Comparing Models For Predicting Autism Using Rapid Miner Tool

R. Ramya[1], B. S. E. Zoraida[2]
[1] Dept of Computer Science
[2]Assistant Professor, Dept of Computer Science
[1, 2] Bharathidasan University, Tiruchirappalli – 620023

Abstract- *Autism is group of developmental disability in human brain. Autism is not a single disorder, but a spectrum of strongly related disorder with a common core of symptoms. Every human being with Autism has troubles to some quantity through understanding, social interactions, communication and flexible behavior but the point of disability and the grouping of symptoms various tremendously from person to person. Autism can be classified as three different levels are Classic Autism (Low), Aspenger's Syndrome (moderate), and Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS (high)). These three different levels of disorders contribute too many of the same symptoms but they differ in their severity and impact. In this paper focuses on diagnosing autism based on ASD dataset. Hence various data mining techniques can be used to predict three different levels of autism using the ASD dataset.*

Keywords- Autism Spectrum Disorder (ASD), Predictive Models, Rapid Miner tool.

I. INTRODUCTION

Diagnosis disease important issues in the health care organization. The autism is one of the complex diagnosing spectrum disorders. Autism is a difficult neurobehavioral situation that includes destruction in social interaction, developmental of language and communication skills combined with repetitive behaviors. Because of the variety of symptoms, this condition is now called Autism Spectrum Disorder (ASD). There are three different forms of autisms classifications are Classic Autism (Low), Aspenger's Syndrome (moderate), and Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS (high)). These three disorders are affecting adult. Children with ASD can be diagnosed within two years. Diagnosing ASD in early stage is not easy. In adults some ASD symptoms of other mental health disorders such as Attention Deficit Hyperactivity Disorder (ADHD). In ASD is usually diagnosed in early childhood, but genetic detection of this brain disorder could mean more timely interventions that improve life for the patient and their career.

The Autism dataset is collected from government website in VAERS. In this website approximately 30,000 VAERS reports are filled in each year and each data is assigned a unique identification number. This paper focuses on diagnosing risk level of autism. By using data mining techniques in Naïve Bayes, K-Nearest Neighbor, Gradient Boosted Tree, Deep Learning and Logistic Regression.

II. LITERATURE REVIEW

There are various works done in the area of ASD for prediction, classification, and analysis for diagnosing the level of autism, analysis of autistic patient, to provide alertness and correct kinds of therapy. The data mining in health care has become increasingly popular because it offers benefits to care providers, patients, health care organization, researchers and insurers. Rapid Miner software supports data mining methodology to store, represent, filter and analyzed the acquire ASD dataset. Margaret H Duham proposed a data mining system is a simple and better understanding for data mining various methods. Data mining is used for predicting the various complex disorders [1]. S. Wheelwright et al. by investigate the association between scores on the Empathy Quotient, Systemizing Quotient-Revised and Autism Spectrum Quotient in both a big sample of typical participants, and a sample of adults with autism spectrum conditions [2]. Laurence Chaby et al. proposed a multidisciplinary approach to multimodal social emotional behaviors in children with autism spectrum disorders and investigations, interfaces between social signal processing and psychopathology is a useful aid to classify disorder characteristics, improved early symptoms, and notify the treatment [3]. E. M. Albornoz et al. describe an soft computing in evolutionary method for automatic features selection of speech in a classification function and based on a genetic algorithm that identify the best combination of phonic features using SVM as classification method [4]. Priyanka Juneja and Anushul Anand proposed ASD patients are analyzed with interpretation value analysis and it is appropriated parameter established fuzzification that will operate the analysis based on some specifications [5]. Ionut Taranu describes Data mining system has big importance for area of healthcare, and it used in

healthcare organizations. Data mining can be used to generate successful decisions that will advance success of healthcare system and health of the patients [6]. Abdollah Ghasempour et al. identify the importance of Emotional Intelligent and its aspects in Autism in the university students. Thus, development of Emotional Intelligent and its dimensions can decrease autism symptoms and signs in person with ASDs and the completed of this study emphasizes the key role of Emotional Intelligent as a reasonable health predictor [7]. Mohana E and Poonkuzhali S used data mining techniques to classify the risk level of autism children. To categorize the autism children dataset is used and different feature selection algorithms- ReliefF, Fisher filtering, stepdisc and Runs filtering are used to mine the required features to describe the risk level of autism. The classification algorithms like Naïve Bayes, K-NN, C4.5, C-RT, BVM, C-SVC, CVM, PLS-LDA ID3, CS-MC4, Linear discriminant, and Rnd tree are tested to the feature selection methods and the accuracy and error rate are intended and work find that BVM, CVM and MLR provides high accuracy and the best classification [8]. Sumi Simon et al. did a comparison on data mining classification methods for ASD prediction. The work aimed at comparing the performance through accuracy of different data mining algorithms. SVM, J48, BVM and decision tree are the most commonly used algorithms are best suited to classify the autistic data and it provides high accuracy and low error rate [9]. M. S. Mythili and A. R. Mohamed Shanavas proposed an advanced ASD predictive method among children using fuzzy cognitive map and feature extraction algorithms [10]. Tanaya Guha et al. describe decreased complexity in facial expression dynamics of discussion with High Functioning Autism relative to their typically rising peers. Major difference is analyzed for expressions related to joy, disgust and sadness [11]. Ashmeet Singh and R. Sathyarai, proposed a comparison on data mining classification algorithms on different datasets methodologies using rapid miner. Naïve Bayes, Random Forest, Decision tree algorithms are compared to find the classification accuracy and it is concluded Naïve Bayes provides highest accuracy [12].

