore, our study aims to estimate ASD (autism spectrum disorder) at a sooner possible time and increase more accuracy than the previous research and reduce medical costs. In our thesis paper, we want to predict and distinguish between autistic and non-autistic children by using a machine learning approach. Firstly, we have gathered data from the surveillance side as much as possible. We also set some particular questions and try to find maximum accurate answers to all questions. Furthermore, supervised learning algorithms are applied to diagnosis whether children meet the symptoms for ASD. Among all applied algorithms KNN and Random Forest shows maximum accuracy and speed to diagnosis. Above all, our final goal is to create an online tool that can provide machine learning-based analysis to a user to detect autism at an early age precisely.

Keywords: ASD; Toddler; Machine learning; Prediction; Treatment; KNN; Random Forest Classifier

Dedication

We would like to dedicate this thesis to our loving parents and teachers.

Acknowledgement

First and foremost, praise and thanks to the Almighty that His blessings bestowed upon us throughout our research works. Our greatest and most sincere gratitude goes to our beloved and respected research supervisor Dr. Mohammad Iqbal Hossain for providing us the opportunity and valuable guidance throughout the research. Furthermore, his sincerity, purity, and liveliness inspired us extremely. without his proper advice, our thesis would not have turned out as we had expected it would be. He always challenged us to think critically and push ourselves beyond our limitations. We are so privileged that we work under his guidance. We are also indebted to our beloved BRAC University for always providing us the equipment, facilities, and resources that we needed during our research time. We would like to thank all the faculty members of the CSE department. we would not have been here today without their precious lessons. To sum up, we are tremendously grateful to our beloved parents for their love, support, and sacrifices they went through which inspire us to work harder and be ready for the future . Also, our love and appreciation go to our peers and classmates for their support and love in our tough times. Without their existence life would not have been smooth.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

ASD Autism Spectrum Disorder

AUC Area Under the Curve

GI Gastrointestinal

KNN K Nearest Neighbours

MLP Multiplier Perception

RBF Redial Basis Function

ROC Receiver Operating Characteristic

SVM Support Vector Machine

WHO World Health Organisation

Chapter 1

Introduction

Autism is also referred to as autism spectrum disorder (ASD) which is a neurodevelopmental disorder. Various impairments in social communication and interaction is caused by it. Also, the existence of unvaried patterns of behavior or activity. ASD appears in childhood and tends to persist into adolescence and adulthood wherein most stories the symptoms are apparent during the very first 5 years. Furthermore, it includes a few disorders that have been called autistic disorders namely pervasive developmental problems, Asperger syndrome, Rett syndrome, and fragile X syndrome. From research by the Center for Disease Control, it is found that an estimated 1 out of 54 children is affected by autism in the United States [5]. In the first year of life, the infants who are prone to get autistic spectrum disorders can show some interruptions. Which can be societal engagement, attention to anything, interaction, and also personality. These can happen before the onset of clinical signs and these can show the signs of early ASD development. Children's experience while engaging with other people in society can lead to many alarming signs. When the age of the child is 2, we can diagnose them by observing the developmental antecedents of autism symptoms. Then when the age becomes 3 the process of diagnosis matures to a stable state. The therapists can get many important info and data that can be helpful for the diagnosis process by doing the downward extension for diagnosing the problem [4]. Besides, various sensory sensitives can help to detect autism developments such as sleep disorders, mental health problems like anxiety, and depression. There are many medical issues too which can make the task a lot easier.

1.1 Motivation

ASD or autism spectrum disorder is a great issue in the whole world. By birth, the disorder begins to develop into a social and mental trauma for the person affected by it and the family of the person. Autism is a social communication and behavior disorder where along with the person the society suffers too. It is mainly a neurological and developmental disorder that starts at the very early stage of life and lasts throughout a person's life. Detecting autism earlier in one life can make a big difference than treating it later. Early autism detection is much needed to improve the behavioral and social communication of the toddler. If it is detected early the toddler can get improvement in his or her communication skill through therapy. From the age of 12 months to 18 months, symptoms start to show, and if detected

earlier and treated accordingly [3]. We aim to detect autism at an early age so that necessary steps can be taken to prevent it from getting worse. Early detection can help in not spending a lot in the future as it has been eliminating those situations such as developing social skills and so on. Treating ASD earlier is always the best option for toddlers as they are still developing.

In very young children with autism, the slow process is very difficult to identify. However, those who received early treatment did a remarkable improvement. A research was conducted where among 1293 children, 58 were evaluated. Out of them, 39 were diagnosed. The process was done using a modified checklist for autism toddlers which contained 23 yes or no items. There were differences found between diagnosed children who had autism and also the opposite. There were six items that were about social relatedness and communication [1]. So considering all the facts and existing research, we decided to take a step forward in this field so that we can lessen problems for the toddlers as much as possible. It is a big deal for our country as well as for the whole universe. Early detection of ASD might bring ease to it.

1.2 Problem Statement

An autistic patient faces several challenges like stress, depression, learning disabilities, movement challenges, difficulties with concentration, and many more. According to WHO every year among 160 children, one is diagnosed with ASD traits all over the world [12].The average prevalence of ASD in South Asia is also higher than previously. In the Bangladeshi perspective, 0.84 percent of children are suffering from Autism Spectrum Disorder (WHO, 2009). Against this huge burden, only 200 psychiatrists and limited professionals are serving. scientists of medical science have developed some screening methods to recognize possible autistic traits at preliminary phase. But a vital amount of time and cost is needed for this diagnosis. Though higher class parents can treat their children by detecting ASD early, most of the time middle class and lower class parents cannot be able to diagnose. However, a premature diagnosis of autism can be very helpful as the patient can have proper medication and treatment at an early age. It also prevents long term complexity and cost. Therefore an ASD detecting tool is required which will give maximum accurate results and will be time-efficient and inexpensive.

Figure 1.1 shows autism spectrum disorder a result and appearance among 255,265 peoples in total who live in 55,492 families of the six unions of Raiganj Upazila in Sirajganj district. The study chose all children whose age gap is between 18 months to 36 months. The obtained statistics was from the family database of the six unions monitoring systems. From the surveillance dataset 5600 children were recognized in total.

1.3 Research Objective

In the literature part, we have seen that measures for detecting the symptoms and working towards coming to a decision have, however, not been done often, and

