EasyChair Preprint

№ 7432

Childhood Autism Spectrum Disorder Screening Using Machine Learning

R Monika, Hamsini R Koushik, B M Divya, Monisha K Shree and M D Deepak

February 7, 2022

Childhood Autism Spectrum Disorder Screening using Machine Learning

Monika R[*], Hamsini R Koushik[*], Divya BM[*], Monisha Shree K[*], Deepak MD[#]

[*]UG Student, Dept. of CSE, K S School of Engineering and Management, Bengaluru 560109
[#]Assistant Professor, Dept. of CSE, K S School of Engineering and Management, Bengaluru 560109

monikavani21@gmail.com, hamsinirk999@gmail.com, divyabm2911@gmail.com, monishanaik08@gmail.com, deepak.m.d@kssem.edu.in

Corresponding Author: deepak.m.d@kssem.edu.in
Affiliated to Visvesvaraya Technological University, Gnana Sangama, Belagavi India

Abstract – A child/person having Autism Spectrum Disorder(ASD) which is a neurological/behavioral disorder will have an effect on their Interaction, Under Sensitivity for Sounds, Smells and touching senses, Over sensitivity, Social Skills, Repetitive and Restrictive Behaviors. Autism Spectrum Disorder(ASD) can be seen in children under or above age of 2 Years andcontinues to be in Adolescent and Adulthood age which is highly heritable. Cause for this disorder might be genetic susceptibility and environmental factors which is said to be a "Behavioral Disease". Symptoms of this disease include Fever, Surgery, Age, Speech, etc. We have applied three algorithms Random Forest, KNN and Naïve Bayes Classifier to detect Autism Spectrum Disorder in Children. Random Forest Algorithm is our applied algorithm which boosts the model and gives the best result based on accuracy and Root Mean Square Error.

Keywords—Autism Spectrum Disorder, Machine Learning, Classification, KNN, Naïve Bayes , Random Forest.

I. INTRODUCTION

Autism spectrum disorder(ASD) is mostly seen in children and it has become more prevailing among young generations in the recent years. Autism Spectrum Disorder (ASD) which is a developmental disability disorder is caused by the difference in the functionality of brain. People with ASD might be different in communicating, interacting, behaving and learning. Signs/Symptoms of ASD begins during childhood and will usually exist for entire life of a person.

Autism is carried out clinically according to DSM-V standards to classify the disorder. These standards are coined/named by some US Mental health professionals by considering their successful diagnostic experience. ADI-R and ADOS are some basic behavioral tests conducted by doctors to detect symptoms of autism. These experiments are carried out in laboratory conditions with clinical atmosphere by certified professionals. Severity of autism in patients is detected based on the combined scores. ASD dataset comprising of 11 attributes which are related to behavior is taken for performing the classification of ASD and Non-ASD patients.

2. LITERATURE SURVEY

Machine Learning plays a significant role among the areas of Artificial Intelligence (AI) [1]. During recent years, Machine Learning (ML) has been attracting many researchers, and it has been successfully applied in many fields such as medical, education, forecasting etc., Right now, the diagnosis of diseases is mostly from expert's decision. Diagnosis is a major task in clinical science as it is crucial in determining if a patient is having the disease or not. This in turn decides the suitable path of treatment for disease diagnosis. Applying machine learning techniques for disease diagnosis using intelligent algorithms has been a hot research area of computer science.

Vaishali R, Sasikala R. have proposed a method to identify Autism with optimum behavioral data sets. In this research paper, they used a dataset with 21 features obtained from UCI machine learning repository which is experimented/conducted with binary firefly feature selection wrapper which is based on swarm intelligence.

The alternative hypothesis of this experiment claims that a machine learning model can achieve a better and accurate classification output with minimum featured subsets [2]. Using Swarm intelligence based single-objective binary firefly feature selection wrapper it is found that 10 features among 21 features of ASD dataset are sufficient to distinguish between ASD and non- ASD patients.

3. METHODOLOGY

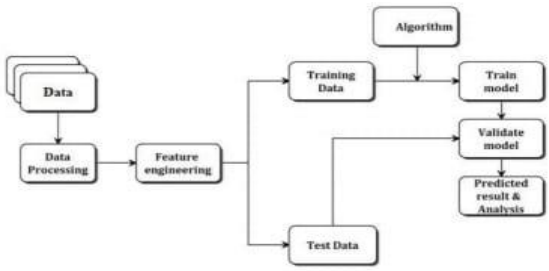

Fig 1: Dataflow Diagram

All the fields of the figure 1 are explained below,

A. Dataset

A dataset is a collection of tabular data, in which every column represents a variable, and each row represents a given record in the dataset. The dataset will have a list of the values for each members and variables of the dataset. Each value is called as datum. Datasets can also consist the collection of documents/files. Here, the Dataset is collected from Kaggle website.

