Exploring The Frequent Symptoms And Predict Learning Disability By Increasing The Classifier Accuracy Using Machine Learning Approach

P. Saraswathi[1], Dr. N. Nagadeepa[2]
[1] Bharathiar University, Coimbatore, TamilNadu, India
[2]Principal, Karur velalar College of Arts and Science, Kuppam(PO), Karur 639 111, India

Abstract- *Learning disability is one of the most important crises for school going children. It impairs the skill to carry out the learning process. Identifying learning disability and interference at the accurate time is vital for the growth of a child. This paper analyzes the symptoms of LD to predict the learning disability. The important parameters of LD prediction are identified using data mining technique. The existing algorithm has difficulty in managing the inconsistent data that leads low classifier accuracy. To overcome the problem, the proposed work develop new association ldfit algorithm that finds an approach that loads the misplaced values as strongly as feasible. Also the attribute reduction method used to decrease the number of attributes that are unrelated to the prediction of LD. The proposed algorithm shows the accuracy of correctly classified instance to build the models. Rules generated with the m-apriori algorithm gives a strong platform for making decisions. To increase the accuracy of prediction, a tool is developed in finding the LD at early stage to avoid academic and social losses.*

Keywords- Classifier, LD tool, association rules, m-apriori algorithm, Learning disability.

I. INTRODUCTION

Data mining used to extract the data from large databases. Educational data mining is one of the most important research areas under mining technique. It is used to analyze the learning behavior of the student. Learning disability is a general term that describes specific kinds of learning problem. It includes the learning disorders like reading (dyslexia), writing (dysgraphia), doing mathematical works (dyscalculia) and even some behavioral problems leads the children to learning disability. They have difficulty in reading, writing, arithmetic skills, paying attention. LD affects the children both academically and socially. Children with LD needs the correct guidance to help them reach high achievers. Data preprocessing is the first step of mining process that involves cleaning, reduction, integration and transformation of data. Classification in mining approach extracts the models by describing the classes of data. In data mining, ARM discovers the relationship between attributes in large data sets. It is used to extract the rules and finding the frequent signs and symptoms of learning disability.

II. REVIEW OF LITERATURE

Pooja Manghirmalani Mishra et al. [1] present the classifier based on SVM and implemented in the samples of LD. Also the paper classifies the data using semi-supervised learning to give the accuracy for predicting LD with children. Dr. Julie M. David et al. [2] present the two classification methods for the prediction of LD in school children. Rule mining is performed and shows that the Rough Set Theory is useful in the inconsistent data. Ambili K et al. [3] analyze various mining technique for the prediction of LD and shows that ANN is the best classification algorithm in the diagnosis of LD compared to other machine learning algorithms.

M. Saraee et al. [4] proposed the methods to identify the students with dyslexia and generate the rules for the progression of each year academic study. Hanumathappa. M et al. [5] discussed the education data mining for the prediction of LD by generating the decision tree algorithm. Margaret Mary. T et al. [6] implement the various classification algorithm to diagnose ADHD in student and extract the rules from the decision tree by adopting new preprocessing algorithm. E. Radhamani et al. [7] propose the accuracy of classifier MLP and SVM to classify the groups and implemented the datasets in Weka tool.

III. METHODOLOGY

The drawback of the existing work is that difficulty in managing misplaced values and the attribute reduction. To overcome the existing problem, proposed work develops a

new association ldfit algorithm to handle the misplaced values and reduced the attributes to enhance the accuracy of classification. Also the existing algorithm rescans the database repeatedly to generate the rules. The proposed algorithm developed the m-apriori algorithm scans the database one time and generates the rules for making decisions.

