Artism and Non-Artism by using Machine Learning

Mallavarapu Swetha Reddy

Department of Computer Science, Sri Venkateswara University, Tirupati

***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 project, 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. KNN algorithm shows maximum accuracy and speed to diagnosis.*

I. 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. 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 [1]. Detecting autism earlier in one life can make a big difference than treating it later. 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 [2]. 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. According to WHO every year among 160 children, one is diagnosed with ASD traits all over the world [3]. Treating ASD earlier is always the best option for toddlers as they are still developing. Against this huge burden, only 200 psychiatrists and limited professionals are serving. 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. 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.

Artificial intelligence (AI) is the ability of a computer program or a machine to think and learn. It is also a field of study which tries to make computers "smart". As machines become increasingly capable, mental facilities once thought to require intelligence are removed from the definition. AI is an area of computer sciences that emphasizes the creation of intelligent machines that work and reacts like humans. Some of the activities computers with artificial intelligence are designed for include: Face recognition, Learning, Planning, Decision making etc., Artificial intelligence is the use of computer science programming to imitate human thought and action by analysing data and surroundings, solving or anticipating problems and learning or self-teaching to adapt to a variety of tasks.

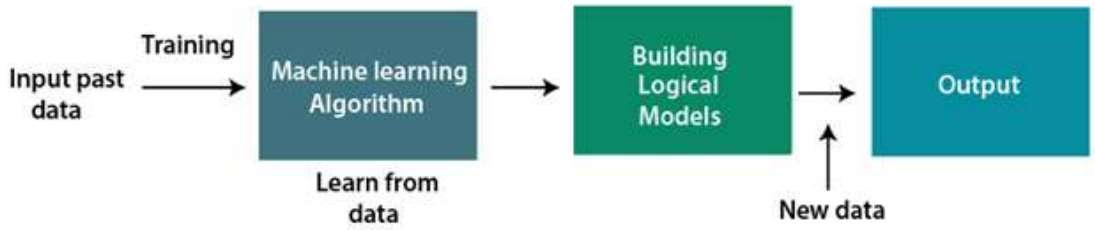
1.1. Machine Learning

Machine learning is a growing technology which enables computers to learn automatically from past data. Machine learning uses various algorithms for building mathematical models and making predictions using historical data or information. Currently, it is being used for various tasks such as image recognition, speech recognition, email filtering, Facebook auto-tagging, recommender system, and many more. Machine Learning is said as a subset of artificial intelligence that is mainly concerned with the development of algorithms which allow a computer to learn from the data and past experiences on their own. The term machine learning was first introduced by Arthur Samuel in 1959. We can define it in a summarized way as:

"Machine learning enables a machine to automatically learn from data, improve performance from experiences, and predict things without being explicitly programmed".

A Machine Learning system learns from historical data, builds the prediction models, and whenever it receives new data, predicts the output for it. The accuracy of predicted output depends upon the amount of data, as the huge amount of data helps to build a better model which predicts the output more accurately.

Suppose we have a complex problem, where we need to perform some predictions, so instead of writing a code for it, we just need to feed the data to generic algorithms, and with the help of these algorithms, machine builds the logic as per the data and predict the output. Machine learning has changed our way of thinking about the problem. The below block diagram explains the working of Machine Learning algorithm:



1.1.1. Features of Machine Learning:

- Machine learning uses data to detect various patterns in a given dataset.
- It can learn from past data and improve automatically.
- It is a data-driven technology.
- Machine learning is much similar to data mining as it also deals with the huge amount of the data.

1.1.2. Classification of Machine Learning

At a broad level, machine learning can be classified into three types:

- Supervised learning
- Unsupervised learning
- Reinforcement learning

Supervised Learning

Supervised learning is a type of machine learning method in which we provide sample labeled data to the machine learning system in order to train it, and on that basis, it predicts the output.

The system creates a model using labeled data to understand the datasets and learn about each data, once the training and processing are done then we test the model by providing a sample data to check whether it is predicting the exact output or not.

The goal of supervised learning is to map input data with the output data. The supervised learning is based on supervision, and it is the same as when a student learns things in the supervision of the teacher. The example of supervised learning is **spam filtering**.

Supervised learning can be grouped further in two categories of algorithms:

- **Classification**
- **Regression**

Unsupervised Learning

Unsupervised learning is a learning method in which a machine learns without any supervision. The training is provided to the machine with the set of data that has not been labeled, classified, or categorized, and the algorithm needs to act on that data

without any supervision. The goal of unsupervised learning is to restructure the input data into new features or a group of objects with similar patterns.

