



A REVIEW ON CLASSIFICATION TECHNIQUES WITH AUTISM SPECTRUM DISORDER AND AGRICULTURE

Premasundari M

Research Scholar

Department of Computer Science

Sri Ramakrishna College of Arts and Science for Women

Coimbatore, India

Dr. C.Yamini

Associate Professor

Department of Computer Science

Sri Ramakrishna College of Arts and Science for Women

Coimbatore, India

Abstract: Agriculture is the backbone of India. The ultimate goal of farming is the cultivation and perfection of human beings. Agriculture acts as a mediator between nature and the human community. Soil is the boon for agriculture and chemical fertilizer is the curse for soil. Food produced by these artificial fertilizers causes many challenges to human lives. Autism Spectrum Disorder is the third most common developmental disorder. The first ever survey shows that more than 10 million children in India suffer from autism. Data Mining is one of the phases involved in Knowledge Discovery in Databases. The goal of Data mining is the discovery of patterns and knowledge from large datasets. There are a number of machine learning techniques such as classification, clustering, association rule etc. involved in data mining. Classification belongs to supervised learning technique of data mining. This research work focuses on finding whether agriculture has any impact on autism. This paper shows the literature review based on the significance of prediction and classification based data mining algorithms in the field of agriculture and autism spectrum disorder.

Keywords: Classification, Autism, Crop yield, Rainfall Prediction, Support Vector Machine, J48, REPTree, Neural Network, Fuzzy Logic

I. INTRODUCTION

Data mining is an interdisciplinary subfield of computer science and deals with the extraction of useful information from the huge amount of data. There are various data mining techniques to discover information and use the same for prediction purposes. Data mining technique is classified into two groups, namely supervised and unsupervised learning. Supervised learning is the data mining task of inferring a function of labelled trained data. Supervised learning is a type of machine learning algorithm that uses a known dataset to make predictions. [1] In supervised learning there are two categories of algorithms namely

- Classification: for categorical response values.
- Regression: for continuous response values.

Data is of two types:

- Numerical Data: In statistical research, Numerical Data is also called Quantitative. This type of data is further divided into two ways:
 - Discrete: Discrete values are countable and it can take only certain values. For example, Total Number of students in a class.
 - Continuous values are measurable and it can take any values within a range. For example, Height and Weight of a Person. Regression can handle these values.
- Categorical Data: It is the data collected in Yes/No fashions. It is grouped into some sort of category like grouping by sex, hair color etc. Some numerical data are considered as categorical data eg. Postal codes, Birth dates etc. Classification can handle these values. This is also known as qualitative data in terms of statistics.

A. Classification

Classification is one of the data mining techniques which intend to analyze the large datasets effectively. Classification involves finding rules that partition the data into disjoint groups. It analyses the training data and constructs a model based on the class label and assigns a class label to the future unlabeled records [2]. The Classification process involves two phases:

- To build classifier or model
- To use classifier for classification.

The classifier is built from the training data set using classification algorithms and the classifier is used for classification [3]. Classification classifies the dataset into predefined groups and building a classification model or classifier to predict the future behavior. Some of the classification algorithms are Support Vector Machines (SVM), Naïve Bayes Classifiers, and Decision Trees etc. Some of the regression algorithms include Linear regression, Nonlinear regression etc.

B. Autism Spectrum Disorder

Autism is the neurodevelopmental delay that begins within the first three years of the child. Autism Spectrum Disorder (ASD) is a group of developmental disorders. Each person has different symptoms. There are three different types of autism spectrum disorder. They are:

Autistic Disorder

Autistic Disorder is the most common type of ASD. This type includes symptoms like delay in speech, difficulties with social communication, impaired intellectual abilities and unusual behavior. This is also known as Classic autism.

Asperger Syndrome

This type of autism includes some of the mild symptoms of autistic disorder such as social interaction and unusual behavior, but not delays in speech and impaired intellectual abilities. Hans Asperger found this in grownup children who generally do not have delays in speech and impaired intellectual abilities. So this type of autism is called Asperger syndrome.

Pervasive Developmental Disorder (PDD)

PDD is also referred to as typical autism. Persons who diagnosed with PDD meet some of the symptoms for either autistic disorder or Asperger syndrome but not all.

