**ABSTRACT**

One of the primary psychiatric disorders is Autistic Spectrum Disorder (ASD). ASD is a mental disorder that limits the use of linguistic, communicative, cognitive, skills as well as social skills and abilities. Recently, ASD has been studied in the behavioural sciences using intelligent methods based around machine learning to speed up the screening time or to improve sensitivity,specificity or accuracy of the diagnosis process. Machine learning considers the ASD diagnosis problem as a classification task in which predictive models are built based on historical cases and controls. These models are supposed to be plugged into a screening tool to accomplish one or more of the aforementioned goals. In this paper, we shed light on recent studies that employ machine learning in ASD classification in order to discuss their pros and cons. Moreover, we highlight a noticeable problem associated with current ASD screening tools; the reliability of these tools using the DSM-IV rather than the DSM-5 manual.Hence the necessity to amend current screening tools to reflect the new imposed criteria of ASD classification with naïve bayes particularly the diagnostic algorithms embedded within these methods.

The purpose of this study is to evaluate the accuracy of two machine learning classification algorithms, Decision Tree and Random Forest, when attempting to classify individuals who may be likely to suffer from Autism Spectrum Disorder.

This study details the data cleaning and pre-processing steps taken to prepare the data-set for classification such as dealing with missing values, outlier removal, variable selection, and the partitioning of the data into training and testing subsets. It also walks through the validation and evaluation methods used when choosing the most appropriate models for each algorithm.

Evaluation of the predictions of the two classification algorithms used indicates that while both algorithms predicted the binary target moderately well, the Random Forest model was significantly more accurate than the Decision Tree model, resulting in less false negative and false positive predictions.

**INTRODUCTION**

In recent times, the application of Machine Learning to cross-disciplinary subjects has been very active and successful, especially in the fields of biology and neurology. Many researchers are interested in creating computational frameworks for automatically generating patterns and trends in large medical data-sets. A learned data representation can help visualise data to assist humans in clinical decision making and predict a target variable from a set of input features (Bone, Goodwin, Black, Lee, Audhkhasi and Narayanan, 2014).

The data selected for this project is the Autistic Spectrum Disorder (ASD) screening data for adults. ASD refers to several related disorders that normally begin in childhood and continue in adulthood. There is no cure for ASD, but treatments can help to improve symptoms. As per HSE (2017), the symptoms can include:

* Social interaction where it is difficult to understand situations and other people’s feelings and emotions.
* Difficulties to communicate, which can involve delayed language development, also not being able to take part in conversations properly.
* Unusual physical behavior such as doing repetitive physical movements, which becomes a routine, then the behavior becomes routine and the individual can get upset if the routine is disrupted.

The ASD symptoms can vary from person to person, and it can classify in three main types. The most typical type is “autistic disorder”, followed by “Asperger syndrome” and “pervasive developmental disorder” (PDD). The third one is also known as ‘atypical autism’. ASD are estimated to affect 1 in every 100 children and boys are more likely to develop ASD than girls by four times (HSE, 2017).

Researches and studies about classification on data related Autism have been conducted mainly by clinical experts and data scientists. A few ASD studies have analysed functional connectivity MRI(fcMRI) scan data to classify whether a data-set is coming from ASD or a typically developing participant solely based on functional connectivity (Chen, Keown, Jahedi, Nair, Plieger, Bailey and Muller, 2015).

Another stream is to develop diagnostic algorithms using machine learning using human behaviour data. The Autism Diagnostic Interview-Revised and the Autism Diagnostic Observation Schedule proved certain level of usefulness of objective machine learning methods for diagnosing autism (Lord, Risi, Lambrecht, Cook, Leventhal, DiLavore, Pickles, M. and Rutter, M., 2000).

To enhance ASD diagnosis accuracy, scholars recently adopted machine learning techniques, i.e. [8, 9, 25, 26]. The main goals of these studies were one or more of the following:

Reducing the screening time

Improving sensitivity and specificity

Identifying the smallest number of ASD codes to simplify the problem

Machine learning methods offer automated efficient and effective classification models for the ASD problem since they adopt a mixture of mathematical and search methods from computer science [24]. Various different machine learning techniques have been recently applied by researchers to the ASD problem, e.g.decision trees [22], support vector machine [, rule classifiers and neural network . ASD diagnosis is considered a typical classification problem in machine learning in which a model is constructed based on previously classified cases and controls. This model can then be employed to guess the new case diagnosis type(ASD, No-ASD).

Recent studies on machine learning in ASD to critically assess improvements in these studies especially the development of new machine learning methods for automatic ASD classification. We show recent results and challenges when machine learning is adopted for ASD classification which future studies can consider in order to improve the quality of the outcome. We believe that machine learning will be the next era in screening tools in which handcrafted classification methods will be replaced with automated predictive models. These models will guide clinical experts with fast yet accurate diagnosis decision.

**LITERATURE REVIEW**

**[3.1] Paper 1**

**Machine Learning for early detection of autism using parental questionnaire and home video screening**

**Aim:**

The aim of this paper is to create a low-cost, quick, and easy to apply autism screening tool that performs better than most widely used standardized instruments. This new tool combines two screening methods into a single assessment, one based on short, structured parent reported questionnaires and the other based on tagging key behaviours from short, semi-structured home videos of children. We further discuss the challenge of text ending machine learning algorithms to conditions beyond autism, and we propose a generalized framework for using machine learning algorithms to simultaneously search for the presence of many different conditions.