B. Data Processing

Data processing will be occurred when data is collected and translated into useful information. Usually Data Processing is conducted by data scientists. it is crucial to perform data processing accurately in order to prevent the effect on the end product or output.

C. Training Data

The training data is an initial dataset which is used to make a program to understand how to apply technologies, such as neural networks to process, learn and produce results. Here in our work, we have used 80% of the data as training data.

D. Test Data

Testing data is a specifically identified data for testing purpose. Testing data may be produced in a focused way or by using other less-focused ways. Testing data may be produced by the tester, or by a program that aids the tester. Testing data may be recorded for re-use. Here in our work 20% of the data is used as the testing data.

E. Predicting Result & Analysis

Predictive analytics is to predictions about future events which is unknown, which is a branch of advanced analytics. Predictive analytics is using many techniques to analyze current data. some of them are data mining, statistics, modeling, machine learning and artificial intelligence The results are reported. The analysis section is where the writer describes about how the data is found.

DATASET

	A	B	C	D	E	F	G	H	I	J	K	L
1	A1_Score	A2_Score	A3_Score	A4_Score	A5_Score	A6_Score	A7_Score	A8_Score	A9_Score	A10_Score	age	target
2	0	0	0	0	0	0	1	1	0	1	11	0
3	1	1	0	0	0	1	1	0	0	0	16	1
4	1	0	0	0	0	0	1	1	0	1	16	1
5	1	1	1	1	1	1	1	1	1	1	3	1
6	1	1	0	1	1	1	1	1	1	1	3	1
7	1	1	0	0	1	1	1	1	1	1	2	1
8	1	0	0	1	1	1	0	0	1	0	14	1
9	0	1	0	0	1	0	1	1	1	1	14	1
10	0	0	0	0	0	0	1	0	0	1	16	0
11	1	1	1	0	1	1	0	1	1	1	1	1
12	1	0	0	1	0	1	1	0	1	1	16	1
13	1	1	1	1	0	1	1	1	0	1	17	1
14	0	0	0	0	0	0	0	0	0	0	7	0
15	1	1	1	1	0	0	1	0	1	1	15	1
16	0	0	0	0	0	0	0	0	0	0	9	0
17	1	1	1	0	1	0	1	1	0	1	12	1
18	0	0	0	0	0	0	0	0	0	0	16	0
19	1	1	1	0	1	1	1	1	0	1	12	1
20	1	0	0	0	1	0	0	0	0	1	10	0
21	1	1	1	0	1	0	1	1	0	1	12	1
22	1	0	0	1	1	1	1	1	1	0	16	1
23	1	0	1	1	1	1	1	0	1	0	16	1
24	1	0	1	1	0	1	0	1	1	1	16	1
25	1	1	1	0	1	1	0	1	1	0	16	1

Autism dataset

Fig 2: Dataset [9 – 13]

In this work, ASD screening dataset has been used for model building, training and empirical evaluation. The two separate datasets are considered from an open-access database of nearly 1758 children/toddlers recorded and given by Fadi Thabtah [3– 8].

There are total 12 attributes recorded for the autism spectrum disorder prediction which are – A1_Score, A2_Score, A3_Score, A4_Score, A5_Score, A6_Score, A7_Score, A8_Score, A9_Score, A10_Score, age, target as shown in the figure 2 above.

A1 score- He/ She often noticed small voices.

A2 score- He/ She might be usually concentrating on the entire image rather than the small details .

A3 score- Conversations of different people can easily be tracked in social groups.

A4 score- He/ She felt going back and forth between various activities was easy.

A5 score- He/ She doesn't know how to maintain conversations with friends

A6 score- He/ She is good at social chatting

A7 score- He/ She might find difficult to know the character's intention or feeling when He/ She reads a story.

A8 score- When He/ She was in preschool, He/she always enjoys playing a game.

A9 score- He/ She might finds it easier to know what someone is thinking by looking at their faces.

A10 score- He/ She might think its difficult to make new friends.