C. Agriculture

Agriculture is the science of farming, growing the crops, rearing of animals to provide food, wool and other products. It plays a vital role in Indian economy. It is the principle means of livelihood in India. Indian Agriculture is one of the largest contributors to the Gross Domestic Product (GDP).

This paper summarizes a literature survey related to some of the classification techniques applied in various fields such as agriculture and medicine in Section II. Then a conclusion is drawn in the Section III.

II. LITERATURE REVIEW

This section provides a detailed discussion about classification techniques and discusses the various works of several authors.

In paper [4], the vital goal of the paper is to review the autism problem, to detect the levels of autism with the help of data mining, classification algorithms and also to study the autism student's performance using data mining methodologies. The autism students' prediction is influenced by various factors like language, social, behavior and the levels of autism is predicted. 100 samples were taken for the implementation using Weka, an open source tool. The classify panel permits the user to use classification algorithms to the dataset and it is used to estimate the accuracy of the resulting autism students' predictive model, and to visualize the model. Support Vector Machine, Neural Network and Fuzzy Logic algorithms were enforced in Weka. 10 fold cross validation is chosen for evaluation.

Mohana E et al., [5] taken Autism Dataset which is formed with M-CHAT-R tool (Modified Checklist for Autism in Toddlers Revised tool) that is valid only for children in the age of 16 to 30 months. The author has applied four feature selection algorithms such as Fisher filtering, ReliefF, Runs filtering and Stepdisc to filter relevant feature from the dataset, and then applied several classification algorithms on the reduced features. Evaluation on performance is made by the author based on the results such as recall, error rate and accuracy and found BVM, CVM and MLR which produced high accuracy of 95.21% using Runs Filtering as the best classification algorithms applied on the dataset.

Sumi Simon et al., [6] performed an empirical evaluation on autistic data and used many classification algorithms like SVM, J48, MLP and IB1 on the behavioural

data to classify the autistic children. The author collected autistic data from National Institute for the Mentally Handicapped and extracted relevant features from the data using Relief and Runs Filtering (Feature Selection algorithms). Out of the four classifiers, SVM is the best classifier to classify the autistic children and it provides high accuracy and low error rate.

Mythili et al., [7] used Weka (Waikato Environment for Knowledge Analysis), an open source machine learning software tool for implementation. The decision tree classifier (J48) and Normalized Polykernel based classifier (SVM) are applied to the dataset in weka explorer. The learning skills of the autistic children are influenced by various factors like handwriting, spelling, read, write, language skills etc. Based on these factors, the rule set is extracted from the dataset by the decision tree classifier and the SVM is also applied to the dataset in the performance of autistic children. The classifier results are compared by time, accuracy, correctly classified instances and incorrectly classified instances. Finally the author concluded that SVM with normalized polykernel is a better classifier for analyzing the autistic children learning skill performance consuming less time with better accuracy, low error rate and highest percentage of correctly classified instances.

Gondy A. Leroy et al., [8] proposed a work to study therapy success with autistic children. Author used ID3 to build the decision tree and association rule to predict the high and low levels of appropriate or inappropriate behaviour and used 10 fold cross validation for evaluation options. The decision tree predicted 69% accuracy when high or low levels of appropriate behavior could occur before treatment (Pre Treatment) and increased accuracy (76%) during treatment and then decreased again after treatment (Post Treatment). Post treatment appropriate behavior cannot be predicted accurately with decision tree. The decision tree provided same accuracy in the levels of inappropriate behavior also. Since the decision tree does not provide accuracy in post treatment, association rule mining is used in the dataset. The rules found in association rule mining confirm the strongest results for predicted levels of appropriate behavior and also add details for predicting levels of inappropriate behavior [8].

Rajshekhkar Borate et al., [9] collected data from Mahatma Phule Krushi Vidyapith Rahuri for research work. The methodology is divided into two major parts namely clustering and classification. Clustering is performed among 20 districts in four ways. Cluster Type-1 is based on 5 environment variables like rainfall, temperature etc, cluster Type-2 on two biotic attributes such as soil pH and soil salinity, cluster Type-3 on irrigated area and cluster Type-4 based on individual crop yields. Rapid Miner Studio software is used to implement k-means clustering for each cluster type 1 to 4 for result and analysis. Prediction results were determined for yields of selected crops, according to the selected input attributes using appropriate classification and regression models in Rapid Miner.