III. PROPOSED WORK

The implementation work is predicting the different level of autism using various data mining models in Rapid Miner tool. The prediction of ASD has been divided into three different levels as Low-functioning autism, moderate autism and High-functioning autism. In figure 1 is show proposed flowchart for the predicting different level of autistic children using valuable data mining classification. The major advantage of this work is higher accuracy of prediction.

A. Rapid Miner tool

Rapid Miner tool is a data science software platform. It is an integrated environment for data preparation, machine learning, deep learning, and text mining and predictive analytics. Rapid Miner is developed on an open core model and written in the java programming language. It is provide data mining and machine learning procedures including data loading and transformation, and preprocessing and visualization, Predictive analytics and statistical modeling, evaluation and deployment.

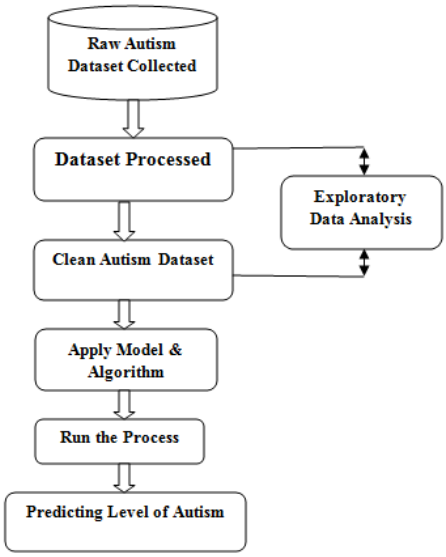

Figure 1: Proposed Architecture

B. Dataset

The raw dataset had been downloaded from their government website in VAERS [12]. In these dataset contained on eight attribute and autism dataset is CSV (Comma Separated) file format. This table 1 represents the description for attribute. Each data had to be provided with certain label (low, middle, high) for level of autism patients. In these Rapid Miner software supports methodology to store, represent, filter and analyzed the acquire ASD dataset.

Table 1: Dataset Explanation

NO.	Attributes	Explanation
1	VAERS_ID	It is each data assigned a unique identification number in VAERS website.
2	Gender	It is defined autism patient gender.
3	Symptom 1	It is defined symptom stage 1
4	Symptom 2	It is defined symptom stage 2
5	Symptom 3	It is defined symptom stage 3
6	Symptom 4	It is defined symptom stage 4
7	Symptom 5	It is defined symptom stage 5
8	Level of Autism	It is label attribute and it is predicting three different levels as low, medium, high.

C. Predictive Modeling

The predictive modeling is a process that uses data mining and probability to forecast outcomes. The prediction is supervised learning where the data are used directly to predict the class value of new instance. These predictions have different steps like creation the machine learn the pattern basically training data. The models to correctly predicting the dataset into different class labels and finally evaluating the classification using test data. This paper have been chooses five classifiers.

i. Naïve Bayes: A Naïve Bayes classifier is a simple probabilistic classifier based on applying Bayes theorem with strong independence assumption. A more descriptive term for the underlying probability model would be 'independent feature model'. In uncomplicated terms, Naïve Bayes predictive models assume the occurrence of exacting characteristic of a class is unrelated to the presence of any other feature [14].

ii. K-Nearest Neighbor: The K-Nearest Neighbor algorithm is based on learning by analogy, that is, by comparing a given test example with training examples that are similar to it. The training examples are described by n attributes. Each example represents a point in an n-dimensional space. In this way, all of the training examples are stored in an n-dimensional pattern space [14].

iii. Gradient Boosted Tree: A Gradient Boosted Tree is a collection of either regression or classification tree models. Together are forward-learning ensemble methods that obtain analytical results through regularly enhanced estimation. Boosting is a flexible nonlinear regression method that helps improving the precision of trees. By serially applying feeble classification algorithms to the incrementally changed data, sequences of decision trees are created an ensemble of weak prediction models. While boosting trees increases their precision, it also decreases speed and human interpretability [14].

iv. Deep Learning (Neural Network): Deep Learning is based on a multi-layer feed-forward artificial neural network that is trained with stochastic gradient descent using back-propagation. The network can contain a large number of hidden layers. Advanced features such as adaptive learning rate, rate annealing, momentum training, dropout and L1 or L2 regularization enable high predictive accuracy. Each compute node trains a copy of the global model parameters on its local data with multi-threading and contributes periodically to the global model via model averaging across the network [14].

v. Logistic Regression: a Logistic Regression is a mathematical modeling approach in which the best fitting, yet least-restrictive model is desired to describe the relationship between several independent explanatory variables and a dependent dichotomous response variable. The learning method can be used for both regression and classification and provides a fast algorithm and good results for many learning task. This operator supports various kernel types including dot, radial, polynomial, neural, anova, epachnenikov, Gaussian combination and multiquadric [14].