**Methodology:**

**Video screener methodology:**

The video screener keys on behavioural patterns typically probed in another diagnostic tool, the Autism Diagnostic Observation Schedule ADOS. ADOS consists of an interactive, highly standardized examination of the child by trained clinicians in a tightly controlled setting. ADOS is widely considered a gold standard and is one of the most common behavioural instruments used to aid in diagnosis of autism, however the cost and time to administer it can be prohibitive. For the Cognoa video screening algorithm, a subset of ADOS questions were identiﬁed as probing features that can realistically be observed in home videos. Separate algorithms were trained for use with pre-verbal children or verbal children, and for each an optimal subset of ten questions were identiﬁed as the most effective for identiﬁcation of autism in that age group. Questions based on these features were answered by a minimally trained analyst after watching two or three one minute home videos of the child’s behaviour taken by their parent. A variety of machine learning approaches were studied before settling on an optimal approach for the video screener. Random forests were then trained to determine the screener output. In order to reduce the impact of bias from non-observable features, missing observations were randomly injected into the training data at a rate that was calculated to minimize the decision making impact of a missing feature in the trees of the forests.

**Parental questionnaire screener methodology**:

This clinical tool consists of a parent interview of 93 multi-part questions with multiple choice and numeric responses which are delivered by a trained professional in a clinical setting. While this instrument is considered a gold standard and gives consistent results across examiners, the cost and time to administer it can be prohibitive. These questions are therefore keyed on by machine learning algorithms when building the parental questionnaire screener for the Cognoa application.

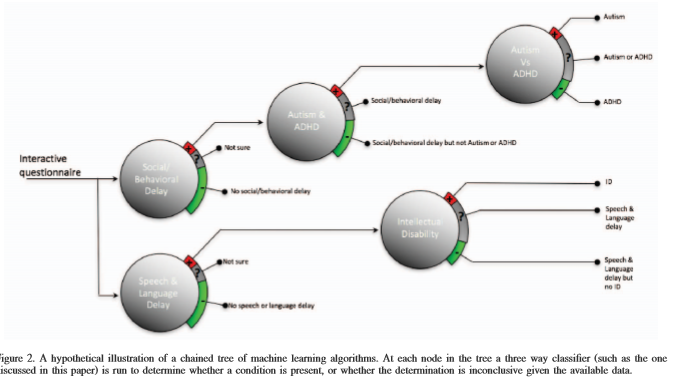
Two machine learning algorithms using random forests were designed, one to screen children between 18 months and three years of age, and the other to screen children between four and six years of age. To make the screening questionnaire easier for parents, the number of questions asked was reduced. From each age group, detailed studies were performed to identify an optimal subset of questions to include for the ﬁnal machine learning algorithm. Questions similar to these, but simpliﬁed and rephrased in order to be easily understood by parents were chosen to be presented in the Cognoa application. Aggregations of the average, extremes, and most common responses were found to signiﬁcantly increase accuracy

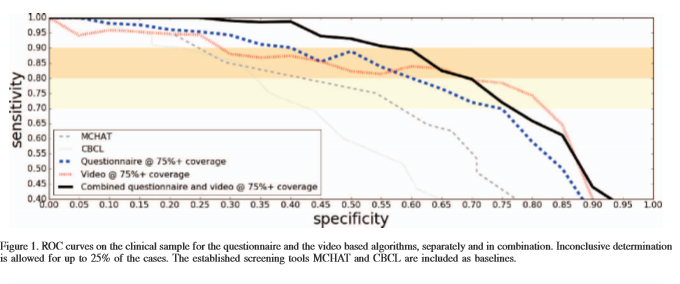
**Autism screening combination:**

The combination of the questionnaire and video screeners is made more challenging by the fact that no training samples are available for children that are known to have both ADI-R and ADOS results. As a consequence, the clinical data itself must be used to build the model to perform the combination. The numerical response of each of the parent questionnaire and video classiﬁers were combined using l2 regularized logistic regression. While some overﬁtting is expected, this is minimal due to the fact that the logistic regression is highly constrained, with only three free parameters. Since screening models were trained for young children and old children, separate combination algorithms were trained per age group. For each combination algorithm, optimal inconclusive output criteria were chosen using the logistic regression response.

**Results:**

ROC curves show the performances of the parental questionnaire-based screener, the video-based screener and the combined screener. The performances of the industry-standard M-CHAT and CBCL autism screeners are also compared. Operating near the commonly used threshold of 80% sensitivity, the combined and video-based screeners presented here have a better speciﬁcity than both the M-CHAT or the CBCL at a 95% conﬁdence level, while the questionnaire alone has a better speciﬁcity than the CBCL screener at a 95% conﬁdence level but is not quite better in speciﬁcity than the M-CHAT screener at a 95% conﬁdence level.