4. SEQUENCE FLOW

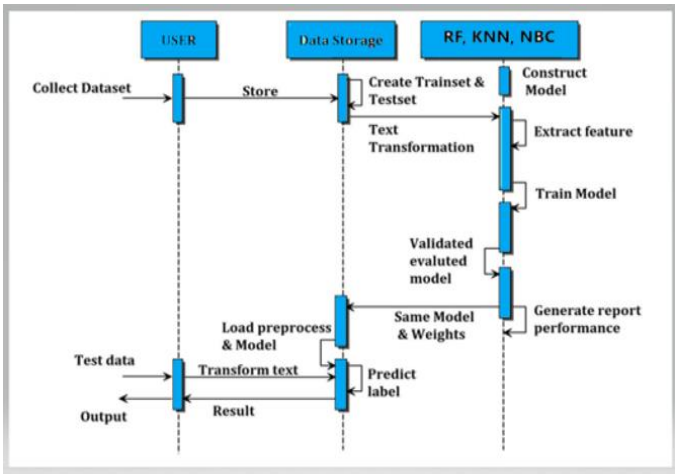

Fig 3: Sequence Diagram

A sequence diagram shows in the figure 3 explains how the processes work and communicate with another process and in which order they work.

Here, the dataset is collected from the user and it is stored [14 – 15]. The data in the dataset is tested and trained. Then various algorithms such as KNN, Random Forest, Naive Bayes are applied. The model is trained, evaluated and validated. The report of the performance is generated. The same weights and model are sent to the storage. Then the model and processes are loaded. The result is showed as output.

5. ALGORITHMS

A. *Random Forest Algorithm*

Random forest is an group method of learning for classification, regression and other tasks that are operated by constructing multitude of decision trees during training time and to give output as average prediction of individual trees. Random forest is an supervised learning algorithm. As in the name, the "forest", it builds a group of decision trees, which will usually be trained with a method called “bagging”. The general idea behind bagging method is the learning models combination that helps to increase the overall result. In this case, the model takes into consideration the A1 to A10 Scores, Age attributes and helps us to predict (Autism) Target of every patient. Here we got the accuracy of 97.63.

B. *K Nearest Neighbours Algorithm*

K-Nearest Neighbors (KNN) algorithm is one of the simpler algorithms used in Machine Learning to perform classification and regression problems. Based on the measures of similarity KNN algorithm will use data and classifies the new data points.

In this case, the model takes A1 to A10 Scores, Age attributes into consideration and helps us to predict Autism of every patient. Here we got the accuracy of 96.20.

C. *Naïve Bayes classifier Algorithm*

Naïve Bayes classifier algorithm is a supervised learning algorithm. It is based on Bayes theorem. This algorithm is used to solve problems on classification. It is mainly used to classify the text which includes an high-dimensioned training dataset. It is one amongst the simpler and effective algorithms on classification that helps to build the fastest machine learning models which are able to make quick predictions. It is also called as a probabilistic classifier because it predicts on the basis of object’s probability. In this case, the model takes A1 to A10 Scores, Age attributes into consideration and helps us to predict Autism of every patient. Here we got the accuracy of 94.63.