In unsupervised learning, we don't have a predetermined result. The machine tries to find useful insights from the huge amount of data.

It can be further classifieds into two categories of algorithms:

- **Clustering**
- **Association**

1.2 KNN(K-nearest neighbor)

The K-nearest neighbors algorithm is a supervised classification algorithm method. It classifies objects dependant on nearest neighbor. It is a type of instance-based learning. The calculation of distance of an attribute from its neighbors is measured using Euclidean distance. It uses a group of named points and uses them on how to mark another point. The data are clustered based on similarity amongst them, and is possible to fill the missing values of data using K-NN. Once the missing values are filled, various prediction techniques apply to the data set. It is possible to gain better accuracy by utilizing various combinations of these algorithms.

- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
- K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data.
- It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.
- KNN algorithm at the training phase just stores the dataset and when it gets new data, and then it classifies that data into a category that is much similar to the new data.

II. LITERATURE REVIEW

[1] TITLE: Autism Classification Based On Logistic Regression Model

AUTHOR NAME: Yuanrui Zheng

DESCRIPTION: Autistic Spectrum Disorder (ASD) is a developmental disability can affect communication and behavior. Existing research has shown that early diagnosis can help doctors to find this disease early and can save significant healthcare costs. With the rapid growth of ASD cases, a ASD related dataset created for scientists and doctors to investigate this disease. Autistic Spectrum Disorder Screening Data for Adult is a well-known dataset, which contains 20 features to be utilized for further analysis. This article developed and test an Autism classification algorithm which based on logistic regression model. The result of this study provided a model can predict the ADS in an average F1 score of 0.97, which displays the superiority of proposed model. Besides, the data visualization part displays several feature distribution images for people to better understand the data and related feature engineering.

[2] TITLE: Autism Spectrum Disorder Detection Using Machine Learning Approach

AUTHOR NAME: Nabila Zaman

DESCRIPTION: It is always a complex procedure to diagnosis autism spectrum disorder (ASD) because there is no particular medical test for autism, like a blood test, to make a diagnosis for the disorder. Autism spectrum disorder is defined by the disability and impairments of participating in social communication or the presence of restricted or repetitive behaviors, or both. It is a behaviorally diagnosed condition. To make a diagnosis, doctors look at the child's developmental history and behavior. Apparently, most children do not attain a proper diagnosis for autism until it is too late. In some cases,

parents are reluctant to accept that their child's mental growth is not developing along with the child's physical growth. This lateness in diagnosis hinders a child's ability to get the help they need to keep developing. It is important to diagnose ASD as early as possible through monitoring, screening, and evaluating a child's development so that we can ensure proper care and support for an autistic child to help them reach their full potential. So, we are developing a system that will have the ability to diagnosis Autism and come up with a reliable and effective conclusion even without the help of a professional. We hope this system will be very helpful for those concerned parents who are worried about their child's growth and activities, at the time same time it will be very useful for the professionals.

[3] TITLE: https://ieeexplore.ieee.org/abstract/document/1017309/ Predictive Analysis of Autism Spectrum Disorder (ASD) using Machine Learning

AUTHOR NAME: Naurin Farooqi

DESCRIPTION: Autism Spectrum Disorder (ASD) is a severe condition related to brain development that impairs communication and interaction with others. It is a "developmental disorder" because symptoms generally appear in the first two years of life or later. Till that age, children behave pretty similar to typically developing individuals. Tracking the symptoms leading to autism is challenging because of the wide variation in the type and severity of the disorder. One of the pitfalls is that many countries, including Pakistan, do not track or report their autism cases. In this paper, we have shown how we performed predictive analysis after collecting data from the only government-run Child Psychiatry Department in Pakistan, Mayo Hospital, Lahore. We also collected data from a private school made explicitly for children with autism in Lahore. Data of a total of 100 autistic children under the age of 12 was collected. To avoid any class imbalance problem, we also collected data of 100 healthy children. After performing various machine learning algorithms on our data, our model managed to classify children with autism with an accuracy of 95%. To the best of our knowledge, no such work has been done regarding autism in Pakistan before. This research aims to create a supervised machine learning model that can classify ASD individuals in much less time and effort compared to currently practiced diagnostic procedures and help us visualize common causes leading up to it so that appropriate lifestyle changes can be made accordingly. In the future, the number of records can be increased, and accuracy can be maximized even further. Our model can be used by various physicians and parents of children with ASD for early prediction and for visualizing common causes leading up to it.