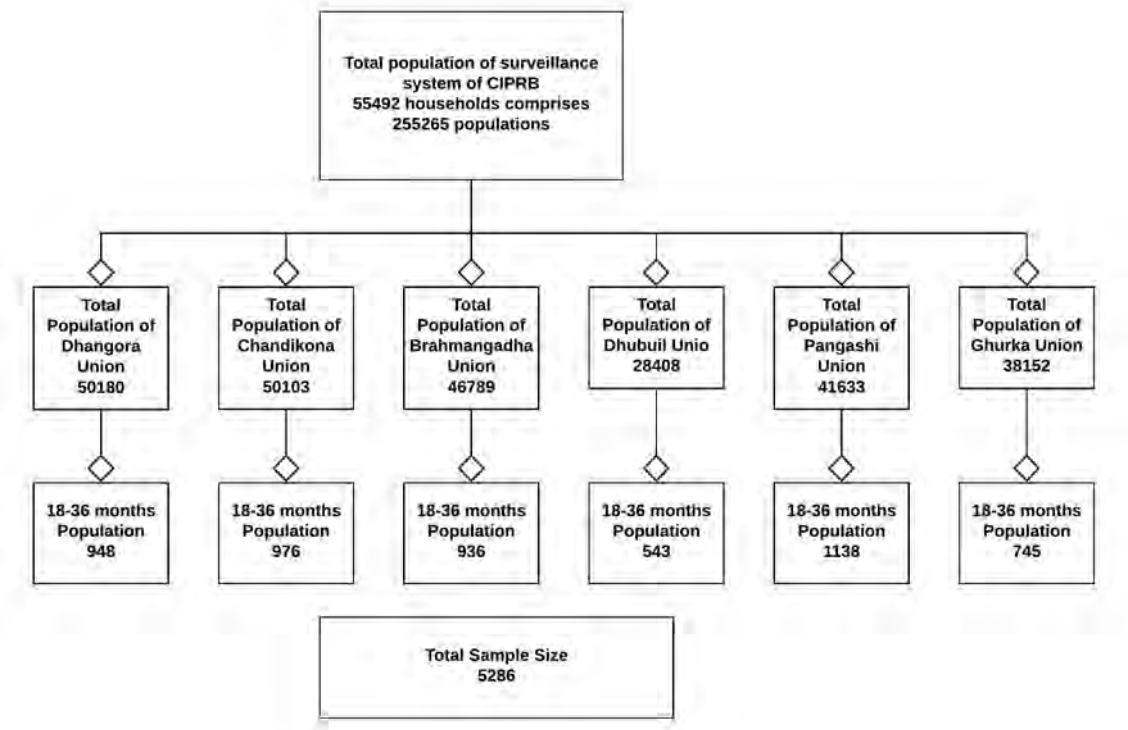


Figure 1.1: Prevalence of Autism Spectrum Disorder (ASD) among the children aged 18-36 months in a rural community of Bangladesh[11]

developing a decision of ASD has scarcely been done. Early detection can create a variation in the lives of an autistic child with autism spectrum disorder (ASD) as well as the parents where those children belong to [2]. Early intercession can improve a child's overall improvement which has been shown by many pieces of research. We should provide proper education and support to these children at their key development stages to obtain fundamental social skills and respond better to society. Essentially, by detecting ASD traits early the autistic child get the opportunity to improve himself or herself for a better life. Parents of autistic children can also discover soon and help their child to grow mentally, emotionally, and physically during the developmental period.

As doctors have to depend on observing the responses of toddlers as well as listening to the concerns of their parents so it is not easy to make an ASD diagnosis at all. That is why the objective of the work is to detect ASD symptoms at a premature age at minimum time and search for maximum accurate dataset to improve the accuracy of previous research and using maximum data. Besides, this work focuses on developing a model using supervised Machine Learning techniques. In the process, the machine will be trained by using data that is well specified which indicates that some data are already tagged with the accurate answer. After that, the user will provide the machine with a new set of examples so that the supervised algorithm critiques the training data(set of training examples) and gives the precise result of the specified data.

Another purpose of our research is to generate a mobile application so that anyone

can use them anywhere and detect whether their child is in the very beginning stage of autism. Some people don't want to go to the doctor or hospital because of the fear of society mostly in Bangladesh. So we thought of making a mobile application based on our model so that it becomes easier for them to test early and take care of their children. We will try to develop it in the future extension of our research.

1.4 Report Outline

The remaining section of the paper has been cataloged as follows. Chapter 2 presents related works and literature reviews loosely based on the dissertation. Moreover, in this section, some of the works done by different teams on predicting ASD on many predicting techniques were shown. That will give us insights about the existing research which will be helpful to describe why our research methodology is a better choice. Also, different types of algorithms that were used to develop our model is also described in this part.

Chapter 3 is basically about the discussion of our proposed model. In this part, the full description of our dataset was discussed and each of the variables was described so that it is understandable. Afterward, the data preprocessing part was presented because there were some modifications needed to test our classifier. In addition to that, this chapter also focuses on features selection and why those were selected. Lastly, this chapter illustrates the model description and how the model ensured accurate results.

In chapters 4 and 5, the implementation of the algorithms and which worked better for our model were discussed. After that, a comparison between the algorithms was shown by comparing different variables gathered from our model. A full representation of the evaluation of performance is recorded.

Finally, chapter 6 concludes with a discussion on some of the limitations we faced and might face. Also how further improvements can be done was explained too.

Chapter 2

Background

2.1 Literature Review

Autism or autism spectrum disorder (ASD) is a disorder in social behavior and communication. Detecting the preliminary symptoms at an early age is very important lest the treatment at a later stage should be more difficult and expensive. In addition, the symptoms of autism are mainly determined by the reaction of children to cognitive function and sometimes distinguishing among small non-autistic and autistic children are very complex. However, it can only be diagnosed by the observation of parents and doctors. In maximum cases, parents are uneducated and cannot show proper symptoms which causes incorrect treatment. Early detection of autism can help the patients with proper prescribing and medication. It can also prevent the condition of the patient from deteriorating further. Not only that, it can help to minimize the long-term cost due to the result of late diagnosis.

In this section we are going to show some of the works done by different teams on predicting ASD on many predicting techniques. For example, one research was conducted on children aged 4-11 , adolescents aged 12-16 years and adults begin from 18 to more , which holds 292, 104 and 704 occurrence one by one [16]. They were able to estimate autism with 92.26%, 93.78%, and 97.10% accuracy in case of child, adolescent and adult persons, one by one with using AQ-10 dataset. In terms of accuracy (77% to 85%) for actual dataset marginal performance showed this result. Because of this insufficient number of real datasets marginal data showed this kind of result. Moreover, merging the Random Forest-CART and Random Forest-ID3 algorithm provided most promising outcome compared to both the Random Forest-CART and Decision Tree-CART algorithm when applied on the dataset. However, the research was done on different age groups of people and did not focus on toddlers for early diagnosis. The work took place only on a small data set of 292 children which is not enough to gain the true sensitivity and specificity of the result.

In another paper the researcher has proposed a system based on machine learning and confabulation theory to effectively predict and diagnose ASD. The proposed framework mainly focuses on using machine learning techniques to reduce diagnosis time, increase the rate of accuracy and reduce input from the database. This paper used the CARMRMR algorithm which uses mRmR and confabulation theory. Using CARMRMR and Apriori is execution time 24.19 and 6.6 [15]. They have tested this

algorithm by 200 autistic children and the identifying rate through the algorithm is 91% but actual diagnosis through autism specialists is 80% [15]. The research done on 200 autistic children is not enough to determine more efficiency and accuracy.

On the other hand in a research paper they chose children with autism spectrum disorder aged 2-4 who are low functioning by a simple upper-limb movement to classify precisely. To differentiate 15 preschool children with ASD from 15 typically improving children they developed a machine-learning method which is supervised to correctly compare by method of kinematic analysis of an easy reach-to-drop task. Maximum classification accuracy of 96.7% was reached by this method with the goal-oriented part of the movement related to seven features [6]. To achieve this goal, on pre-school children with ASD in comparison to their mental-age-matched was applied by validated supervised machine-learning procedure to the kinematic analysis of a simple reach, grasp and drop task performance.

ASD can be long and cost-prohibitive if it is diagnosed in a clinic by specialists. To provide better accuracy, sensitivity, efficiency and specificity in diagnostic discovery Machine Learning provides techniques that can be automated classifiers which can be utilized by everyone. A new machine learning method called Rules-Machine Learning (RML) was introduced in this paper which uncovers autistic traits of cases and controls. Speciality of this is that it also proposes users knowledge bases (rules) that can be employed by domain experts in grasping the causes behind the classification RML proposes classifiers with higher predictive sensitivity, accuracy, specificity and harmonic mean than those of other ML approaches such as Boosting, Bagging, decision trees, and rule induction shows on empirical results on three dataset containing children,adolescent and adults. In this paper, they collected data by using ASDTests mobile application. RML is compared with PRISM, CART, AdaBoost, Bagging, Nnge, RIDOR, and RIPPER algorithms. Their datasets contain 704 cases for adults, 104 for adolescents and 292 for children. The result of accuracy was 92.5% to 95% for adults, 84% to 91% for children and 67% to 86% for adolescents. Datasets for adolescent and child were small and the accuracy which is not satisfying [17].

