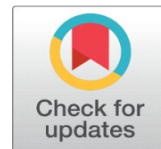# NON-PARAMETRIC RANDOMIZED TREE CLASSIFIER FOR DETECTION OF AUTISM DISORDER IN TODDLERS

Prajwala T R [1] ✉

[1] Assistant Professor, Department of Computer Science and Engineering PES University, India.

ABSTRACT

Autism is a behavioral disorder seen in toddlers and adolescents. It is a disorder which concerns behavior of child, speech, social interaction of child as well as nonverbal communication of child is affected. The parents of affected children find it very cumbersome to manage the child. Detection of such anomalies is really important at early stages. This paper mainly focuses on early detection of autistic behavior in toddlers. There are various machine learning and deep learning algorithms. Non parametric Extreme randomized classifier is one such technique which helps in early detection of autistic behavior in toddlers. The various performance evaluation metrics used are Jaccard score, ROC Curves and Mean Squared Error. The Feature selection is done using spearman correlation to identify the features affecting the child most and represented in form of Heat map. Extra tree classifier proves a better algorithm in detection of autism at early stages of child development.

Keywords: Jaccard Score, ROC Curves, Spearman Correlation, Gini Coefficient, Extra Tree Classifier and Negative Mean Absolute Error

1. INTRODUCTION

Autism disorder is a behavioral and developmental disorder. The occurrence of disorder can't certain abnormalities in brain development which can be structural abnormality or functional abnormality. The symptoms of disorder include lack of eye contact, lack of communication or lack of speech and interaction also involuntary bowel movements. It is important for parent to identify the autism spectrum disorder at early stages to help child. Rasool Azeem Musa et al (2020).

The machine Learning algorithms help the autistic detection. One such algorithm is Extreme randomized classifier. The classifier is a non-parametric learning technique which uses the randomness of the decision tree in detection and classification of dataset as autistic or not. The randomness of tree and Gini Index helps in better classification of samples. The Spearman correlation coefficient identifies the features affecting the toddler the most. Spearman correlation gives a measure of strength between the target variable of detection of autistic behavior and other features considered. American Psychiatric Association (2000) The Jaccard Score is a measure of similarity as well as diversity in data samples. The ROC curves a plot of true positive and false positive rate gives the classification measure of autistic data set. The learning curves and scalability curves are measure of cross validation scores of the classifiers. McClellan (2020)

The work focuses on detection of autism at early stages in toddlers which help the guardians of children in taking care of them. The Non parametric classifiers of machine learning are used to detect the autistic spectral disorder. The performance evaluation metrics are Jaccard score ROC curves, Precision and Recall.

2. EXTRA TREE CLASSIFIER

The Extra Tree classifier is implemented for Autistic data set. The autism data set has 1055 data samples. The target variable is categorical with type yes/no for autism detection. The attribute type is categorical where there are 17 attributes considered. The data set doesn't have missing values. The data set is preprocessed so that all variables are categorical for the classifier to detect the presence or absence of autism. The data set includes following features: sex, family history presence or absence of Jaundice and quantitative checklist. The quantitative checklist accumulates the scores for the 10 questions. Alarifi and Young (2018) The 10 questions mainly focus on whether child responds to his or her name, child's social wellbeing, child's gestures.

The Extra tree classifier is similar to random forest ensemble technique and deviates from random forest classifier in mode of construction of ensemble of tree. The de correlation of trees is due to random selection of trees. The Gini Index is measure of purity of node in extra tree classifier Electrical, Computer and Communication Engineering (2019). The optimization element of extra tree classifier is still an issue though randomness of classifier gives best results for the autistic dataset.

This paper focuses on application of extra tree classifier to detect autism in toddlers at early stages. The Gini importance is used to identify the feature importance of autistic dataset Fadi Thabtah (2018). The correlation between the features is represented using Heat map. The spearman correlation coefficient is used to identify the correlation between the features for prediction of autism in child. The Jaccard score, ROC curves and accuracy are the primary evaluation metrics for validating the prediction of autism in the toddlers at early stage.

3. FEATURE SELECTION

The feature selection is implemented using spearman correlation coefficient. The spearman correlation coefficient defines the strength of variables. Since the data set is categorical spearman correlation coefficient judges the relation between variables best.