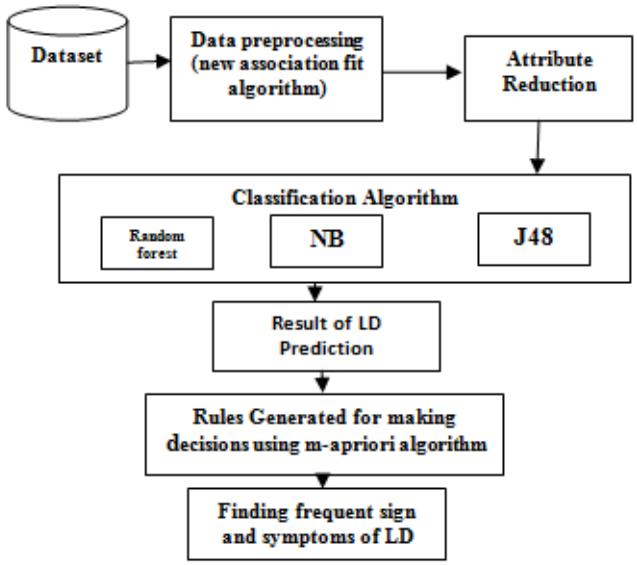

Fig 1. Framework to predict LD using classifier accuracy and find the frequent signs and symptoms of LD behavior

A. Data Collection

The real world datasets are collected in the form of questionnaire from school going children, parent, and teacher and with the clinical way therapist. The collected datasets contains 1080 instances with the attributes related to learning disability. The sample contains 229 attributes and it is applied to the filtering techniques and extracted 16 top most attribute with common signs and symptoms affects the learning disability is identified.

B. Data Distribution

Among 229 attributes, factors affecting learning disability attributes is filtered. Using the filtering technique, top most factors affecting 16 attributes is extracted. In that 68% of children having learning disability (LD=True) and 32% of children are not having learning disability (LD=False). The most frequent attribute is listed using the filtering technique.

C. Data Normalization

The dataset contains non-numeric values which are extracted from different sources. With the use of normalization technique, it is converted into Boolean values.

Table 1. Normalized value of most frequent attributes that affects the student with LD

PRD	NRI	IHT	GTH	HME	ME	STM	MTP	CD	LSH	OMI	TPT	LAP	PAM	DE	DSP
0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1
0	0	1	1	0	0	0	1	0	0	1	0	0	0	0	1
0	1	1	1	0	0	0	1	0	0	1	1	0	1	1	1
0	0	0	1	0	1	0	1	1	0	0	1	0	0	1	1
1	1	0	1	0	1	1	0	0	1	1	0	1	1	1	1
0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0
0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	1	0	0	1	1	1	0	0	1	0	1	0	1	1
0	0	0	0	0	0	1	0	0	1	1	1	0	0	1	1

D. Data Preprocessing

The main drawback in the existing work is no solution for handling the unwanted data in dataset. So, the classifier accuracy is low. To increase the accuracy of classification new preprocessing algorithm ie new association ldfit algorithm is proposed to enhance the accuracy of prediction. To overcome the above disadvantages two approaches are used. First, by replacing the missing attribute value existing value of same magnitude in another cases. The second approach finds the location of the values to decrease the attributes. Among huge number of attributes, the most common attribute is extract and unrelated attribute is detached using the attribute reduction algorithm. The reduced attribute applied to ranking algorithm to rank the attributes. The dataset is preprocessed using the above two approach and applied to the classifier to generate the accurate accuracy.

Algorithm – new association ldfit algorithm

Input: Dataset
Output: new inclusive dataset
Step 1: Read the Dataset
Step 2: Find the misplaced Value in Dataset
Step 3: Compute the association value with antecedent and consequent attribute of misplaced value
Step 4: if antecedent value is greater than consequent value, then fill it with antecedent attribute value
Step 5: else fill it with consequent attribute value

E. Classification accuracy using the new preprocessing algorithm

The modified dataset used in different classifier to correctly classify the instances. The dataset after the preprocessing technique is applied to NB, J48, and Random Forest classifier. The classification accuracy is identified with the help of WEKA open source tool to predict the LD of school going children.