This paper [10] is about constructing a structure of procedure to find,record and label the patterns of behavior of children with Autism Spectrum Disorder. This structure works with 2 types of sensors. The first one is a wearable sensor; it detects behavioral patterns of a thing as it works with an accelerometer. The second one is static sensors that can capture photos and videos as it is based on a microphone and built-in camera sensor. The sensors are provided with a camera which is the proof of any occurrence and it will mark that part of the video where autism behaviour is shown. As it is a camera it will of course store the memory and the time of the video when autistic movement is detected. To spread out features, a Time-Frequency model is used and for analyzing the signal of the accelerometer Hidden Markov models have been used. Applying these techniques, they managed to get 91.5% of classification rate for behavioral patterns. It is good but the method they needed many resources and most importantly, it is complex. Also, it is not suitable to use in every field.

As existing technologies are expensive, time-intensive, this paper tried to use a low-cost and quick screening tool. They combined two screening methods. The first one is based on organized parent-reported questionnaires. The second one is in form of home videos of children [9]. Significant accuracy was collected in a clinical study where the trial is of 162 children. They also discussed the problem of extending machine learning algorithms to conditions beyond autism. The app Cognoa was used to develop their screeners. Patterns of behavioral keyed off of features which are probes by clinical instruments which are established were used to train their Screening algorithms. The instruments are ADI-R and ADOS. Gold standard instruments that have high accuracy are measured by these. The video-based screener, the questionnaire-based screener and both performances are shown in their ROC curve. Better specificity at 95% confidence level than both the M-CHAT and the CBCL was shown in screener on the combined and the video-based. Though having the same confidence level of 95% questionnaire alone has a higher specificity than CBCL but is not superior than the M-CHAT screener [9]. They used a sample of 162 children which is the largest drawback of this paper.

2.2 Algorithms

2.2.1 Random Forest

The algorithm for Random Forest is a supervised method for classification. Random forest algorithm works with the help of multiple decision trees. The distinction between the Random Forest and the decision tree is that the method of analyzing the root node and separating the component nodes would run randomly in the Random Forest.

A large number of selected decision trees that function as a unit are composed of random forests. A class prediction is thrown out by every single tree in the random forest and the class containing the maximum number of similarity is our model's prediction. This model involves two core concepts for which the method is called random, instead of merely measuring the prediction of trees.

Firstly, when trees are built random sequencing of training instances are created. Secondly, random subsets of functions are considered when dividing nodes. Each tree learns from a representative sample of data sources in a random forest when training. The samples, classified as bootstrapping, are developed with substitution, which means that such components can be used multiple times in a single tree. The principle is to train any tree on different samples, but with respect to a comparable collection of training samples, each tree may have a high variability. After the training the data is tested to see if it matches the prediction. During testing time, an average of the observations of each decision tree is created by predictions. This very method of training each individual case on various bootstrapped data samples and then averaging the observed values is called bagging [14].

There are more or less four main advantages of using the Random Forest algorithm. The most important one is for classification and regression tasks, it can be used. Overfitting is another crucial issue that can make the outcomes worse, but if there

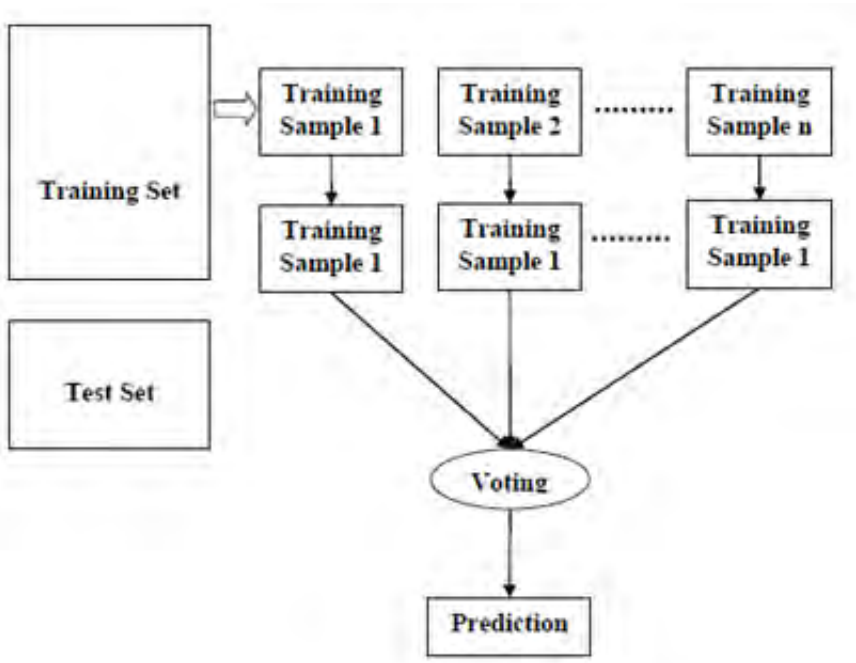


Figure 2.1: Random Forest Algorithm Model[19]

are enough trees in the forests for the Random Forest algorithm, the classification algorithm will not cause errors to the model. This model also has the capability to handle missing values. Furthermore, categorical values can be modeled on the Random Forest classifier [14].

In our research we used a random forest classification method as we are dealing with a huge number of dataset, from which we need to get insured on which class our final result of the dataset belongs to. As classifiers help to predict the destination class where the resultant value of the model will belong, it searches through predetermined lists or probabilities and finds the required best fitting class. Random forest classifiers do these predictions accurately because of the privilege it has over regression. For instance, Due to the number of decision trees involved in the process, random forests are considered a highly accurate and robust system [14]. It also takes all assumptions on average, which cancels the biases. In reality, classification trees work very well compared to regression in specifying specific classes. Thanks to their hierarchical structure, they are responsive to outliers, flexible and able to naturally form non-linear decision boundaries. As we need to know whether or not there is a probability of a child being positive with autism it is better to use Random forest classification method to determine if it will belong to the class as "positive" or "negative".

In the above diagram, there are X number of datasets containing independent class values. The Random forest classifier will go through the different classes and pick the repetitive values from them. It will examine the probability of a class getting a certain value and predict the majority of the values being repeated. This will help projecting the final class values from the given X number of datasets.

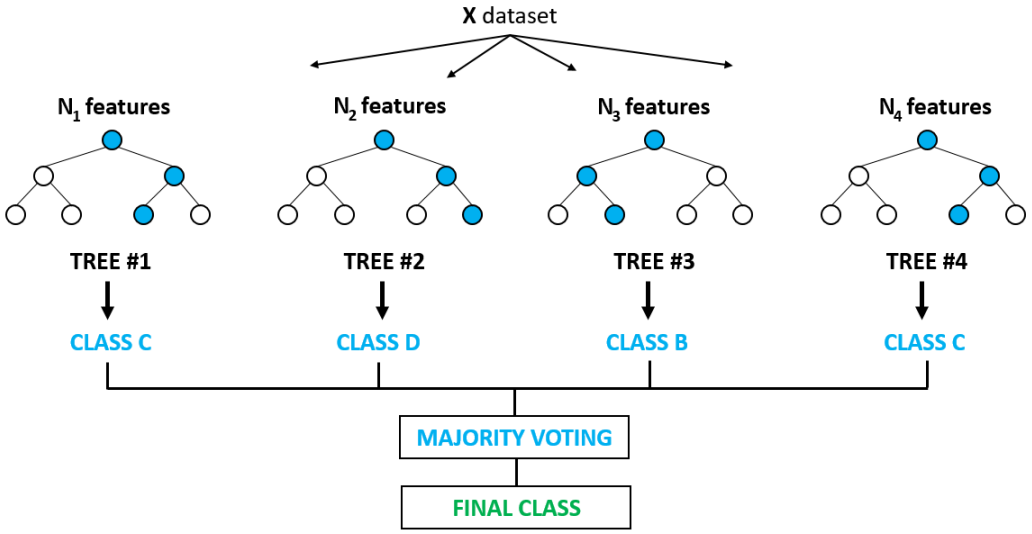


Figure 2.2: Random Forest Classification Model[13]

2.2.2 Kernel SVM

A support vector machine (SVM) is a supervised classification model that uses classification techniques in two groups for classification problems. In order to transform data, it uses a method originally called the kernel trick, and then finds an efficient threshold between the possible outputs based on those representations. After presenting an SVM model collection of identified training samples for each category, the two classes will interpret new knowledge. This machine learning method is a reliable classification algorithm that performs quite well with minimal volume of information. Since this paper deals with toddlers' autism detection data of not more than two thousand using Kernel SVM algorithm will be beneficial for the process.