**[3.2]PAPER 2**

**Smart Autism - A mobile, interactive and integrated framework for screening and conﬁrmation of autism**

**Aim:** The aim of this paper is a mobile, interactive and integrated framework is proposed to screen and conﬁrm autism in different age group (0to17years) with 3layers of assessment process. Firstly, it screens by evaluating the responses of pictorial based screening questionnaire through mobile application. If autism is suspected, then in virtual assessment process, the child watches a video, its reaction is recorded and uploaded to the cloud for remote expert assessment. If autism is still suspected, then the child is referred to the nearest Autism Resource Centre (ARC) for actual assessment. Analyzing these results, the integrated framework conﬁrms autism automatically and reduce user’s ARC visit. It is expected that the proposed framework will bring changes in autism diagnosis process and create awareness

**Methodology:**

Smart Autism is a system for screening and conﬁrmation of autism. In order to use the system, one needs a smart device (smart phone, tablet) with internet connection. The user gets decision from the framework through a sequence of activities which are organized as follows,

1. Screener downloads the Smart Autism application in their smart device and registers the child
2. User enters the birth date of the child and app selects screening method according to the age. Then user responds to all the interactive screening questionnaires of the app.
3. The algorithm of the screening method analyzes the responses in the app and informs the user about screening decision instantly.
4. Then the cloud manager stores all the responses and decision of the Screening process in online database
5. The cloud manager initiates the Conﬁrmation process by sending a video stream to the user’s device. The video is played in front of the child and the camera of the device records the child’s reaction and behavior. Then the app uploads the video in the cloud.
6. An expert observes the video and virtually assesses the child’s condition. If the reaction and behavior of the child gives an indication of autism, then the child is referred to an ARC for conﬁrmation otherwise the child is kept in hold for 3month store-do the screening operation. This is the Virtual Assessment process.
7. The cloud manager notiﬁes the nearest ARC for the Actual Assessment process and sends its address to the user.
8. The user and the child visit the ARC and an expert or specialist evaluates the child, conducts the conﬁrmation process and uploads the decision to the system.
9. After assessing the results of Screening, Virtual and Actual Assessment process, the cloud manager notiﬁes user about autism. Then according to the decision the user is notiﬁed for further intervention plans



**Results:**

Smart Autism is a framework that involves an approach to solve problems regarding autism detection for different age group of people in one platform. In this framework people can add local screening tools to make it an effective and culturally sensitive system. The 3 layer assessment process will screen and conﬁrm autism effectively. Moreover, the cloud manager monitors and manages the entire system centrally to store information for future reference and assist the user. It is expected that people of resource limited countries will be beneﬁted by the framework and children with autism will get better service and care.

**[3.3] PAPER 3**

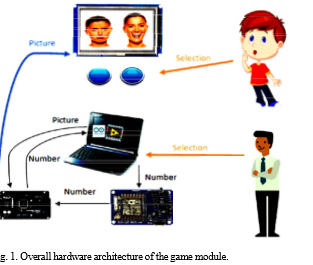
**A General Purpose Game Module for Children with Autism Spectrum Disorder**

**Aim:** The aim of this paper is to present a digital gaming module, which has been designed to help in the therapeutic process of children with ASD. The module involves displaying images on a touch screen through Arduino MEGA platform with Wi-Fi module connected to a graphical user interface to design different types of activities. The module can display various types of images as needed in the activity. The displayed images involve basic facial expressions, geometric and other objects, and text in such a way that the children could select the images corresponding to questions asked by the care-giver

**Methodology:**

The care-giver will first select the desired face expressions by pressing on the switch in the Lab VIEW interface which will transmit a number corresponding to the selection using a WIFI link. This connects to a mega microcontroller which is connected to Arduino IDE software. Depending on the number received, the program on Arduino IDE will select which image to load from SD card connected to screen. The image will be displayed and care giver will ask the question. Child should respond by pressing on one of the two buttons. The response will be transmitted to the Arduino IDE program which will tell the child if the answer is correct or not by displaying False/True sign on the screen.

In this basic configuration, the designed system is capable of displaying one or two images, such as the happy or the sad face, and corresponding switches can be used for child’s response in terms of detecting the correct emotion. This game module attracts the attention of children to be able to teach them how to identify face expressions. One can easily see that it would be extremely straightforward to modify this simple scheme for any similar game.

****

**Results:**

In this paper, a general purpose digital module has been presented. The module comprises of a small touch LCD screen with Bluetooth connectivity, through a microcontroller base and optional SD card memory shield. The device is remotely connected to a teacher’s PC running a simple graphical user interface for the teachers to modify the activity as needed. The module was designed to be small enough in size so that it could be embedded in various types of toys and gadgets acceptable for the activities planned by the teachers and other caregivers. The module provided various functionalities for the children, such as touch screen for drawing, selecting objects displayed on the screen, color recognition, shape recognition, vocabulary, etc. Initial tests have shown remarkable acceptance from children, willingness to play for longer durations, fewer instance of loss of interest, and participation enthusiasm. It was also observed that the retention was improved during the training sessions. Further studies are being conducted to collect more credible data and to produce quantified results.