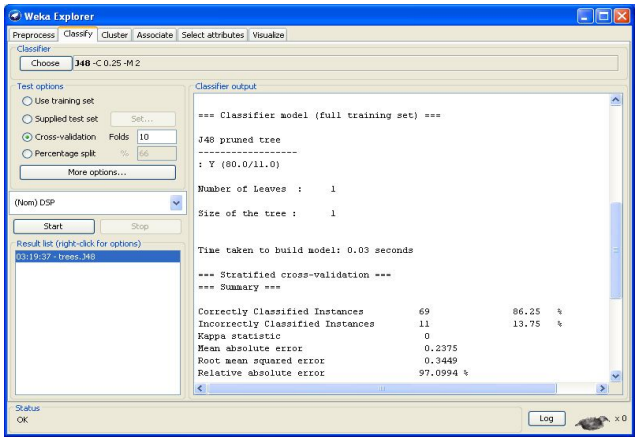

Fig 2: Classifier accuracy in Weka Tool

The Classifier accuracy is compared and shows the accuracy to predict the LD. From the Table 1 shows the classified instances for the classifier Naïve Bayes, J48 and Random Forest.

Table 2. Classifier accuracy for the new preprocessed dataset

Classifier	Naïve Bayes	After new association fit algorith	J48	After new association fit algorith	Random Forest	After new association fit algorith
Correctly Classified Instances	75%	91%	86.25%	92%	68%	90%
Incorrectly Classified Instances	25%	9%	13.75%	8%	12%	10%
Kappa statistic	0.1412	0.1815	0.1545	0.1678	0.0235	0.0104
Mean absolute error	0.2754	0.2812	0.2375	0.3147	0.2475	0.3105
Root mean squared error	0.4084	0.5012	0.3449	0.5174	0.3886	0.6521
Relative absolute error	112.59%	101%	97.10%	91.78%	101.19%	95.41%
Root relative squared error	118.39%	124.52%	99.96%	97.50%	112.64%	98.43%

The confusion matrix for the classifer is given below..

Confusion Matrix

a,b <- classified as

68	1	a=Y
11	0	b=N

F. Finding the frequent signs and symptoms of LD using m-apriori algorithm.

The Learning disability prediction process considered as a decision making process. The rules generated using the dataset gives a strong platform for making decisions. The m-apriori algorithm used to generate the rules for making decisions and finding the frequent signs and symptoms of learning disability.

m-apriori

========

Minimum support: 0.5 (1080 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 10
Generated sets of large itemsets:
Size of set of large itemsets L(1): 16
Size of set of large itemsets L(2): 55
Size of set of large itemsets L(3): 37

Best rules found:

1. MT=N LAP=N 44 ==> PR=N 42 conf:(0.95)
2. DMD=N DSP=Y 42 ==> LSE=N 40 conf:(0.95)
3. MT=N PCD=N 47 ==> PR=N 44 conf:(0.94)
4. DNRI=N LAP=N 44 ==> PR=N 41 conf:(0.93)
5. LSE=N LAP=N 44 ==> PR=N 41 conf:(0.93)
6. PCD=N LAP=N 43 ==> PR=N 40 conf:(0.93)
7. LSE=N DMD=N 43 ==> PR=N 40 conf:(0.93)
8. DMEA=N LSE=N 43 ==> DSP=Y 40 conf:(0.93)
9. LSE=N DMD=N 43 ==> DSP=Y 40 conf:(0.93)
10. LAP=N 52 ==> PR=N 48 conf:(0.92)