A set of mathematical functions known as the kernel are used by SVM algorithms. The role of the kernel is to take data and translate it into the necessary form as an input. There are different types of Kernel approach. Among all the Kernel approaches features of Sigmoid kernel is very likely to give the optimal result. Sigmoid kernel also known as Hyperbolic tangent kernel can be said as MLP kernel. Here in our work we tried to use the sigmoid kernel SVM method to get the best-case result.

Kernel SVM is an approach where the intention is to find the most possible separation hyperplane that provides the highest marginal distance between the nearest points of the two groups. In general, this method ensures that the greater the margin is, the smaller the mean absolute error of the classifier is. When such a hyperplane exists, it is obvious that it offers the best boundary of separation between the two classes and is recognized as the maximum-margin hyperplane and the maximum-margin classifier.

The SVM kernel works by computing the distance into a high-dimensional feature space so that even if the variables can not be segregated linearly, data points can be categorized. There is a transition zone between the classes, so the data is inter-

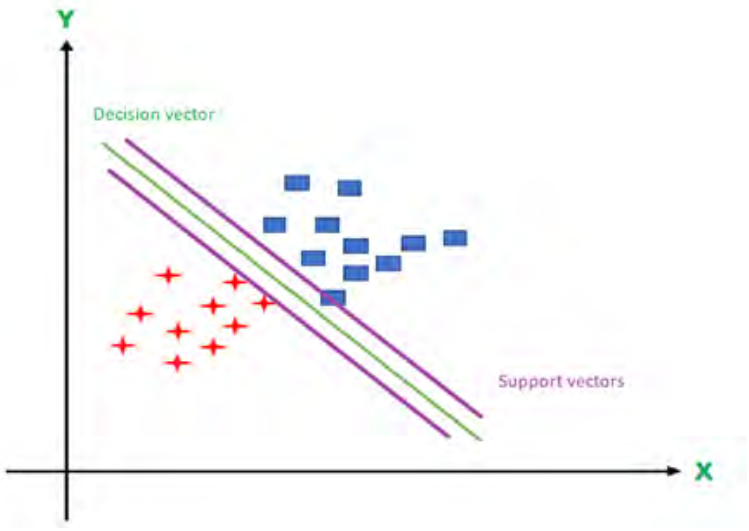


Figure 2.3: Support Vector Machine Model

preted and it is possible to draw the transformer as a hyperplane. In other words, the algorithm produces a suitable hyperplane in which growing trends are classified given the training samples. Hyperplane is a line that splits a region into two parts for two dimensional spaces where everything lies on either side of each row. This line (hyperplane) is the limit of judgment, we can identify everything that falls against one side of it as blue, and red for the points of the opposite side. The vectors which touch each side of the separate data on both sides of the hyperplane are called support vectors.

Another advantage of using Kernel Support vector machine classifiers is that it can work both on linear and non-linear data. In any event of linear data the separation of various classes of data can be done more easily by finding the hyperline between them, as shown in "figure-5". If there are multiple groups of data then from both groups we find the points nearest to the line according to the SVM algorithm. These points are known as support vectors. Now we measure the distance between the support vectors and the line which is known as the margin. The objective is to optimize the margin. The ideal hyperplane is the hyperplane where the margin is maximum.

2.2.3 Naive Bayes

The Naive Bayes algorithm is known as a classifier. It is mainly assumed independent amongst predictors. Though this classifier is a family of probabilistic classifiers, this model is simple to create and particularly advantageous for massive data sets. It also performs even higher in advanced classification methods. The algorithm relies on Bayes theorem and it is an equation that describes the relationship between the conditional probabilities of statistical quantities. The goal is to find the likelihood of a mark in the Bayesian classification, provided some marked features, which can be written as, $P(L|features)$. We can learn by the theorem of Bayes way to convey this in terms of quantities that can be measured more clearly:

$$P(L|features) = \frac{P(features|L)P(L)}{P(features)}$$

If the question is about choosing between L1 AND L2 labels, then one way to make this choice is to measure the posterior probability ratio for each mark:

$$\frac{P(L_1|features)}{P(L_2|features)} = \frac{P(features|L_1)P(L_1)}{P(features|L_2)P(L_2)}$$

All that is needed now is a model for each mark by which we can compute $P(features|Li)$. Since the conditional random process forming the defined data this model is called a generative model. The key part of the training is to define this generative model for every mark[20]. We can make the generative version of the training step simpler through naive assumptions, a rough estimation of the generative model can be found for each class and then get going with the classification of the Bayesian. Gaussian Naive Bayes classifier is the most suitable method for the data set size that is used in this paper. This model is adapted solely to get the mean and standard deviation of the points in each label, which is all needed to explain such a distribution[20]. We have a good formula to measure the probability of $P(features|L1)$ for any data point with this generative model in place for each class, so we can easily calculate the posterior ratio and evaluate which is most likely for a given point.

Bayes theorem provides a way of calculating posterior probability $P(c|x)$, $P(x|c)$ is likelihood, $P(x)$ is predictor prior probability, $P(c)$ is class prior probability. Here posterior probability will be the result of multiplication of likelihood and class probability which will get divided by predictor prior probability.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$$P(c|x) = P(x_1|c) \times P(x_2|c) \times ..... \times P(x_n|c) \times P(c)$$

Here:

- $P(c|x)$ = Posterior Probability
- $P(x|c)$ = Likelihood
- $P(c)$ = Class Prior Probability
- $P(x)$ = Predictor Prior Probability

In the first place, we need to train datasets and corresponding target variables. then we have to turn the dataset into a frequency table. By finding probabilities like overcast probability we have to create a likelihood table. Later, to compute the

posterior probability for each class the Naive Bayes equation is applied. Lastly, the result of prediction relies on the class with the highest posterior probability .

Naive Bayes theorem is straightforward and quick to anticipate a class of data sets. It also functions well in the categorical input variables compared to numerical variables. Naive Bayes classifiers can also be interpreted quite easily and have very few customizable parameters[20]. As we are working with a medium-size dataset Gaussian Naive Bayes classifier could give us a nearly perfect test result. Besides, they are extremely useful if naive assumptions match the data. Again for very well-separated categories, when model complexity is less important, which perfectly represents our dataset.

2.2.4 KNN

K-nearest neighbor algorithm is applied mainly to predict classifications puzzles. However, It can also be used for both regression and classification predictive problems. There are two properties for defining KNN algorithms namely Lazy learning algorithm and Non-parametric learning algorithm.KNN algorithm does not any specialized training phase and all trained data are used to classify so it is called a lazy algorithm. Besides, due to not assuming any underlying data KNN is also recognized as a non-parametric learning algorithm. It also applies "feature similarity" to anticipate the value of new data points which indicates that based on how closely it resembles the point in the training set new data will be allocated a value.

The KNN classifier uses some parameters when applied. It takes several parameter values to give the most optimal result. These parameters change according to the type of data that is used for the algorithm and also on the sort of output we want to receive. Though there are a number of parameters there are mainly two that are used frequently. they are, the n_neighbours and metric parameters. n_neighbors holds the number of nearest neighbors that the classifier would calculate [14]. On the other hand, matrices are important for choosing Euclidean distance. There are a number of metrics to choose from; each of the metrics acts differently and has a default p parameter which defines the power of the used metrics.