**[3.4] Autism Spectrum Disorder Screening: Machine Learning**

**Adaptation and DSM-5 Fulfillment**

**Aim:**

One of the primary psychiatric disorders is Autistic Spectrum Disorder (ASD). ASD is a mental disorder that limits the use of linguistic, communicative, cognitive, skills as well as social skills and abilities. Recently, ASD has been studied in the behavioural sciences using intelligent methods based around machine learning to speed up the screening time or to improve sensitivity,specificity or accuracy of the diagnosis process. Machine learning considers the ASD diagnosis problem as a classification task in which predictive models are built based on historical cases and controls. These models are supposed to be plugged into a screening tool to accomplish one or more of the aforementioned goals. In this paper, we shed light on recent studies that employ machine learning in ASD classification in order to discuss their pros and cons. Moreover, we highlight a noticeable problem associated with current ASD screening tools; the reliability of these tools using the DSM-IV rather than the DSM-5 manual.Hence the necessity to amend current screening tools to reflect the new imposed criteria of ASD classification in the DSM-5 particularly the diagnostic algorithms embedded within these methods.

**Methodology:**

**ADOS AND ADI SCREENING TOOLS**

ADOS is one of the popular screening tools for Pervasive Development Disorder (PDD) and ASD that can be applied to

children, adolescents and adults based on a structured set of activities with a certain module. Large numbers of clinical

practices utilise ADOS for clinical diagnoses of ASD and PDD due to its reliability, validity, sensitivity and specificity. ADOS was designed by as a semi-structured test that mainly evaluates an individual’s behaviour related to language, social

interaction and play (imaginary) to assess PDD and Autism traits and levels on any individuals under examination. There are four main modules developed in ADOS for children and adults in which each is applicable to a certain population based on the behavioural and language levels ranging from verbally fluent to non-verbally fluent. Usually the examiner selects the right module for each case under examination based on two factors: the chronological age and language proficiency .

Examiners during an ADOS session observe the individual behaviours related to the primary development areas of ASD via standardised activities. Social, language communication and other behavioural activities are evaluated when cases under consideration respond to the activities assigned to them and examiners record responses. The observation normally lasts around 30-45 minutes in which responses are recorded after the session by the examiner through coding them into a computerised tool. This tool contains a handcrafted diagnostic algorithm that computes a score

**CLINICAL TOOLS AND MACHINE LEARNING: A CRITICAL ANALYSIS**

The Use of Machine Learning claimed that machine learning methods such as decision trees can be employed to construct a model that contains fewer features than items found using ADOS-R (Module 1). Therefore, the time associated with the medical diagnosis is shortened without negatively influencing sensitivity, specificity, and validity of the test. The authors sought to identify the least number of items in ADOS-R to classify ASD cases via constructing decision tree classifiers in WEKA by using information gain filtering. In particular, they have applied a number of machine learning methods (decision tree based) on an ASD dataset aiming to identify the best classifier.

A better approach toward achieving fewer numbers of features should result from investigating the complete features set significance on classification performance using feature selection methods. This may derive smaller features sets that are generic and not algorithm or data sensitive. One clear shortcoming of the data set(s) used in [25, 26] is the fact that it is clearly unbalanced and a third class/category of ASD is discarded which may simplify the problem of classification to either severe autism or no autism at all. This surely does not reflect the complexity of the problem, which contains overlapping data examples among class labels. In fact, some cases are hard to determine since they may belong to multiple categories which may confuse the algorithm. Eliminating these data examples may cause a generating of simple biased classifiers and therefore unreliable performance in terms of sensitivity, specificity and accuracy.

The process of clinical ASD diagnosis often takes between 30 to more than 120 minutes depending on:

1) The case complexity to be diagnosed

2) The clinical diagnosis procedure

3) The expertise of the clinical professional

**RESULTS:**

The classification process within machine learning is automated and ‘on the fly’, not a standalone problem with a static training data. Rather, it is a complex dynamic process integrated with a screening tool in the presence of appropriate medical staff inside a clinical environment. Unfortunately, recent studies on the use of machine learning in ASD research focusing on the diagnosis separates the machine learning from the diagnostic tool and deals with the ASD problem statically, whereas existing machine learning algorithms are merely applied on an historical dataset of cases and controls. In this paper, we focused on recent machine learning studies that tackled ASD as a classification problem and critically analysed their advantages and disadvantages. Moreover, we showed the necessary steps required to claim the development of intelligent diagnostic tools based on machine learning by replacing the handcrafted rules inside the ASD screening tools with a predictive model. Lastly, we highlighted the urgency of updating ASD clinical screening tools to reflect changes proposed in the DSM-5 manual. The dissemination of the DSM-5 demanded a change in the way that the diagnostic algorithm coded within the ASD screening tool behaves in the process of classifying cases. There is a need to re-examine questions or features within the ASD diagnostic tools to fulfil the new criteria of the DSM-5. This requires mapping the new ASD criteria to the features or attributes used in the clinical diagnosis tool as well as evaluating the way the diagnostic algorithm works. The adjustment will direct researchers to how the different Pervasive Development Disorders (PDD) overlap in the new DSM-5 criteria and this will help with improving current diagnostic tools.

## METHODOLOGY

## **CLASSIFICATION ALGORITHMS**

Classification is a technique to predict what group a certain instance is going to be. To create classifiers, we use from the given learning data set and evaluate on the test samples, so it is possible predict what class the group is following to. For Witten and Eibe (2017), “classification is sometimes called supervised because, in a sense, the scheme operates under supervision by being provided with the actual outcome for each of the examples.”