D Ramesh et al., [10] presented a brief analysis of crop yield prediction using Multiple Linear Regression and Density Based Clustering technique for the East Godavari district of Andhra Pradesh in India. The estimated results using an MLR range between -14% and +13% for 40 years interval and using density based clustering technique, it

ranges between -13% and +8% for 6 cluster approximation. Finally, comparison is done between Exact production and Estimated values using MLR technique and Density-Based Clustering technique.

Zekarias Diriba et al., [11] proposed a study which uses data from the Ethiopian Economic Association (EEA). Three classifiers are employed namely J48, REPTree and Random Forest on the data. Out of total 90 attributes, only 13 attributes are selected. The data is experimented using N-fold cross validation for evaluation. Experiments are made using classifiers. The three algorithms show a slight difference with Performance evaluation and the first performed is REPTree, next is Random Forest and then J48. With the Efficiency evaluations, REPTree and Random Forest are more efficient than J48 in all measures except ROC.

M Kannan et al., [12] collected five years of data during three months September, October, November (winter Season) from the Statistical Department of Tamil Nadu, Chennai. Using this data, the values are computed for rainfall fall in the ground level using Karl Pearson

correlation coefficient and predicted rainfall for further three years by Multiple Linear Regression. Since predicting values lie below the computed values, the results are not accurate but approximate.

Hari Ganesh et al., [13] proposed a study on soil classification depending on the fertility class of the soil. Naïve Bayes, J48 and JRip algorithms are the classification techniques used for classification of soil with the help of Weka tool. These techniques are evaluated and compared on the basis of time, error rate and accuracy, True Positive rate and False Positive rate. The experiment used 10 fold cross validation for testing options. Finally, the experimental results shown that J48 algorithms worked out to be well defined classifier for soil samples.

III. INFERENCES

The following table has inferred from the research works related to Autism Spectrum Disorder and Agriculture.

Table I. Inferences from ResearchWorks

Sr. No.	Research Work	Data Mining Technique	Best algorithm	Data Source	Attribute
1	A Study on Autism Spectrum Disorders using Classification Techniques	Artificial Neural Network (perceptron), Support Vector Machine, and with Fuzzy logic	-	-	-
2	Categorizing The Risk Level Of Autistic Children Using Data Mining Techniques	Feature selection algorithm: Fisher filtering, ReliefF, Runs filtering and Stepdisc Classification algorithm: BVM, C4.5, C-RT, CS-CRT, CS-MC4, C-SVC, CVM, ID3, K-NN, Linear Discriminant Analysis, Multilayer perceptron, Naïve Bayes continuous, Multinomial Logistic Regression, PLSDA, PLS-LDA and Rnd Tree	BVM, CVM AND MLR	M-CHAT-R tool (Modified Checklist for Autism in Toddlers Revised tool)	Numerical (Discrete)
3	Empirical Evaluation of Data Mining Classification Methods for Autistic Children	SVM, J48, MLP and IB1 Probabilistic Neural Network (To classify pre-verbal vocalization of autistic children)	SVM	National Institute for the Mentally Handicapped	Numerical
4	A Novel Approach to Predict the Learning Skills of Autistic Children using SVM and Decision Tree	J48 SVM	SVM	Data from children database	-
5	Data Mining Techniques to Study Therapy Success with Autistic Children	ID3 (Decision Tree) Association Rule Mining	Association Rule	Claremont Autism Center	-
6	Applying Data Mining Techniques to Predict Annual	Clustering: K Means clustering	Based on the results,	-	Numerical (Continuous)