To implement the KNN algorithm at first we have taken a dataset and we have also placed the trained and examined data. Furthermore, the value of K is chosen. Thirdly, we estimate the interval between test data and each row of training data for each point in the analysis data. Here we can use any methods from Euclidean, Manhattan or Hamming distance. Among these tree methods we use the Euclidean method which is mostly used to calculate distance.

$$D(a,b) = \sqrt{\sum_{i=1}^{n}(b_i - a_i)^2}$$

Now depending on distance value we sort them in ascending order. Then, it chooses the highest k rows of the sorted array. Based on the most prevalent class of these

rows it assigns a class to the test point. We are assuming that we have a dataset

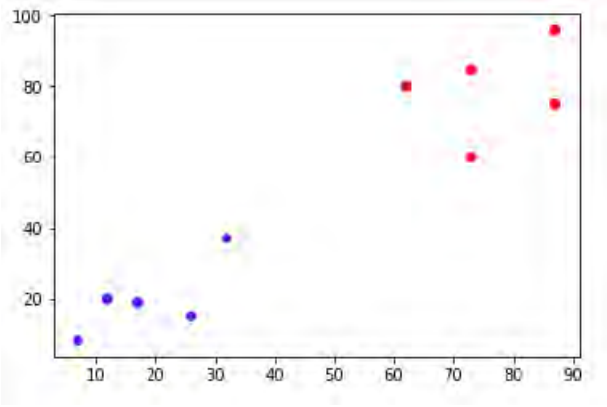


Figure 2.4: Plotted Dataset[7]

which has been plotted as in the figure 8. Now we have to classify a new data point for example here in the following figure 2.2 at point (60,60) a black point can be seen in between blue and red class. Suppose K=3. In figure 2.2 it is shown that three nearest data points with a black dot would be found. Here it is showed that among those three two of them reclines in the red class so the black dot will be specified in red class. KNN is a very simple algorithm that is helpful for nonlinear data as it

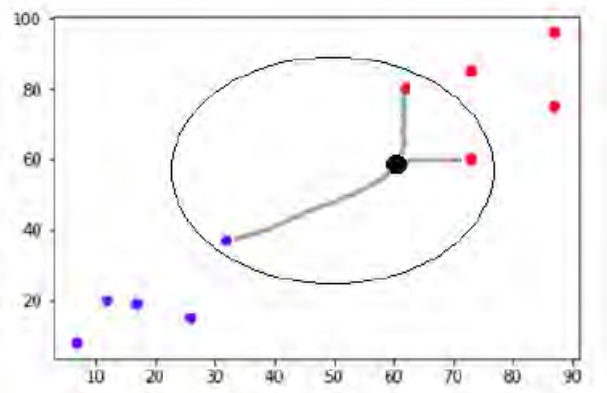


Figure 2.5: Finding K-nearest neighbours[7]

does not assume data. On the other hand, KNN comparatively bit expensive than other algorithms we have used because it stores all the training data so high memory accommodation is expected and prediction is slow in case data is larger.

Chapter 3

Proposed Model

3.1 Dataset Description

To apply supervised algorithms, we have used 1054 datasets. It has been categorized into three areas including medical, health and social science. Besides, the attribute type is categorical, continuous and binary. The queries of the Q-CHAT and AQ tools are 10. Each item value assigned from Kaggle and UCI ML repository [18]. The contained datasets are for toddlers and it represents 30.76% female and 69.24% male in toddlers. There are a total of 1054 toddler case values. In addition, the following ten questions were asked to the parent, self, caregiver and medical staff. The Q-CHAT questions possible answers are- "Always, Usually, Sometimes, Rarely and Never" which are selected as "0" or "1" in the dataset. If their reply was always /usually/ sometimes then "1" is allocated. Otherwise it is 0 for 10 (A10). Again for question 0-9 (A0-A9), the response was recorded as "1" for the answer Sometimes / Rarely / Never. If any child cuts more than 3(Q-CHAT-10-score) then the child will be detected with ASD otherwise no ASD traits are detected. Nevertheless, figure 3.1 and figure 3.2 describe the feature of data in different columns [18].

Figure 3.1 values were collected based on the Q-CHAT questionnaires. Figure 3.2 are the data that we will actually use to feed the ML algorithms to acquire the target result. Here, most of the data are boolean or binary type which will be fitting for the classifiers to compute and will not need preprocessing. Besides these there are also some integer and string type value fields which will need conversion before we use them in any classifiers. Otherwise we will not be able to reach the optimal result. The figure 3.2 contains the data type of each field and also the explanation of the features along with the data acquiring method.

3.1.1 Data Preprocessing

The acquired dataset needed some modifications before we could test our classifier algorithms on it. We preprocessed the dataset in such a way that it was able to provide prime output. Previously we found that unsorted and unprocessed data affected our result scores hugely. To get reed of the problems we followed some particular steps so the algorithms would be able to give more precise results. Most of our values in the dataset were binary and boolean values. They were based on polar questions primarily. However, few of the question criteria required non integer

Variable in Dataset	Corresponding Toddler Features
A1	Does your child look at you when you call his/her name?
A2	How easy is it for you to get the eye contact of your child?
A3	Does your child point to indicate that s/he wants something?
A4	Does your child point to share interest with you?
A5	Does your child pretend? (e.g. care for dolls, talk on toy phone)
A6	Does your child follow where you're looking?
A7	If you or someone else in the family is visibly upset, does your child shows signs
A8	Would you describe your child's first word?
A9	Does your child use simple gestures? (e.g. wave goodbye)
A10	Does your child stare at nothing with no apparent purpose?

Figure 3.1: Details of variables mapping to the Q-Chart-10 screening methods

and non boolean type answers. These data were recorded in string format. For our algorithm to give the optimal result we needed to convert these string type values to binary. We used one hot encoding to transform the values to binary. This technique was applied on only the values of the column "Ethnicity". One hot encoding is a method that converts categorical variables into a type which can be given to ML algorithms to do a better job of projection. For this process we first needed to switch the string values of "Ethnicity" column by default then applied one hot encoding on these values to get the unique binary correspondents. We also used Standard deviation in 'Age_mons' and 'Qchat-10-Score' columns. Lastly, we dropped two column values that indicated the case number and who completed the Q-CHAT questionnaires test as they were not needed for our analysis. After this we were able to read data for our experimentation.

Next we decided the ML algorithms which we will use for our experimentation based on our data size, the data features and the target set of results we are looking for. Initially we choose the following algorithms:

- Random forest Classifier.
- Kernel SVM.

Feature	Answer type	Description
A1: Question1 Answer	Binary (0,1)	The answer of the question based on the screening method used
A2: Question1 Answer	Binary (0,1)	The answer of the question based on the screening method used
A3: Question1 Answer	Binary (0,1)	The answer of the question based on the screening method used
A4: Question1 Answer	Binary (0,1)	The answer of the question based on the screening method used
A5: Question1 Answer	Binary (0,1)	The answer of the question based on the screening method used
A6: Question1 Answer	Binary (0,1)	The answer of the question based on the screening method used
A7: Question1 Answer	Binary (0,1)	The answer of the question based on the screening method used
A8: Question1 Answer	Binary (0,1)	The answer of the question based on the screening method used
A9: Question1 Answer	Binary (0,1)	The answer of the question based on the screening method used
A10: Question1 Answer	Binary (0,1)	The answer of the question based on the screening method used
Age	Number	Toddlers (months), children, adolescent, and adults(year)
Score by Q-chart-10	Number	1-10(less than or equal 3 no ASD traits > 3 ASD traits)
Sex	Character	Male or female
Ethnicity	String	List of common ethnicities in text format
Born with jaundice	Boolean (true, false)	Whether the case was born with jaundice
Family member with ASD history	Boolean (true, false)	Whether any immediate family member has a PDD
Who is completing the test	String	Parent, self, caregiver, medical staff, clinician etc.
Why are you taken the screening	String	Use input textbox
Class variable	String	ASD traits or No ASD traits (automatically assigned by the ASD Tests app). (yes/No)

Figure 3.2: Features collected and their descriptions

- Decision Tree Classifier.
- Gaussian Naive Bayes Classifier.
- K Nearest Neighbor.
- Logistic Regression.