**Naive Bayes Algorithm**

Naive Bayes algorithm is the algorithm that learns the probability of an object with certain features belonging to a particular group/class. The Naive Bayes algorithm is called “naive” because it makes the assumption that the occurrence of a certain feature is independent of the occurrence of other features.

It gives us a method to calculate the conditional probability, i.e., the probability of an event based on previous knowledge available on the events. More formally, Bayes’ Theorem is stated as the following equation:

*P*(*A*|*B*)=*P*(*B*|*A*)*P*(*A*)/*P*(*B*)

Let us understand the statement first and then we will look at the proof of the statement. The components of the above statement are:

* *P*(*A*|*B*): Probability (conditional probability) of occurrence of event *A*  given the event *B* is true
* *P*(*A*) and *P*(*B*): Probabilities of the occurrence of event *A* and *B* respectively
* *P*(*B*|*A*): Probability of the occurrence of event *B*  given the event *A* is true

The terminology in the Bayesian method of probability (more commonly used) is as follows:

* *A* is called the **proposition** and *B* is called the **evidence.**
* *P*(*A*) is called the **prior** probability of proposition and *P*(*B*) is called the **prior** probability of evidence.
* *P*(*A*|*B*) is called the **posterior.**
* *P*(*B*|*A*) is the **likelihood**.

This sums the Bayes’ theorem as

Posterior=(Likelihood).(Proposition prior probability)/Evidence prior probability

D : Set of tuples

Each Tuple is an ‘n’ dimensional attribute vector

X : (x1,x2,x3,…. xn)

Let there be ‘m’ Classes : C1,C2,C3…Cm

Naïve Bayes classifier predicts X belongs to Class Ci iff P (Ci/X) > P(Cj/X) for 1<= j <= m , j <> i

Maximum PosterioriHypothesis

P(Ci/X) = P(X/Ci) P(Ci) / P(X)

Maximize P(X/Ci) P(Ci) as P(X) is constant.

With many attributes, it is computationally expensive to evaluate P(X/Ci). Naïve Assumption of “class conditional independence”

P(X/Ci)= ∏k=1 to n[ P( xk/Ci )]

P(X/Ci) = P(x1/Ci) \* P(x2/Ci) \*…\* P(xn/ Ci)

Naive Bayes classifier applies to the learning class where each instance x is described by a conjunction of attribute values and the target function can take any value

Now, if a new instance is presented in Naive Bayes described by attribute value (a1,a2,.....,an) then Naive Bayes should predict the class of the target variable for that new instance.

Naive Bayes is based on assumption that the attribute value of the instance the probability of observing the conjunction of a1,a2,a3,..........,an is just the product of the probabilities for the individual attributes .

P(a1,a2,.....,an/Vj)= ∏i=0 to n [ P(ai/Vj)]

* Naive Bayes is very simple, easy to implement and fast.
* If the Naive Bayes conditional independence assumption holds, then it will converge quicker than discriminative models like logistic regression.
* Even if the Naive Bayes assumption doesn’t hold, it works great in practice.
* Need less training data.
* Highly scalable. It scales linearly with the number of predictors and data points.

### **DECISION TREE**

The decision tree is a visual representation that is used as part of a selection criteria, or even to support the selection of specific data, considering the overall structure. It represents choices and its results in the form of a tree. It can start with simple questions that will have 2 or more answers, leading to a further question, and so on. It will support to identify and classify the data. Decision trees are mostly used in Data Mining applications using R and Machine Learning.

A decision tree will divide the data into leaf nodes and each one of them will represent an attribute. In a nutshell, decision tree is a splitting method that is applied to demonstrate every possible outcome of a decision

The name decision tree already implies the meaning of the technique. From root to leaves it can predict and classify outcomes, leading to a new question. The author sustain that the tree is placed upside down, with the leaves indicating the outcomes, and the root at the top, which represents the original data-set Zangh , affirms the following: “Because the parent population can be split into in numerous patterns, we are interested in the one with the greatest purity. In technical terminology, purity can be described by entropy.”

To control the size and to select the optimal tree size the complexity parameter (cp) is used. It will stop the tree construction in case a new variable need to be added and its value is above the cp. It stipulates how the cost of a tree is penalized considering the number of terminal nodes .

## **RESULTS AND DISCUSSION**

## **THE DATA-SET**

Raw data-set contains ten binary variables representing the screening questions (A1\_Score to A10\_Score).

|  |  |
| --- | --- |
| VARIABLES | DESCRIPTION |
| Age: | Age in years |
| Gender: | Male or female |
| Ethnicity: | List of common ethnicities in text format |
| Born with Jaundice: | Whether case was born with jaundice |
| Family member with PDD: | Whether any immediate family member has a PDD |
| Who is completing the test: | Parent, self, caregiver, medical staff, clinician, etc |
| Country of Residence: | List of countries in text format |
| Used the screening app before: | Whether the user has used screening app |
| Screening Method Type: | Type of screening method chosen based on age category |
| Question 1 Answer: | I often notice small sounds when others do not |
| Question 2 Answer: | I usually concentrate more on the whole picture, rather than the small details |
| Question 3 Answer: | I find it easy to do more than one thing at once |
| Question 4 Answer: | If there is an interruption, I can switch back to what I was doing very quickly |
| Question 5 Answer: | I find it easy to read between the lines when someone is talking to me |
| Question 6 Answer: | I know how to tell if someone listening to me is getting bored |
| Question 7 Answer: | When I’m reading a story I find it difficult to work out the character’s intentions |
| Question 8 Answer: | I like to collect information about categories of things (e.g. types of cars, types of bird, types of train, types of plant, etc) |
| Question 9 Answer: | I find it easy to work out what someone is thinking or feeling just by looking at their face |
| Question 10 Answer: | I find it difficult to work out peoples intentions |
| Screening Score: | Final score obtained based on scoring algorithm of screening method used |