[**4] TITLE:** Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques

AUTHOR NAME: Kaushik Vakadkar

DESCRIPTION: Autism Spectrum Disorder (ASD) is a neurological disorder which might have a lifelong impact on the language learning, speech, cognitive, and social skills of an individual. Its symptoms usually show up in the developmental stages, i.e., within the first two years after birth, and it impacts around 1% of the population globally. ASD is mainly caused by genetics or by environmental factors; however, its conditions can be improved by detecting and treating it at earlier stages. In the current times, clinical standardized tests are the only methods which are being used, to diagnose ASD. This not only requires prolonged diagnostic time but also faces a steep increase in medical costs. To improve the precision and time required for diagnosis, machine learning techniques are being used to complement the conventional methods. We have applied models such as Support Vector Machines (SVM), Random Forest Classifier (RFC), Naïve Bayes (NB), Logistic Regression (LR), and KNN to our dataset and constructed predictive models based on the outcome. The main objective of our paper is to thus determine if the child is susceptible to ASD in its nascent stages, which would help streamline the diagnosis process. Based on our results, Logistic Regression gives the highest accuracy for our selected dataset.

[5] TITLE: Identification of Autism Spectrum Disorder through Feature Selection-based Machine Learning.

AUTHOR NAME: Marian Binte Mohammed.

DESCRIPTION: Autism Spectrum Disorder (ASD) is a developmental disability that is likely to be perceived at a young age, persisting throughout a lifetime. The goal of this study is to detect ASD more efficiently with the use of Machine Learning methods. In our paper, we worked with the AQ-10 Adult dataset. Multiple steps have been taken to perform the data preprocessing. We have used different data synthesization techniques and a few feature selection techniques and eventually implemented them with other classifiers. Although throughout our analysis, we can see that the usage of Neural Network has some significant effect due to a smaller data set, the best-performance was provided by the combination of classifiers and feature selection methods to develop the prediction model. After evaluation, We deduced that a model with Principal Component Analysis (PCA) feature selection method using the AdaBoost classifier gave the best results.

III. PROBLEM STATEMENT

- In recent years, machine learning algorithms (ML) has been successfully applied in many engineering fields which include computer systems, vision, finance, hospital and medicine, transportation, telecommunications, heuristic classification, aviation, gaming, data mining, speech recognition, and heavy industry.
- The use of such powerful machine learning techniques in Artism prediction could result on early prediction that protects and avoid loss from cure the disease.
- In addition it will reduce the time that needed to find the disease manually.
- The Roc graph is plotted.

Disadvantage

- It will take too much resource for the processing.
- Take too much processing time.
- Accuracy is less for the prediction in testing.

IV. DEVELOPMENT PROCESS

4.1 Requirement Analysis and Specifications

The requirement engineering process consists of feasibility study, requirements elicitation and analysis, requirements specification, requirements validation and requirements management. Requirements elicitation and analysis is an iterative process that can be represented as a spiral of activities, namely requirements discovery, requirements classification and organisation, requirement negotiation and requirements documentation.

4.2 Input Requirement and Output Requirements

Input Design

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

Objectives

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.
2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.
3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus, the objective of input design is to create an input layout that is easy to follow

Output Design

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how

the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2. Select methods for presenting information.

3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

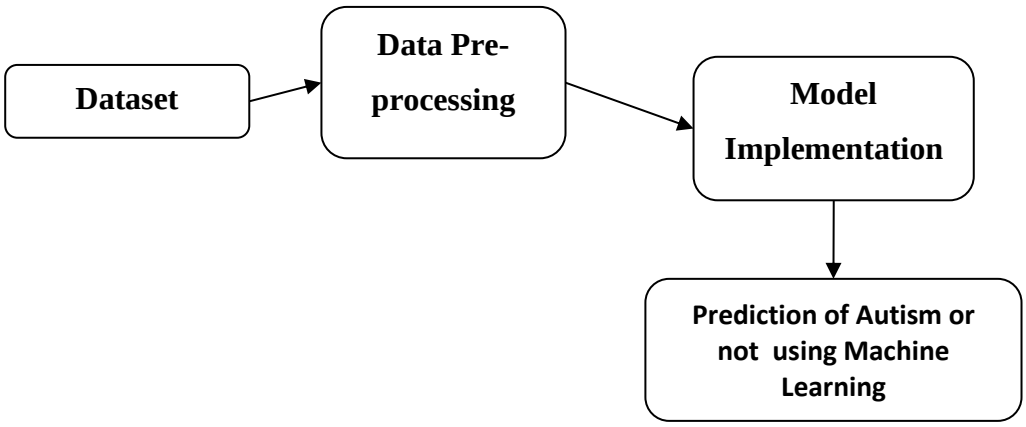
- ❖ Convey information about past activities, current status or projections of the
- ❖ Future.
- ❖ Signal important events, opportunities, problems, or warnings.
- ❖ Trigger an action.
- ❖ Confirm an action.