For these algorithms we need to import the required libraries individually for each of the classifiers. Our goal by applying these classifiers was to find the accuracy,

precision and recall scores for which we imported the necessary libraries. For KNN, Random Forest and Naive Bayes classifiers we imported AdaBoost classifier and GradientBoosting classifier along with the other required ones. Some of the common libraries all of the libraries used were pandas, numpy, plt, plot ROC curve etc. Our experiments were done using Google colab notebook. We imported libraries and the preprocessed data on colab to compute our target result. At the beginning of the process we set the features of the data to X and the output to Y. We then splitted our data set into a test and train set, where 20% of the dataset were used for testing and 80% were used for training.

3.2 Model description

The proposed model ensures a more accurate result to the research about autism detection at an early age of autism. We used different machine learning algorithms to get more accurate results like the kernel Support Vector machine, Random Forest, Naive Bayes and KNN, Logistic Regression and Decision Tree. First, we collected the questions by following a standard. Those who are required to answer them we collected their answers and sent them to data preprocessing. Data preprocessing is adjusting if it is a non-matrix format type or not, how many attributes and instances are there and how many we need to run in a specific algorithm. In our case there are 1054 instances and 18 attributes including the class variables.

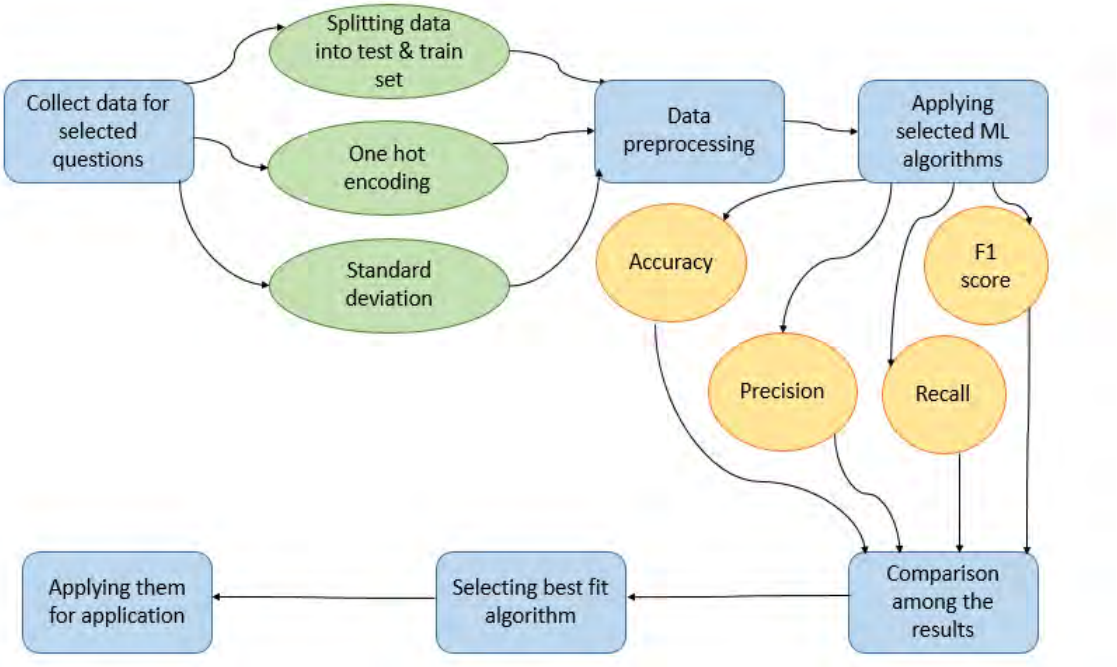


Figure 3.3: Workflow

We started working by collecting data of question sets to train our machine learning algorithm. For that we imported the necessary libraries. After that we allocated 80% of our dataset for training and 20% for testing. We have set the X matrix for feature and Y matrix for output. By using one hot encoding we have transformed a column "ethnicity" to unique binary values. For data preprocessing we have also

used standard deviation in “Age_mons” and “Qchat-10-Score” which transforms the value within the range of -3 to 3. After preprocessing we started applying our ML algorithms. We used six different supervised algorithms of which Logistic regression and Decision tree have been later excluded as they overfit our dataset. But the other algorithms resulted in a good way, they did not overfit or underfit as our dataset is less for the other algorithm. Accuracy of all the algorithms has been compared to each other for better results. Furthermore we found the score for precision and recall. Precision does not depend on accuracy as for precision the values are not always the same. After getting precision,accuracy and recall we compared the results and selected the best fit algorithm for our model. Lastly we apply them for the application of our model. Our purpose is to improve the accuracy of the outcome in relation to the other autism detection research. Besides this our second target is to get as much data possible to train our model. We have planned to offer a mobile application of this model in the future and for that we will be needing a fund sponsor.

Chapter 4

Experimentation

Firstly, to build a model the important thing is to find the goodness of a model. And the most valuable work is going to be how good the predictions are. In here, the following diagram shows the outcome of a model. This is called a ROC curve. In a ROC curve, it can be determined how an algorithm performs by observing the AUC.

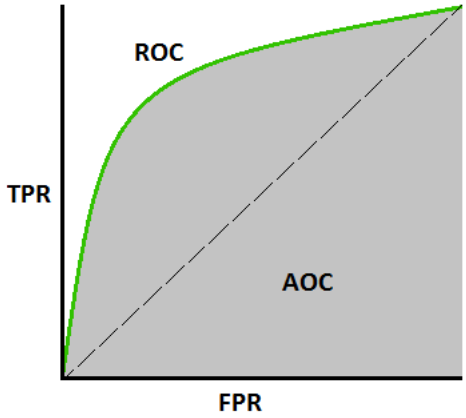


Figure 4.1: ROC curve example [14]

Also by using Confusion matrix, it can be determined of the goodness of a model. Now a confusion matrix which basically is for defining the results of a classification algorithm. Accuracy, Precision, Recall, F1 score and other measures except AUC can be calculated from the left most four values. And these four values/parameters are true negative, true positive, false negative and false positive which are the answers of class(output). In the following table the true positives and true negatives are shown in green color because these are correctly predicted. Also, false positives and false negatives are shown in red color because these are incorrectly predicted. These red color values are needed to be minimized. The details of these values are-

True Positive(TP): True positives are correctly found values. This means the genuine result and the anticipated result both are yes.

True Negative(TN): True negatives are also correctly found results. This means the genuine and anticipated result both are no.

	Predicted Class		
Actual Class		Class = Yes	Class = No
	Class = Yes	**True Positive**	**False Negative**
	Class = No	**False Positive**	**True Negative**

Figure 4.2: Confusion matrix parameters

False Positive(FP): This means the real result is no but the anticipated result is yes.

False Negative(FN): This means the real result is yes but anticipated result is no.

To identify how good a model has performed can be measured by some parameters namely Precision, Accuracy, F1 Score and Recall. By using these four values we can calculate Precision, Accuracy, F1 Score and Recall. These scores are:
Accuracy: Accuracy means how close the measured value is compared to the standard value. Proportion of the total number of predictions that are actually correct.

$$Accuracy = (TP + TN)/(TP + FP + FN + TN) \tag{4.1}$$

Precision: It indicates to the adjacency of two or more measurements to each other. If we take any measurement 3 times and every time we get the same value though the value is not close to the standard value then the measured value will be considered as precise but not accurate. So precision does not depend on accuracy.

$$Precision = TP/(TP + FP) \tag{4.2}$$

Recall(Sensitivity): Recall actually determines how many of the real positives a model captures by labeling it as true positive. So, recall is accurately anticipated positive inspection ratio to all the inspections in the real table.

$$Recall = TP/(TP + FN) \tag{4.3}$$

F1 Score: F1 takes both false negative and false positive in the count and takes a weighted average. F1 is usually more useful than accuracy.

$$F1Score = 2 * ((Precision * Recall)/(Precision + Recall)) \tag{4.4}$$

So, these parameters and ROC curve show the performance of a model. [8].
We have implemented four supervised machine learning classifier algorithms in this paper. These are K-NN, Naive Bayes, SVM and Random Forest. For every experiment we used 20% data for testing and 80% for training the model.