**Source Code:**

mydata<-read.csv(file.choose())

str(mydata)

dim(mydata)

tindex = sort(sample(nrow(mydata), nrow(mydata)\*.7))

mtraining<-mydata[tindex,]

mtesting<-mydata[-tindex,]

library(e1071)

NB<-naiveBayes(Class.ASD~.,data=mtraining)

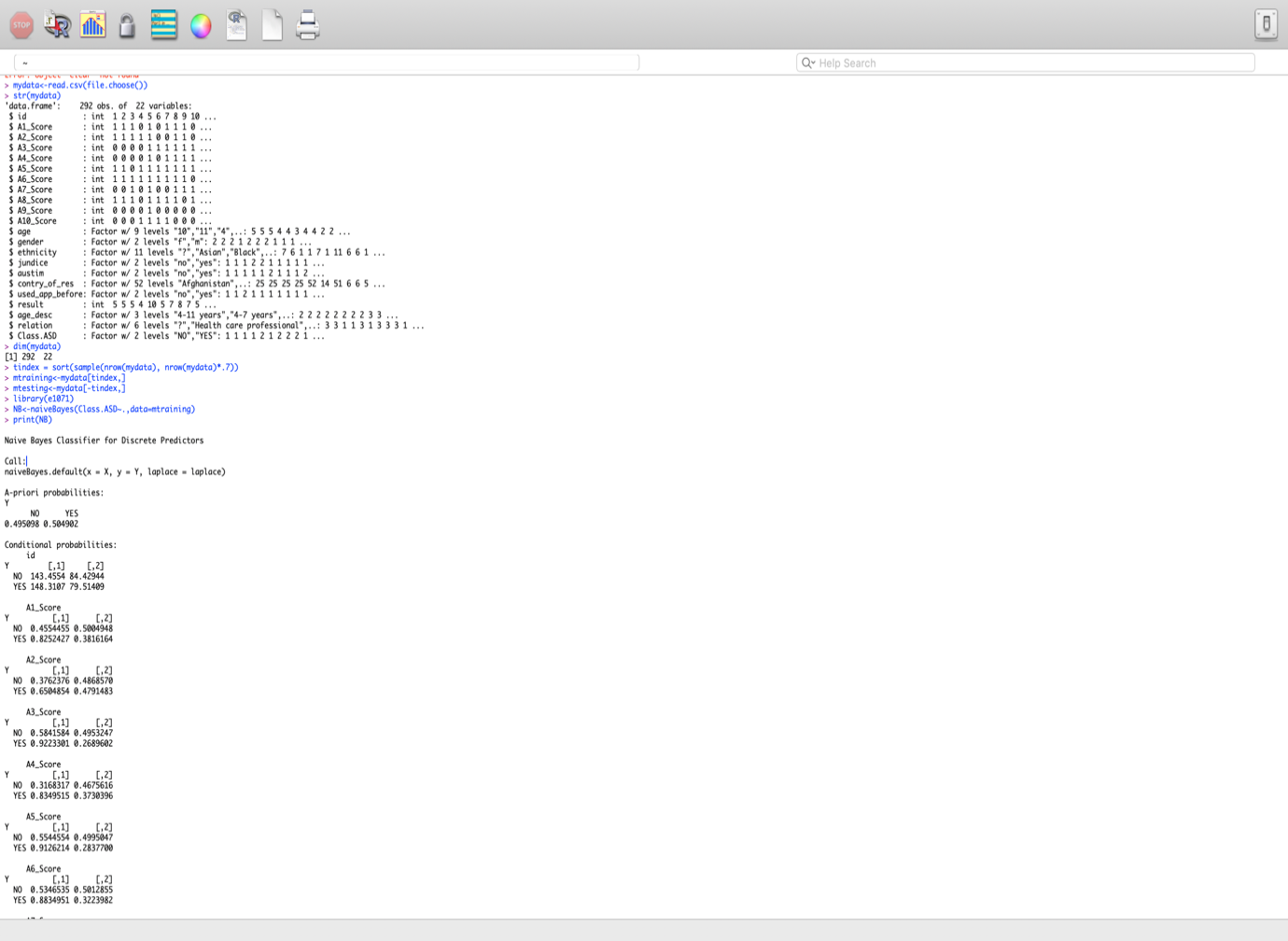
print(NB)

predNB1<-predict(NB,mtesting)