6. RESULTS AND EVALUATION

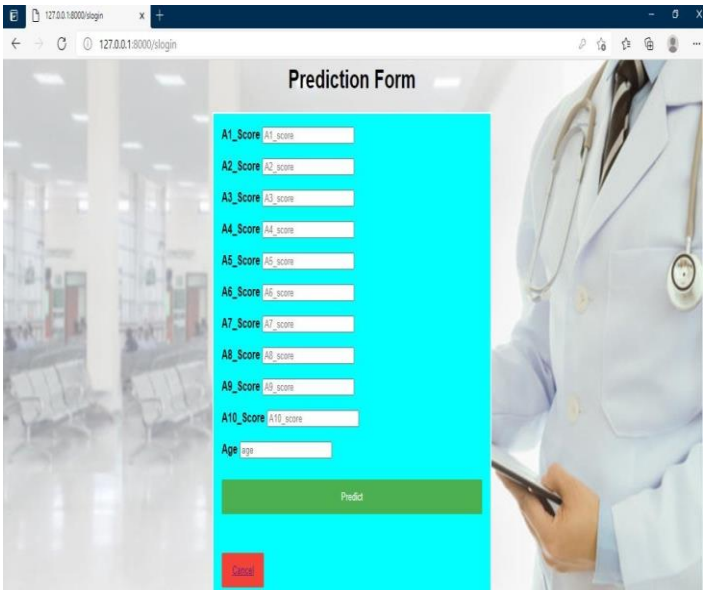

Fig 4: Prediction Form asked in Front End

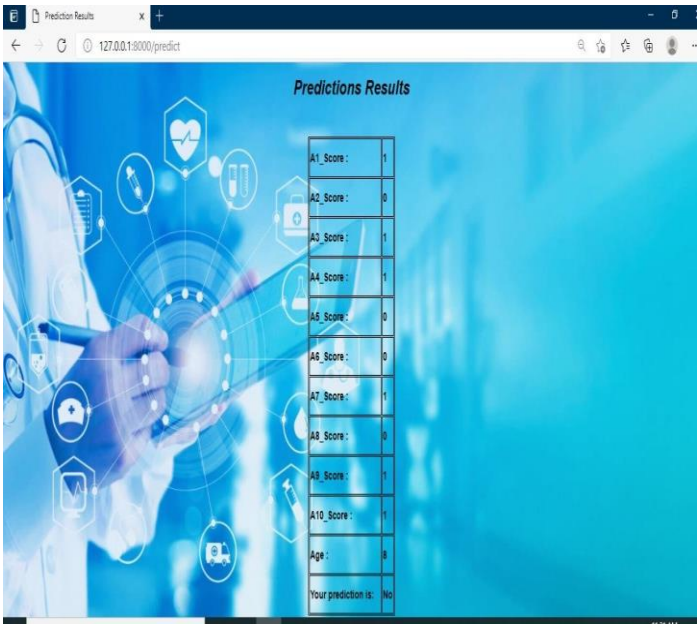

Fig 5: Predicted Results Displayed on Front End

Above figures, fig 4 and fig 5 represents the front-end pictures of our application.

```
In [27]: #plot for algorithms comparision
         import seaborn as sns

         algorithms=['KNN','Random Forest','Naive bayes']
         RMSE=[19.47,15.39,23.15]
         sns.barplot(algorithms,RMSE)

Out[27]: <matplotlib.axes._subplots.AxesSubplot at 0x27d50816308>
```

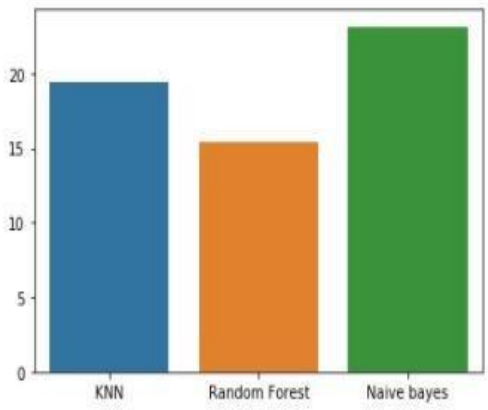

Fig 6: Algorithm's comparison based on RMSE value

```
In [23]: #plot for algorithms comparision
         import matplotlib.pyplot as plt

         algorithms=['KNN','Random Forest','Naive bayes']
         accuracy=[96.2,97.63,36.51]
         plt.plot(algorithms,accuracy)
         plt.xlabel('algorithms')
         plt.ylabel('Accuracy')
         plt.title('Algorithm comparison based on Accuracy')

Out[23]: Text(0.5, 1.0, 'Algorithm comparison based on Accuracy')
```

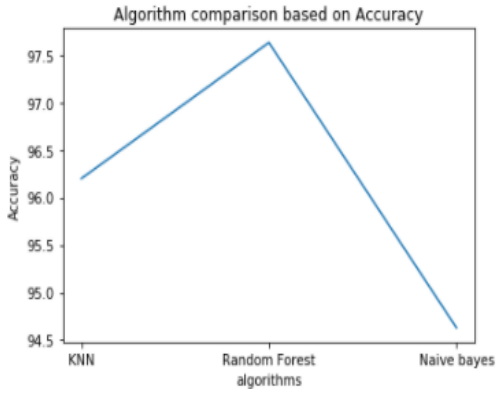

Fig 7: Algorithm's comparison based on Accuracy

By applying Random Forest algorithm, K-Nearest Neighbors (KNN) algorithm and Naïve Bayes Classifier algorithm we found out that Random Forest gives the root mean square error of 15.39372679517698 and has the accuracy of 97.6303317535545, K-Nearest Neighbors (KNN) gives the root mean square error of 19.47169534044942 and has the accuracy of 96.20853080568721 and Naïve Bayes Classifier gives the root mean square error of 23.15766833089332 and has the accuracy of 94.63722397476341. Hence Random forest is more appropriate for this model. The same is shown in the above figures fig 6 and fig 7.