The Spearman Rank correlation coefficient is defined as:

$$R = 1 - 6\frac{\sum d^2}{n(n-1)}$$

Where,

n is number of samples which is 1055 samples in current dataset.

d is difference of ranks between various observations made in autistic data set.
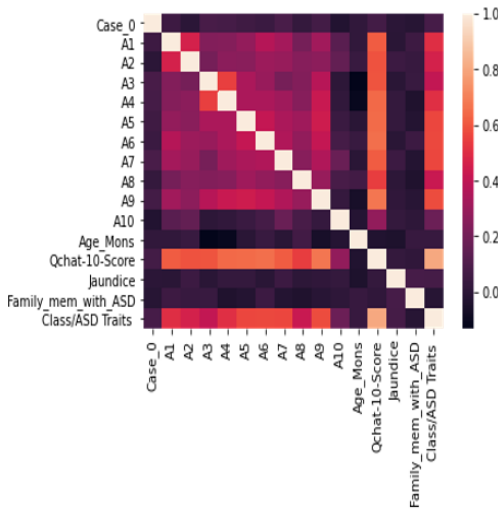
Figure 1 Heat map of Spearman correlation coefficient for autistic dataset

Figure 1 is a heat map. The heat map is generated by using spearman correlation coefficient which shows which attribute has maximum correlation for detecting autism spectrum disorder. Based on the above correlation coefficient the features of child able to identify things like toys, child able to respond to his or her name and child able use gestures like hi or goodbye. Among the 17 features these features are found to be more corelated to identification of autism in child. The Gini Coefficient identifies the feature importance using following formula

$$G = \frac{\sum_{i=1}^{n} \sum_{J:1}^{n} |x_i - \mathbb{Z}_j|}{2n^2 \bar{x}}$$

Where,

xi is the target class variable

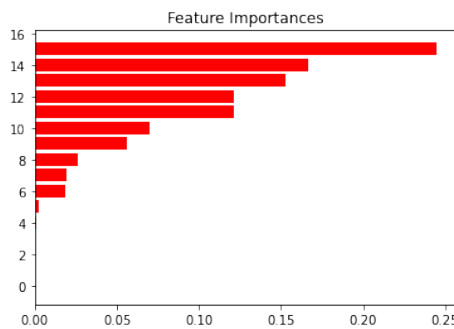x is the mean of the data

n is number of samples.

Figure 2 Relative feature importance calculated using Gini coefficient for autistic toddler data set.

The Figure 2 shows the result of feature importance using Gini Coefficient. Accordingly, the following inference can be made: The family environment is an important feature affecting the behavior of autistic child, child able to identify things

like toys, child able to respond to his or her name and child able use gestures like hi or goodbye.
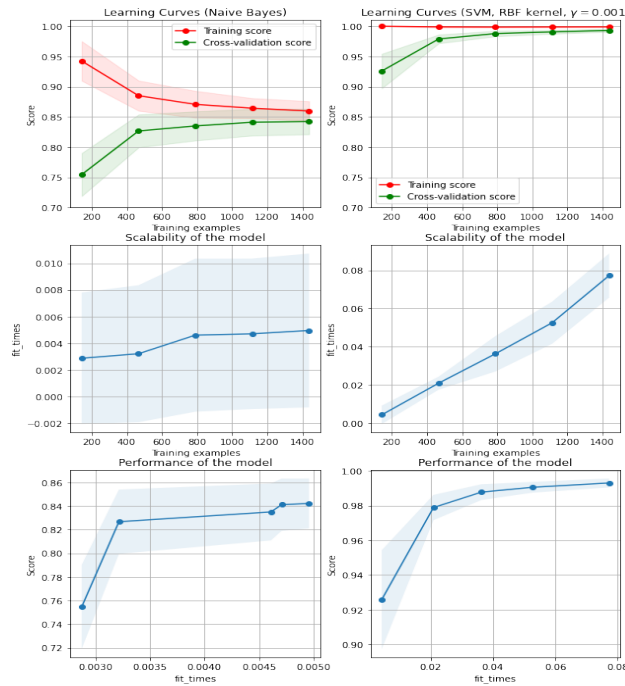

Figure 3 Learning curves using Naïve base classifier, SVM, scalability of model, Performance of model (left to right)**.**

The Extra tree classifier is compared against Naïve bayes and SVM model. The training score and cross validation score is as shown in Figure 3. The scalability and performance of the model is also shown in Figure 3. But the Extra tree classifier proves it is the best algorithm. The evaluation metrics used are Jaccard score, accuracy and ROC curves.

4. RESULTS AND DISCUSSION

The extra tree classifier is implemented to find the best accuracy in detection of autism in toddlers at early stage. The randomness of extra tree classifier makes it possible to handle the autistic data. The Jaccard score measures the diversity as well as similarity in autistic data set of 1055 samples. The Jaccard score is calculated as follows:

$$J = \left(\frac{(A_P \cap p_N)}{ApUA_{\mathbb{N}}} \right) * 100$$

Where,

A_p is number of samples which exhibit autistic disease.

Higher the value of index better is the feature selection. The Jaccard score for the autistic data set is found to be 0.974. It indicates that the feature selection is more accurate compared over Naïve bayes of SVM which is 0.856 and 0.899 respectively.