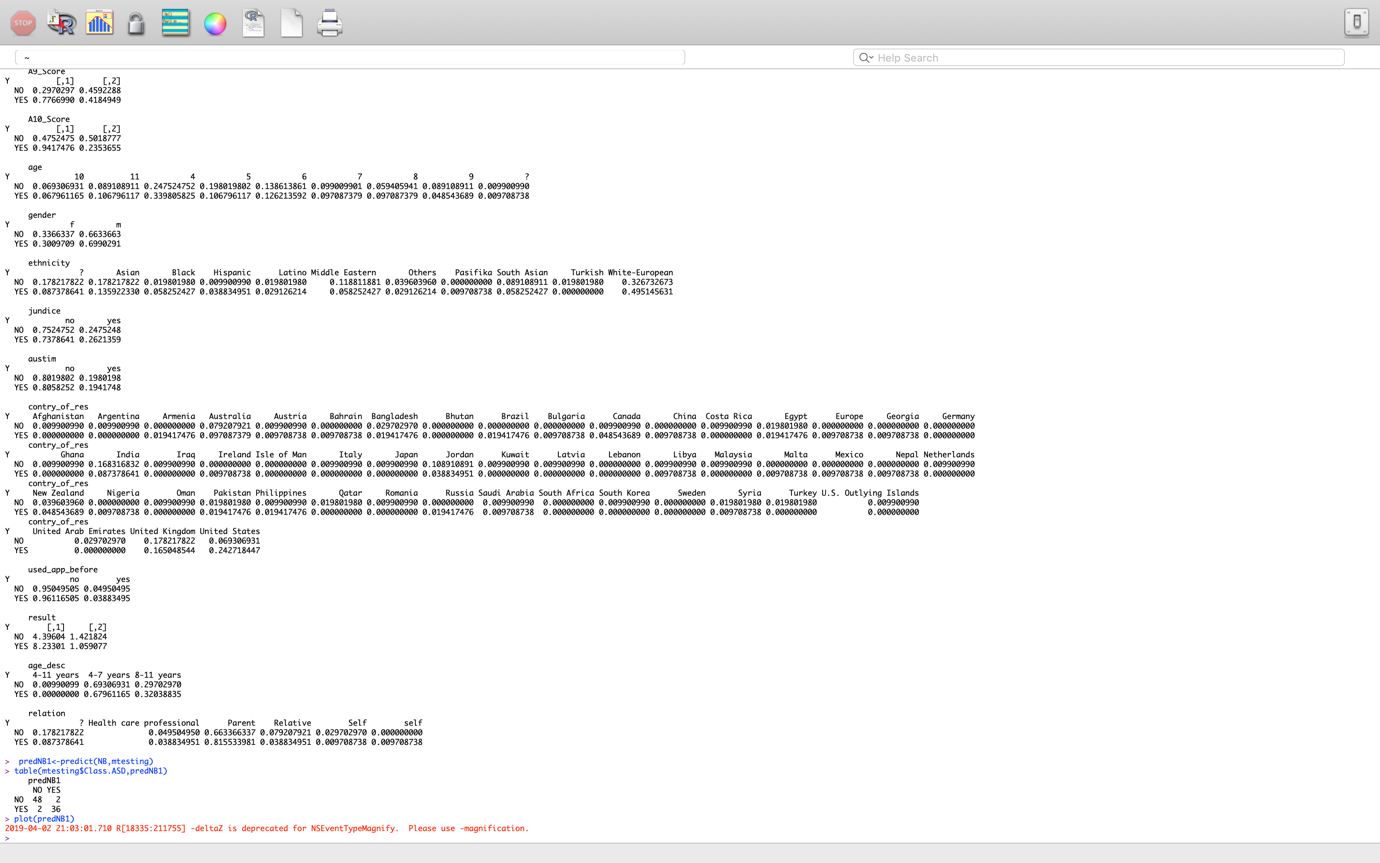
table(mtesting$Class.ASD,predNB1)

plot(predNB1)