6. CONCLUSION

A person with Autism Spectrum Disorder(ASD) will need an early diagnosis and treatment. As early the ASD is diagnosed, that better the results will be in long term. This paper, is based on the behavioral dataset in which we adopted different algorithms of machine learning to find the accurate models for each of the datasets.

By analyzing the data collected by us, we were able to create an accurate model which could help doctors/nurses to detect the autism spectrum disorder of patients based on physiological/behavioral factors, here we have considered A1 to A10 Scores, Age and found that Random Forest model seems to yield the highest accuracy and its feature sets performed better than other algorithms.

ACKNOWLEDGMENT

We would like to thank our institute K S School of Engineering and Management (KSSEM) for providing us this opportunity and for guiding us to make this possible.

REFERENCES

[1] E. Leblanc, P. Washington, M. Varma, K. Dunlap, Y. Penev et al., "Feature replacement methods enable reliable home video analysis for machine learning detection of autism," Scientific Report, vol.10, no. 21245, pp. 1–11, 2020

[2] M.D. Hossain, M.A. Kabir, Detecting child autism using classification techniques., Studies in health technology and informatics 264, 1447 (2019). https://doi.org/10.3233/shti190477

[3] Fadi Thabtah, Firuz Kamalov and Khairan Rajab, "A new computational intelligence approach to detect autistic features for autism screening," International Journal of Medical Informatics, vol. 117, pp. 112-124, sep 2018. https://ur.booksc.eu/book/81952624/1299c4

[4] F. Thabtah, "autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment," in ICMHI '17: Proc. of the 1st Int. Conf. on Medical and Health Informatics 2017, Taichung, Taiwan, pp. 1–6, 2017. https://doi.org/10.1145/3107514.3107515

[5] F. Thabtah, “Machine learning in autistic spectrum disorder behavioral research: A review and ways forward,” Informatics for Health and Social Care, vol. 44, no. 3, pp. 278–297, 2019.

[6] F. Thabtah, F. Kamalov and K. Rajab, “A new computational intelligence approach to detect autistic features for autism screening,” International Journal of Medical Informatics, vol. 117, pp. 112–124, 2018. https://doi.org/10.1016/j.procs.2020.03.399

[7] Fadi Thabtah, “Autistic Spectrum Disorder Screening Datasets,” UCI machine learning repository, 2017. [Online]. Available: https://archive.ics.uci.edu/ml.

[8] Thabtah, Fadi. "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. (2018) " Informatics for Health and Social Care: 1-20. https://doi.org/10.1016/j.procs.2020.03.399

[9] K. S. Ramana, M. S. Lakshmi and M. Janardhan, “Machine learning based novel autism spectrum disorder screening,” Turkish Journal of Computer and Mathematics Education, vol. 12, no. 3, pp. 4866–4879, 2021.

[10] H. S. Alarifi and G. S. Young, “Using multiple machine learning algorithms to predict autism in children,” in Proc. of Int. Conf. Artificial Intelligence, New York, NY, United States, pp. 464–467, 2018.

[11] S. M. Lundberg and S. Lee, “A unified approach to interpreting model predictions,” in NIPS’17: Proc. of the 31st Int. Conf. on Neural Information Processing Systems, New York, United States, pp. 4768–4777, 2017.

[12] M. T. Ribeiro, S. Singh and C. Guestrin, “Why should I trust you?” explaining the predictions of any classifier,” in KDD ‘16: Proc. of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, New York, United States, pp. 1135–1144, 2016.

[13] H. S. Nogay and H. Adeli, “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, 2020.

[14] A. Zunino, P. Morerio, A. Cavallo, C. Ansuini, J. Podda, F. Battaglia, E. Veneselli, C. Becchio, V. Murino, in 2018 24th International Con1wference on Pattern Recognition (ICPR), vol. 2018-August (IEEE, 2018), vol. 2018- August, pp. 3421–3426.

[15] Abbas, H., Garberson, F., Glover, E., & Wall, D. P. (2017). Machine learning approach for early detection of autism by combining questionnaire and home video screening. arXiv preprint arXiv:1703.06076.