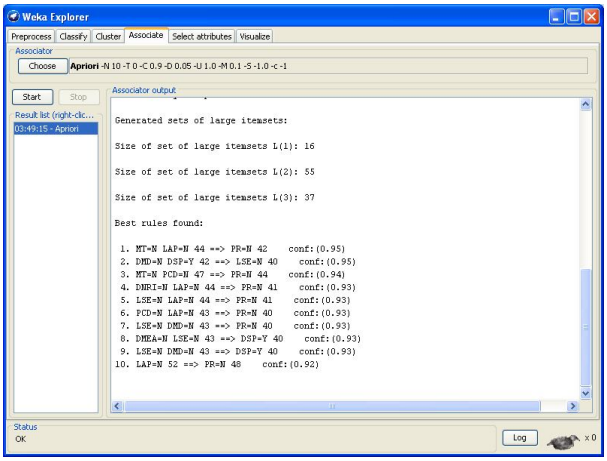

Fig 3. Best rules found by the algorithm m- apriori algorithm

G. Prediction of Learning Disability using LD Tool

Based on the above study, developed a new LD Tool in MATLAB help the child to assess the status of learning disability effectively. Also the tool assess the category of LD prediction ie Dyslexia, Dyscalculia, Dysgraphia etc. It is user friendly tool and can be used by parent/teacher/ clinical

therapist to predict the learning disability of student. The early diagnosis of LD at the right time to avoid the academic and social loss of students.

IV. EXPERIMENTAL RESULT

From the above table 2, shows the classifier accuracy for the prediction of LD using the new association ldfit algorithm. From the fig 4. shows the accuracy of the algorithm for the prediction of learning disability.

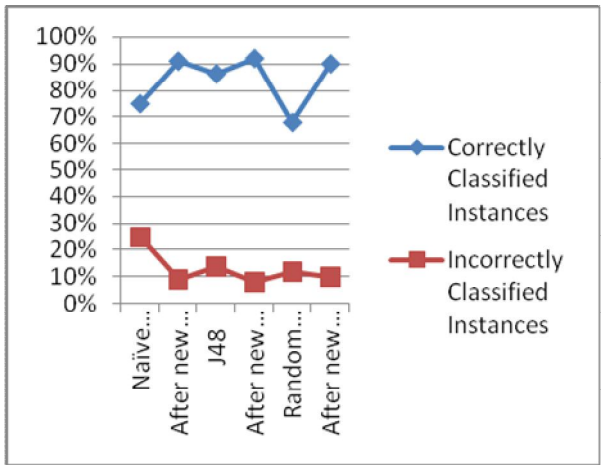

Fig. 4 Accuracy of the classifier for the prediction of LD

V. CONCLUSION

Based on the experimental result the new association ldfit algorithm in preprocessing stage shows the correctly classified instance for the classifier Naïve Bayes, J48 and Random Forest to build the model. Also the study finds the reading, writing, math and behavioral problem are the significant symptoms in the prediction of learning disability. The m-apriori algorithm generate the rules shows the accuracy of decision making improved by applying these rules. The developed tool used to predict the learning disability with percentage and the category of LD.

REFERENCES

[1] Pooja Manghirmalani Mishra, Dr. Sushil Kulkarni, " Classification of data using semi-supervised learning", International Journal of Computer Engineering and Technology, Vol 4, Issue 4, July-August 2013.

[2] Dr. Julie M. David, Dr. Kannan Balakrishnan, " Comparative study of Decision trees and rough sets for the prediction of learning disabilities in school-age children", International journal of innovative research in computer and communication engineering, Vol 2, Issue 7, July 2014.

[3] Ambili K, Afsar P, " A framework for learning disability prediction in school children using Artificial Neural Networks", International journal of advanced research in science, engineering and technology, Vol 3, Issue 6, Juber 2016.

[4] M Saraee, C. Edwards, A. Peers, A. Redfern, P. Wilkinson, " Identifying trends concerning computer science students with Dyslexia: A data mining approach", University of salford Manchester, 2005.

[5] Hanumathappa. M, "Educational data mining and prediction of learning disabilities", Asia pacific journal of research, October 2013.

[6] Margaret Mary. T, Hanumathappa. M, " Diagnosis of learning disabilities in school going children using data mining techniques: a survey", International Journal of Innovative Science, Engineering & Technology, Vol 1, Issue 8, October 2014.

[7] E. Radhamani and K. Krishnaveni, " Diagnosis and evaluation of ADHD using MLP and SVM Classifiers", Indian Journals of Science and Technology, Vol 9, May 2016.