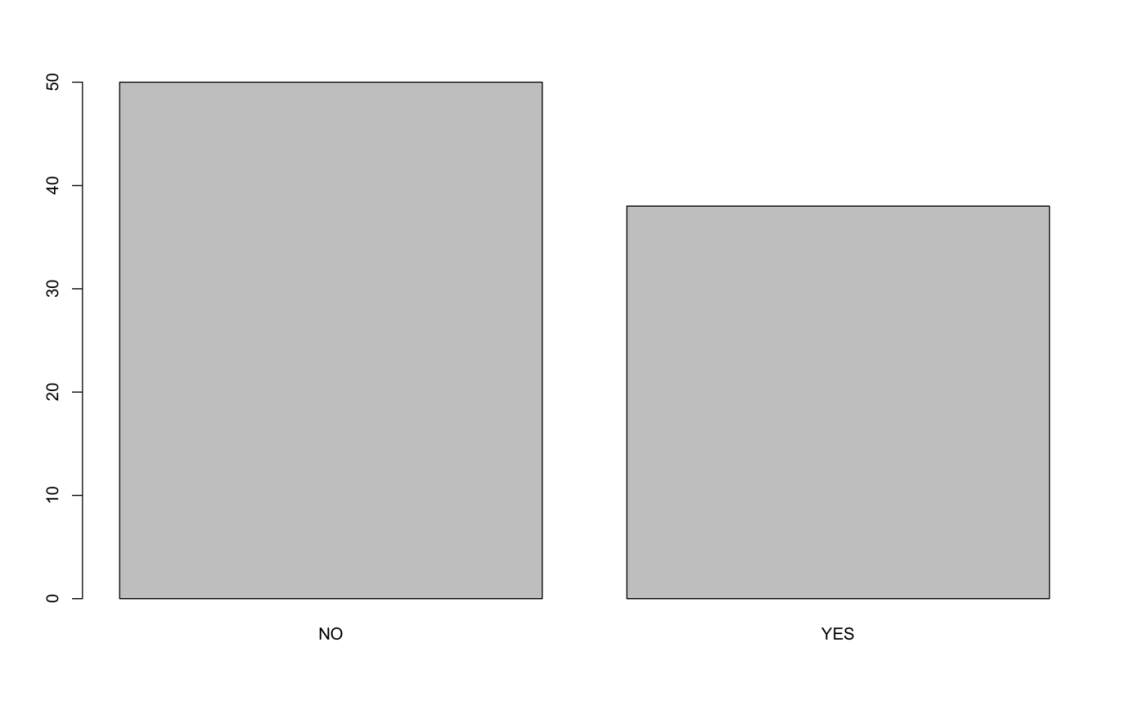
**Output:**

****

This output shows the brief details of the dataset , the dimensions of the dataset .After that the data is divided into training data and testing data ,then we will apply the naïve bayes algorithm .Then we got the prior probabilities of no and yes and the conditional probabilities of each attribute in the dataset.

****

This output shows the conditional probabilities of each and every attribute in the dataset and the prediction table of no and yes .It shows that 48 are no’s which are correctly predicted and 36 yes are correctly predicted.

****

This graph shows the predication of no and yes from the given dataset and how accurate they are. No are 48 predicted correctly and Yes are 36 predicted correctly.

**SOURCE CODE:**

library(caret)

library(rpart.plot)

mydata<- read.csv(file.choose())

str(mydata)

head(mydata)

set.seed(3033)

intrain<- createDataPartition(y=mydata$ASD, p=0.7, list=FALSE)

training<- mydata[intrain,]

testing<- mydata[-intrain]

dim(training)

dim(testing)

anyNA(mydata)

summary(mydata)

trctrl<- trainControl(method = "repeatedcv", number = 10, repeats = 3)

set.seed(3333)

dtree\_fit\_gini<- train(ASD ~., data = training, method = "rpart",

parms = list(split = "gini"),

trControl=trctrl,

tuneLength = 10)

dtree\_fit\_gini

prp(dtree\_fit\_gini$finalModel, box.palette = "Blues", tweak = 1.2)

**OUTPUT:**

**Description: A screenshot of a social media post

Description generated with very high confidence**

In this output screenshot we are adding the library of caret and rpart.plot after that reading the input file .then the contents of the dataset are displayed .Then the dataset in splited into training dataset and testing dataset .then displaying the dimensions of both the datasets.

**Description: A screenshot of a computer

Description generated with very high confidence**

In this we are displaying the summary of the dataset. then we are applying the gini index for the construction of the decision tree.

**Description: A screenshot of a cell phone

Description generated with high confidence**

Decision tree for the given dataset

So by using Naïve bayes we got accuracy of 0.9545

By using decision tree we got accuracy of 0.88889

So we can say that naïve bayes algorithm gives more accuracy then decision tree

**CONCULSION**

Finally this project is successful in implementing all the constraints which are required for solving this problem and it gives a correct output We will give data with missing values and later we use machine learning as a source to the problem we predict in what cases Autistic Spectrum Disorder takes place and in which situations it effects the health and we also done screening tests for the Autistic Spectrum Disorder, Thus our project gives result with maximum accuracy . we focused on recent machine learning studies that tackled ASD as a classification problem and critically analysed their advantages and disadvantages. Moreover,we showed the necessary steps required to claim the development of intelligent diagnostic tools based on machine learning by replacing the handcrafted rules inside the ASD screening tools with a predictive model.

**REFERENCES**

1. Abdelhamid N. and Thabtah F. 2014. Associative Classification Approaches: Review and Comparison. Journal of Information and Knowledge Management (JIKM), 13(3). Incorrect
2. American Psychiatric Association. 2013. Diagnostic and statistical manual of mental disorders: DSM-5. Washington, D.C: American Psychiatric Association.
3. American Psychiatric Association 2000. Diagnostic and statistical manual for mental disorder (4th edn), Text revision. Washington, DC: American Psychiatric Association.
4. Berument, S. K., Rutter, M., Lord, C., Pickles, A., Bailey, A., MRC Child Psychiatry Unit, & Social, Genetic and Developmental Psychiatry Research Centre 2000. Autism Screening Questionnaire: Diagnostic validity. British Journal of Psychiatry, 175, 444–451.
5. Bolton P, Macdonald H, Pickles A, Rios P, Goode S, Crowson M, Bailey A, Rutter M. 1994. A case-control family history study of autism. J Child Psychol Psychiatr 35:877–900.
6. Duda M., Ma R., Haber N., Wall D.P. 2016. Use of machine learning for behavioral distinction of autism and ADHD. Translational Psychiatry 9(6), 732.