4.1 SVM Classifier Experiment:

Supporting vector machines is a different type of machine learning classifier that shows some different kind of result than other models. SVM is a model which works best for limited data processing. As we are dealing with less than two thousand data, this algorithm was quite fitting for the task. In our experiment of SVM, we basically used the kernel SVM that performed very well. Before some proper preprocessing and without using the kernel it showed almost 71-74% accuracy which was not so good. After preprocessing and by using the kernel it showed us 83% accuracy. We used gamma as 0.7 and C as 1.0 for parameters in applying the classifier. Then the result of its accuracy showed 83% almost. The precision score it showed is 89%. Moreover, 88% and 88% of Recall and F1 score. These scores can be seen in the following results table. The time it took while generating the result was 1.91 seconds which is quite higher than some other models.

4.2 Naive Bayes Classifier Experiment:

In Naive bayes experiment we used the Gaussian Naive bayes which showed some promising results and the results were better than Supporting Vector Machine classifier. We kept the parameters for Naive Bayes as default. We did not apply any different parameters for Naive bayes as it showed some good results in default parameters. It gave 89% of accuracy which is better than the first algorithm and the precision score is 100% which is impressive. Also, Recall score was 84% which is not so good as before and the F1 score it showed is 91%. It took 1.53 seconds to complete the experiment and to produce the result which is the lowest among all other algorithms.

4.3 Random Forest Classifier Experiment:

The third algorithm we used is Random Forest algorithm which performed as one of the best performers. We kept parameters of random forest as best as possible. Then the results it showed are better than the first two algorithms. It showed 93% of accuracy which is the second best result we have got. Precision was 92% which is good enough also. Then the recall and F1 score was 100% and 96% which is also better than others. All the results of this algorithm were good except one thing and that is the time it took was higher than any other algorithms. 2.30 seconds is the highest time it took than other algorithms.

4.4 K-NN Classifier Experiment:

This algorithm was the last successful implementation of our experiments and it showed the best result among all other algorithms. We used parameters of n_neighbors as 5, metric as 'minkowski' and p equal to 2. It performed in this metric minkowski is better than others. The accuracy it showed is 98% which is the finest among all results we got. The scores of Precision, Recall and F1 score as 100%, 97% and 99%. All are good scores as they had higher results than others. The Precision score is better than some others. And Precision indicates to the adjacency of two or more

measurements to each other. Also, high precision rate means low false positive rate which means the score is high and this is good. Moreover, it should be mentioned that this algorithm did a good time management like it took only 1.55 seconds to complete the calculation.

The result of applying the algorithms are given below:

Method	Accuracy	Precision	Recall	F1 Score	Time
SVM	0.83	0.89	0.88	0.88	1.91 sec
Naive Bayes	0.89	1.0	0.84	0.91	1.53 sec
Random Forest	0.93	0.92	1.0	0.96	2.30 sec
K-NN	0.98	1.0	0.97	0.99	1.55 sec

Table 4.1: Results after applying Classifier algorithms

Chapter 5

Result Analysis

Firstly, to obtain a good result from a dataset we had to do some preprocessing that is necessary for a machine learning model to perform well. Also, in machine learning we should do some proper data preprocessing for future proofing that means the acceptability of models and to perform well depends on preprocessing also. So, what we did in preprocessing is, we kept all of the values of features in binary except three features values. In those three features one was "Ethnicity" and it was in string type. For the string type data we did encoding which is named one hot encoding to transform the string type data to a binary value but unique. We could have used only simple numerical values but numerical values for 11 types of ethnicity will not give better results after applying algorithms. Now, one hot encoding is different from simple binary code because it represents a value as a unique value and there is no dependency or serial like numerical values. And for the other two datas, "Age_Mons" and "Qchat-10-Score" we did standard deviation that transformed the datas to the range of -3 to 3 which were previously numerical values without any range. These data preprocessing helped really well in algorithms performance of obtaining better results. For instance, Support Vector Machine gave accuracy results of 71-73% highest without the preprocessing of datas and when the values of "Ethnicity" were converted into numerical values only. But after applying preprocessing the algorithm "SVM" did a better job like the accuracy increased by 10%.So, that is why we used these preprocessing materials for all the algorithms.

Secondly, we experimented a total of six algorithms on our dataset. The algorithms were Random Forest, SVM, K-NN, Naive Bayes, Logistic Regression and Decision Tree classifier. Every one of them were classifier algorithms. From these six algorithms, two of the algorithms were named Logistic Regression and Decision Tree classifier which resulted in overfit of data. So, we dropped these two algorithms and kept the best performed algorithms in our experiments. From the rest 4 algorithms K-NN showed us the most promising result and took one the lowest possible time while calculating compared to others. Then the second best performed classifier algorithm was Random Forest Classifier. In algorithms we chose K-NN because closeness of features is the main basement of K-NN. A data point is classified by it based on how classification of its neighbours is done. It works better on data which are not complex and our dataset is complication free(noise free) as we preprocessed the data. Also, it works well in small dataset. K-NN normally calculates the euclidean distance of unknown data points from all the points to find the near-

est neighbours values. Then by default it takes 5 neighbours values of euclidean distance to classify. We used 33 as neighbours value because that is the square root of our total data. Square root of total data is one of the best to take as the nearest neighbour at k value. And the 33 nearest values of euclidean distance did help to classify that a child has autism or not. So these are the reasons why K-NN performed really well in our model.

Thirdly, the second best result given algorithm is Random Forest and we used it because Random Forest shows no overfitting of data in result. By using multiple trees it reduces the risk of overfitting. Also, it takes less training time though it did the opposite in terms of timing in our case. Random forest method operates by constructing multiple decision trees when it is in the training phase. And from those decision trees, random forest takes the majority voted result and that is why it can give a higher accuracy in results.

After that, the Naive Bayes classifier has given the third good result which is 89% accuracy. Naive Bayes classifier works on the basis of contingent probability from the Theorem of Bayes. It is very easy to implement and simple. It is not sensitive to irrelevant features and that is what we found it would have given good results even if we had irrelevant features. It needs less training data and we chose it as we have less amount of data which is less than two thousand. Moreover, it is fast and it takes 1.53 seconds of time to train and calculate the results which is the lowest compared to other algorithms. But it gave less accuracy than the Random Forest and K-NN classifier because those algorithms could obtain better results as those have got some advantages in our dataset.

In the case of Support Vector Machine, it gave good results but the lowest among other algorithms. Support Vector Machine creates a decision boundary by dividing the data into two categories. In splitting the data, the boundary should be in a way where maximum space is found that separates the two classes. In our case the Support Vector Machine could not place the boundary in the best place where it separates the two classes in maximum space means the support vector and the hyper-plane was not as far as possible, that is why it showed less better results than others.

The ROC curve in the following figure demonstrates the true positive rate versus false positive rate curve of algorithms. Here we can see that the Support Vector Machine has the lowest area under the curve and that is why the orange curve looks like this. And the other curves show that the GaussianNB is a little bit less accurate than the K-NN and Random forest classifier algorithms. So, It means the curve demonstrates Random Forest and K-NN is showing better performance than others.