	Yield of Major Crops and Recommend Planting Different Crops in Different Districts in India	Classification and Regression Models: k-NN, ANN Linear Regression	recommendations are made with 3 possible crops for farming across all agricultural districts		
7	Analysis Of Crop Yield Prediction Using Data Mining Techniques	Multiple Linear Regression and Density Based Clustering technique		East Godavari district of Andhra Pradesh in India	Numerical (Continuous)
8	Application of Data Mining Techniques for Crop Productivity Prediction	J48, REPTree and Random Forest	REPTree and Random Forest	Ethiopian Economic Association (EEA).	-
9	Rainfall Forecasting Using Data Mining Technique	Karl Pearson correlation coefficient, Multiple Linear Regression	Results are not accurate, but approximate	Statistical department of Tamil Nadu, Chennai.	-
10	Data Mining Technique to Predict the Accuracy of the Soil Fertility	Naive Bayes, J48, JRip algorithms	J48	Agricultural Soil datasets	Numerical

IV. CONCLUSION AND FUTURE WORK

Most of the classification algorithms such as Decision tree, Naïve Bayes etc. can handle both numerical and categorical data. If these approaches do not work, ensemble techniques like Random Forest can be used for handling mixed data (both numerical and categorical attributes). The best classifiers dealt with autism spectrum disorder are Support Vector Machine and Multinomial Logistic Regression. Random Forest, REPTREE and J48 are the best classification algorithms for predicting crop yield and soil fertility. In future, the findings will be pointed to any food related cause for Autism Spectrum Disorder. This is taken into consideration to carry out an interdisciplinary research.

V. REFERENCES

- [1] <https://in.mathworks.com/discovery/supervised-learning.html>
- [2] Arun K Pujari, "Data Mining Techniques", University Press 2001.
- [3] https://www.tutorialspoint.com/data_mining/dm_classification_prediction.htm
- [4] M.S.Mythili, A.R.Mohamed Shanavas, "A Study on Autism Spectrum Disorders using Classification Techniques", International Journal of Soft Computing and Engineering (IJSCE), Vol.4 Issue-5, Pp.88-91, November 2014.
- [5] Mohana E, Poonkuzhali.S, "Categorizing The Risk Level Of Autistic Children Using Data Mining Techniques", International Journal of Advance Research In Science And Engineering IJARSE, Vol. No.4, Special Issue (01), Pp.223-230, April 2015.
- [6] Sumi Simon, Chandra J and Saravanan N, "Empirical Evaluation of Data Mining Classification Methods for Autistic Children", [7] Special Issue Published in International Journal of Trend in Research and Development (IJTRD), Pp.7-10, February 2016.
- [8] M.S.Mythili, A.R.Mohamed Shanavas, "A Novel Approach to Predict the Learning Skills of Autistic Children using SVM and Decision Tree", International Journal of Computer Science and Information Technologies, Vol. 5 (6), Pp.7288-7291, 2014.
- [9] G. Leroy, A. Irmscher, and M.H. Charlop-Christy, "Data Mining Techniques to Study Therapy Success with Autistic Children", 2006 International Conference on Data Mining, 26 - 29 June 2006, Monte Carlo Resort, Las Vegas, USA.
- [10] Rajshekhar Borate, Rahul Ombale, Sagar Ahire, Manoj Dhawade, and Mrs. Prof. R. P. Karande, "Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in India", International Journal of Novel Research in Computer Science and Software Engineering Vol. 3, Issue 1, Pp.34-37, January-April 2016.
- [11] D Ramesh, B Vishnu Vardhan, " Analysis of Crop Yield Prediction using Data Mining Techniques", International Journal of Research in Engineering and Technology, Vol. 04 Issue 01, Pp.470-473, Jan-2015.
- [12] Zekarias Diriba, Berhanu Borena, "Application of Data Mining Techniques for Crop Productivity Prediction", HiLCoE Journal of Computer Science and Technology, Vol. 1, No. 2, Pp.152-155.
- [13] M.Kannan, S.Prabhakaran, P.Ramachandran, "Rainfall Forecasting Using Data Mining Technique", International Journal of Engineering and Technology, Vol.2 (6), Pp.397-401, 2010.
- [14] Dr. S.Hari Ganesh, Mrs. Jayasudha, "Data Mining Technique to Predict the Accuracy of the Soil Fertility", International Journal of Computer Science and Mobile Computing, Vol.4 Issue.7, Pp. 330-333, July- 2015.