IV. RESULT AND ANALYSIS

The predicting level of ASD used to compare the result of different five data mining techniques in rapid miner tool. K-Nearest Neighbor (K-NN), Naïve Bayes, Gradient Boosted Tree (Tree), Deep Learning (Neural Network), and Logistic Regression models are executed and compare to check the accuracy, classification error, recall and Precision. In this figure 2 X-Axis represent classifiers components and Y-Axis represents the predictive models. In these are all models are analyzed on classifier components using statistical analysis. The Deep Learning model is best suited to predict the Autism dataset and it provides highest Accuracy, Recall, Precision and Lowest Classification error.

In figure 3 gives the accuracy parameters and classification error of different predictive models. The graphical representation for Accuracy, Classification-Error, Recall and Precision is shown in figure 3.

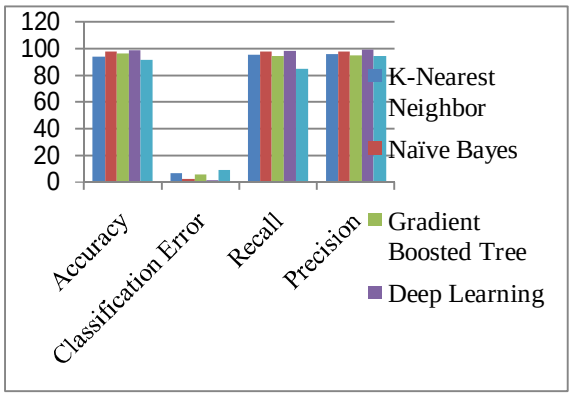

Figure 2: Compare Different models Accuracy and Error-Rate.

V. CONCLUSION

In this paper an attempt is made to diagnosing level of autism with the dataset using different models like, Naïve Bayes, K-Nearest Neighbor (K-NN), Gradient Boosted Tree (Tree), Deep Learning (Neural Network), and Logistic Regression models are evaluation to find the accuracy, classification error, recall and precisions. Among the above models deep learning model outperform the other models.

REFERENCES

[1] Margaret H. Dunham, data mining: Introductory and Advanced Topics, 2nd ed., Pearson Education, 2006.

[2] S. Wheelwright, S. Baron-Cohen, N. Goldenfeld, J. Delaney, D. Fine, R. Smith, L. Weil, A. Wakabayashi, "Predicting Autism Spectrum Quotient (AQ) from the Systemizing Quotient-Revised (SQ-R) and Empathy Quotient (EQ)", 2006.

[3] Laurence Chaby, Mohamed Chetouani, Monique Plaza, David Cohen, "Exploring Multimodal Social-Emotional Behaviors in Autism Spectrum Disorders" , 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust.

[4] E. M. Albornoz, L. D. Vignolo, C. E. Martínez & D. H. Milone, "Genetic Wrapper Approach for Automatic Diagnosis of Speech Disorders related to Autism" 14th IEEE International Symposium on Computational Intelligence and Informatics (CINTI), nov, 2013.

[5] Priyanka Juneja and Anshul Anand, " International Journal of Computer Science and Mobile Computing", Vol.3 Issue.7, July- 2014, pg. 585-593

[6] Lonut Taranu, "Data mining in healthcare: decision making and precision", Database Systems Journal vol. VI, no. 4/2015.

[7] Abdollah Ghasempour, Abutaleb Tavakoli, Mansour Mahmodi Aghdam, Maede Mosavi, "Predicting Autism Spectrum Disorders on the Basis of Emotional Intelligence Among University Students", Health Scope. 2015 August; 4(3): e22441.

[8] Mohana E & Poonkuzhali S, " Categorizing the Risk Level of Autistic Children using Data Mining Techniques," International Journal of Advance Research In Science and Engineering, vol.04, pp. 223-230, I. Simon, Chandra J and Saravanan N, "Empirical Evaluation of Data Mining Classification Methods for Autistic Children", International Journal of Trend in Research and Development (IJTRD) in 2016.

[9] M. S. Mythili and A. R. Mohamed Shanavas," An Improved Autism Predictive Mechanism Among Children Using Fuzzy Cognitive Map And Feature Extraction Methods (Feast), VOL. 11, NO. 3, FEBRUARY 2016.

[10] Tanaya Guha, Member, IEEE, Zhaojun Yang, Student Member, IEEE, Ruth B. Grossman, Shrikanth S. Narayanan, Fellow, IEEE, "A Computational Study of Expressive Facial Dynamics in Children with Autism", IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, VOL. XX, NO. X, MARCH 2016.

[11] Ashmeet Singh1, R Sathyaraj, "A Comparison Between Classification Algorithms on Different Datasets Methodologies using Rapidminer", International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 5, May 2016.

[12] "Vaccine Adverse Event Reporting System" (VAERS), https://vaers.hhs.gov/data/data.

[13] "Rapid Miner tool", http://rapidminer.com/