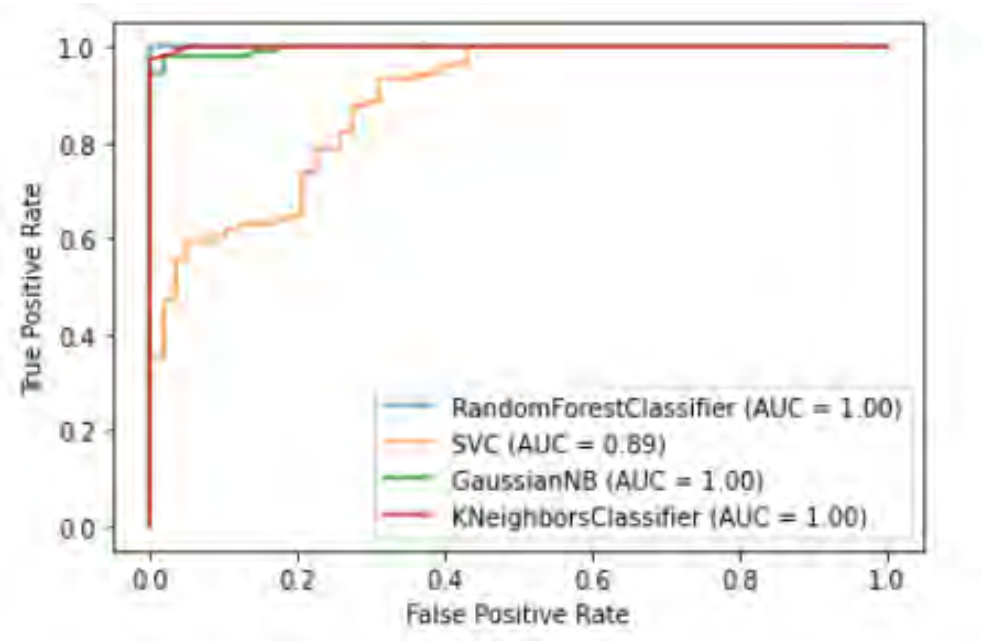


Figure 5.1: ROC curve comparing Random Forest Classifier with KNN, Gaussian NB and kernel SVM.

Chapter 6

Conclusion

6.1 Conclusion and Future work

Our proposed model allows us to have a more accurate result in terms of detecting autism at an early age. Questions which are provided to the parents to identify if their children are in danger or out of danger is set in a way to maintain their privacy. Using the dataset from Q-CHAT and AQ tools our proposed model can predict using Kernel SVM, Random Forest, Naive Bayes and KNN with 83%, 93% ,89% and 98% accuracy in case of toddlers. Algorithms which are supervised are selected to run our dataset after preprocessing it. We used Kernel SVM, Random Forest, Naive Bayes and KNN to get our output more accurately. This outcome showed better performance compared to the others. Our result showed marginal performance in terms of accuracy (93%, 98%). The only limitation of our model is the lack of enough large data to train our model.

Previously, many tried to detect autism at different ranges of age but we tried to emphasize on early ASD detection. The purpose of choosing such an age limitation in our model is to get as accurate results as possible. With the help of more accurate results and more data to train our model we can get tremendous work done. Many countries are struggling to detect autism as early as possible but with our model and the set of questionnaires we collected the problem can be solved effortlessly. Our objective for future work is to collect as much possible data from various sources and enhance the accuracy more. Moreover, we are thinking of building a user-friendly mobile application for end users based on our proposed model so that any individual can use the application to predict the autism test effortlessly and efficiently. Since diagnosing autism is quite a costly and lengthy process, it has been postponed for countless children. To conclude, with the help of our proposed model individuals can be guided at a very early age that will limit the situation from getting any worse and reduce costs associated with delayed diagnosis.

The limitations on the characteristics of design or methodology that impacted or influenced the interpretation of our thesis research, the prime candidate is the lack of dataset. While implementing two algorithms we found that because of the lack of our dataset our result was overfitting. This modeling error occurred because a function corresponds too closely to a particular set of data. As a result, it failed to fit additional data and affected the accuracy of predicting future observations and

we had to drop those two algorithms. In addition, it was not helpful even after preprocessing our data.

Bibliography

[1] D. L. Robins, D. Fein, M. L. Barton, and J. A. Green, "The modified checklist for autism in toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders", *Journal of autism and developmental disorders*, vol. 31, no. 2, pp. 131–144, 2001.

[2] S. Maestro, F. Muratori, M. C. Cavallaro, F. Pei, D. Stern, B. Golse, and F. Palacio-Espasa, "Attentional skills during the first 6 months of age in autism spectrum disorder", *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 41, no. 10, pp. 1239–1245, 2002.

[3] S. E. Bryson, L. Zwaigenbaum, and W. Roberts, "The early detection of autism in clinical practice", *Paediatrics & child health*, vol. 9, no. 4, pp. 219–221, 2004.

[4] S. J. Webb and E. J. Jones, "Early identification of autism: Early characteristics, onset of symptoms, and diagnostic stability", *Infants and Young Children*, vol. 22, no. 2, p. 100, 2009.

[5] J. Baio, "Prevalence of autism spectrum disorders: Autism and developmental disabilities monitoring network, 14 sites, united states, 2008. morbidity and mortality weekly report. surveillance summaries. volume 61, number 3.", *Centers for Disease Control and Prevention*, 2012.

[6] A. Crippa, C. Salvatore, P. Perego, S. Forti, M. Nobile, M. Molteni, and I. Castiglioni, "Use of machine learning to identify children with autism and their motor abnormalities", *Journal of autism and developmental disorders*, vol. 45, no. 7, pp. 2146–2156, 2015.

[7] T. Point, *Knn algorithm - finding nearest neighbors*, 2016. [Online]. Available: https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_knn_algorithm_finding_nearest_neighbors.htm.

[8] E. Solutions and *. Name, *Accuracy, precision, recall amp; f1 score: Interpretation of performance measures*, Nov. 2016. [Online]. Available: https://blog.exsilio.com/all/accuracy-precision-recall-f1.

[9] H. Abbas, F. Garberson, E. Glover, and D. P. Wall, "Machine learning for early detection of autism (and other conditions) using a parental questionnaire and home video screening", in *2017 IEEE International Conference on Big Data (Big Data)*, IEEE, 2017, pp. 3558–3561.

[10] C.-H. Min, "Automatic detection and labeling of self-stimulatory behavioral patterns in children with autism spectrum disorder", in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2017, pp. 279–282.

[11] S. Akhter, A. E. Hussain, J. Shefa, G. K. Kundu, F. Rahman, and A. Biswas, "Prevalence of autism spectrum disorder (asd) among the children aged 18-36 months in a rural community of bangladesh: A cross sectional study", *F1000Research*, vol. 7, 2018.

[12] P. Mesa-Gresa, H. Gil-Gómez, J.-A. Lozano-Quilis, and J.-A. Gil-Gómez, "Effectiveness of virtual reality for children and adolescents with autism spectrum disorder: An evidence-based systematic review", *Sensors*, vol. 18, no. 8, p. 2486, 2018.

[13] A. R, *Applying random forest (classification) - machine learning algorithm from scratch with real...* Jul. 2018. [Online]. Available: https://medium.com/@ar.ingenious/applying-random-forest-classification-machine-learning-algorithm-from-scratch-with-real-24ff198a1c57.

[14] N. DONGES, "A complete guide to the random forest algorithm", *Built In*, vol. 16, 2019.

[15] S. R. Dutta, S. Datta, and M. Roy, "Using cogency and machine learning for autism detection from a preliminary symptom", in *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, IEEE, 2019, pp. 331–336.

[16] K. S. Omar, P. Mondal, N. S. Khan, M. R. K. Rizvi, and M. N. Islam, "A machine learning approach to predict autism spectrum disorder", in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, IEEE, 2019, pp. 1–6.

[17] F. Thabtah and D. Peebles, "A new machine learning model based on induction of rules for autism detection", *Health informatics journal*, vol. 26, no. 1, pp. 264–286, 2020.

[18] ——, "A new machine learning model based on induction of rules for autism detection", *Health informatics journal*, vol. 26, no. 1, pp. 264–286, 2020.

[19] N. Donges, *A complete guide to the random forest algorithm.* [Online]. Available: https://builtin.com/data-science/random-forest-algorithm.

[20] J. VanderPlas, *In depth: Naive bayes classification.* [Online]. Available: https://jakevdp.github.io/PythonDataScienceHandbook/05.05-naive-bayes.html.