V. PROPOSED METHOD

- The main contribution for this project. The one is to apply powerful and popular machine learning techniques of KNN classifier to classify and detect the person has Autism or not.
- Supervised learning algorithms are applied to diagnosis whether children meet the symptoms for ASD.

Advantages

- An improved version of Machine learning parameters was proposed in this proposed method to improve the accuracy of classification.
- Based on the experimental results, our method was superior to the techniques described in the conventional methods.



Modules Used

1. Datasets Collection.
2. Pre-processing.
3. Model Implementation.
4. Classification.

1. Datasets Collection:

A data set (or dataset) is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question.

A Collection of data is called datasets. Here, the datasets collection is image file.

2. **Preprocessing:**

Pre-processing routines prepare the data for analysis. Before we start the actual processing, the data has to be pre-processed to remove the detector effects. Preprocessing is the most important aspect of data processing. Hence, data filtering, data ordering, data editing and noise modeling play an important role in any data preprocessing.

A Pre-Processing is one of the techniques is used to reduce the noise in the file format of the system. Smoothing and de-trending are processes for removing noise and linear trends from data, while scaling changes the bounds of the data. Grouping and binning methods are techniques that identify relationships among the data variables.

3. Model Implementation:

KNN(K-nearest neighbor)

The K-nearest neighbors algorithm is a supervised classification algorithm method. It classifies objects dependant on nearest neighbor. It is a type of instance-based learning. The calculation of distance of an attribute from its neighbors is measured using Euclidean distance. It uses a group of named points and uses them on how to mark another point. The data are clustered based on similarity amongst them, and is possible to fill the missing values of data using K-NN. Once the missing values are filled, various prediction techniques apply to the data set. It is possible to gain better accuracy by utilizing various combinations of these algorithms.

4. Classification:

Classification neural networks used for feature categorization are very similar to fault-diagnosis networks, except that they only allow one output response for any input pattern, instead of allowing multiple faults to occur for a given set of operating conditions. Classification is a process related to categorization, the process in which ideas and objects are recognized, differentiated and understood the system. Classification is a term used both about the process to classify the condition of the fault wire. A classification process can be obtained to training the network of the system. Here, classification techniques can be obtained by using Machine Learning model to classify the result of the model. This project finally predicts whether the person has artism or not.

VI. FUTURE ENHANCEMENT

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 early autism symptoms effortlessly so they can seek professional help if needed. 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.

VII. CONCLUSION

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 are set in a way to maintain their privacy. Using the dataset from Q-CHAT and AQ tools, our proposed model can predict using KNN higher accuracy in case of toddlers. Algorithms which are supervised are selected to run our dataset after preprocessing it. This outcome showed better performance compared to the others. The limitations on the characteristics of design or methodology that impacted or influenced the interpretation of our project, the prime candidate is the lack of dataset. This modeling error occurred because a function corresponds too closely to a particular set of data. 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.

REFERENCES

[1] 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.

[2] 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.

[3] P. Mesa-Gresa, H. Gil-Gomez, J.-A. Lozano-Quilis, and J.-A. Gil-G ´ omez, ´ “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.

[4] 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.

[5] N. DONGES, “A complete guide to the random forest algorithm,” Built In, vol. 16, 2019.

[6] E. Solutions and . Name, “Accuracy, precision, recall amp; f1 score: Interpretation of performance measures,” Nov 2016. [Online].

[7] P A Filipek, P J Accardo, G T Baranek et al., "The screening and diagnosis of autistic spectrum disorders[J]", Journal of autism and developmental disorders, vol. 29, no. 6, pp. 439-484, 1999.

[8] L T Curtis and K. Patel, "Nutritional and environmental approaches to preventing and treating autism and attention deficit hyperactivity disorder (ADHD): a review[J]", The Journal of Alternative and Complementary Medicine, vol. 14, no. 1, pp. 79-85, 2008.

[9] C Gold, T Wigram and C. Elefant, "Music therapy for autistic spectrum disorder[J]", Cochrane Database of Systematic Reviews, no. 2, 2006.

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