Table 1 Evaluation Metrics for Extra tree classifier over Autism data set	
Evaluation metrics	Value
Accuracy	0.981
Precision	0.98
Recall	0.99
F1 score	0.98
Roc accuracy score	0.998
Negative mean squared error	0.189

The Table 1 clearly depicts that extra tree classifier is better in terms of accuracy, precision, recall, F1 score and negative mean squared error. ROC (Receiver Operating curves) is a measure of true positive and false positive rate. The Figure 4 shows the ROC curve plotted for autistic dataset. The Figure 4 shows that the classifier maps the close to ideal ROC curve.
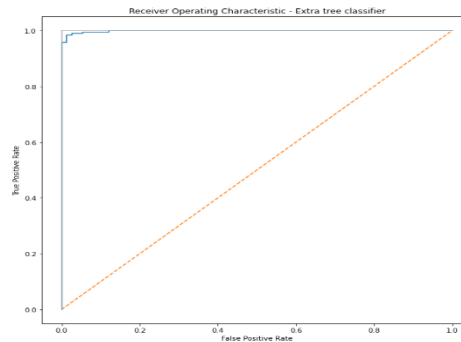
Figure 4 ROC curve of classifier over autistic dataset.

The extra tree classifier proves that it is one of the best algorithms which can be applied to detect autism in toddlers. The randomness of tree is one of the major reasons why the algorithm predicts better over Navie Bayes and SVM classifier.

5. CONCLUSION AND FUTURE WORK

The autism spectral disorder is a behavioral disorder keen to be seen in toddlers. It is very difficult for the toddler and the parents to identify the disorder at early stages. The Extra tree classifier is an algorithm which is applied to obtain an accuracy of 98% to detect the autistic disorder in early stages. The Feature selection and correlation is done using Gini Correlation and Spearman correlation coefficient. The Negative mean squared error is very less which is 0.189. The ROC curves plotted are close to ideal results. Jaccard score is also high which enables to get high accuracy and similarity in autistic data set. The future work can be to identify the effect of the diet of toddlers on behavior of autistic children.

REFERENCES

American Psychiatric Association. (2000) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision. Washington, DC: American Psychiatric Association

Electrical, Computer and Communication Engineering (ECCE) Cox's, (2019) pp. 1-6, IEEE, Bazar, Bangladesh.

Fadi Thabtah, (2018) Firuz Kamalov, Khairan Rajab, A new computational intelligence approach to detect autistic features for autism screening, International Journal of Medical Informatics, Volume 117, Pages 112-124, ISSN 1386-5056. Retrieved from https://doi.org/10.1016/j.ijmedinf.2018.06.009

K. S. Omar, P. Mondal, N. S. Khan, M. R. K. Rizvi, and M. N. Islam, (2019) "A machine learning approach to predict autism spectrum disorder," in International Conference on. Retrieved from https://doi.org/10.1109/ECACE.2019.8679454

Kayleigh K. Hyde et.al (2019) "Applications of Supervised Machine Learning in Autism Spectrum, Disorder Research: a Review", Review Journal of Autism and Developmental Disorders 6:128-146. Retrieved from https://doi.org/10.1007/s40489-019-00158-x

McClellan, David (2020) A.Data Analysis and Classification of Autism Spectrum Disorder Using Principal Component Analysis, Advances in Bioinformatics, Retrieved from https://doi.org/10.1155/2020/3407907

Nadire Cavus et.al (2021) "A Systematic Literature Review on the Application of Machine-Learning Models in Behavioral Assessment of AutismSpectrum Disorder", J. Pers. Med, 11, 299. Retrieved from https://doi.org/10.3390/jpm11040299

Nogay, Hidir Selcuk and Adeli, Hojjat. (2020) "Machine learning (ML) for the diagnosis of autism spectrum disorder (ASD) using brain imaging" Reviews in the Neurosciences, vol. 31, no. 8, pp. 825-841. Retrieved from https://doi.org/10.1515/revneuro-2020-0043

Rasool Azeem Musa et al (2020) "Predicting Autism Spectrum Disorder (ASD) for Toddlers and Children", ICMAICT

S. Alarifi and G. S. Young, (2018) "Using Multiple Machine Learning Algorithms to Predict Autism in Children," in Proceedings on the International Conference on Artificial Intelligence (ICAI), pp. 464-467. Retrieved from https://www.proquest.com/openview/fb51efb158b86219eb72ad711681 5b22/1?pq-origsite=gscholar&cbl=1976349

Thabtah, Fadi. (2018) "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. " Informatics for Health and Social Care : 1-20. Retrieved from https://doi.org/10.1080/17538157.2017.1399132

uman Raj, Sarfaraz Masood, (2020) Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques, Procedia Computer ScienceVolume 167,Pages 994-1004, ISSN 1877-0509. Retrieved from https://doi.org/10.1016/j.procs.2020.03.399