

# INTRODUCTION OF DECISION TREE TECHNIQUE OF DATA MINING TO PREDICT LEARNING DISABILITY IN SCHOOL - AGE STUDENTS

*G. Angeline Prasanna\*, P.J. Anu*

## Abstract

These learning difficulties are due to genetic and/or life science factors that alter brain function in a way that affects one or more processes related to psychological characteristics of learners. These problems hinder the process of learning basic skills such as reading and writing or math, etc. Also they will additionally interfere with high-level skills such as ranking, time design, abstract design, long-term or short-term memory, and so on. It is necessary to know that learning difficulties have an influence on each person's life in addition to academics and can affect family members, friends and even the work itself.

This paper aims to put forward some data extraction tools that predict learning disability in children of school age. Decision Tree is a powerful tool that used for data classification and data mining methods. The aim is to achieve better performance and success more effectively by effectively using the technical tools of education, which will deliver a change to students, faculty and academic institutions.

**Keywords:** Index Terms - accuracy, clustering, data mining, decision tree, prediction

## I. INTRODUCTION

Data mining is a set of techniques. It refers to discovering interesting and unexpected patterns in data, which brings about the invention part and potentially useful pattern or knowledge from huge amount of knowledge.[1] First it's a non-trivial process; which suggests that it's not obvious the

knowledge isn't Obvious it's to be extracted. It's inexplicit the sense that the knowledge is inbuilt within the data you extract it only from the information. There's a novelty part which implies that the Knowledge should be a brand new knowledge and unknown knowledge previously. And eventually, it's to be potentially useful this has got to be useful knowledge looking on the applying. So, the knowledge often takes the shape of patterns in data, some regularity or some reasonably structure within the data and from huge amount of knowledge that's also a very important aspect. Data mining is used in various areas such as customer relationship management, marketing, technology, expert forecasting, medicine field, analysing patterns and trends in crime and disorder, web browsing and mobile computing [2]. Many areas of medical services also use information extraction, such as predicting the effectiveness of surgical procedures, medical tests, drugs, and thereby discovering links between clinical and diagnostic data [3]. Classification can be defined as data manipulation function that assigns objects in a collection to categories or classes for focus. The purpose of classification technique is to exactly predict the target class for each case in the data. For eg: you have gone to a doctor and every patient is separated and labeled by some attribute say has high temperature or not, has cough or not, has fever or not; and based on this signs of sickness and attributes the doctor classifies the patient into a disease say influenza or flu or dangerous disease puts it into a class. Predicts the class for a how does the doctor does it, she has looked at various influenza and malaria patients in the past; she has trained over them. And from this previous experience she has spitted a model and for a new patient comes, using this model she tries to predict the class of that new person. You can have different type of classifiers, rule based methods, Support

Department of Computer Science,  
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India  
\*Corresponding Author

Vector Machines, memory reasoning , neural networks, decision tree methods etc. Here in this paper we are going to discuss how decision tree help for finding Learning Disability in School age students.

## **II. LEARNING DISABILITY**

Learning disabilities do not represent information or intelligence level. Children with the disorder are more likely to have difficulty performing certain types of skills or tasks when they have to practice them alone or when taught in the traditional way. Learning disabilities cannot be cured or fixed [4]. There are a few suggestions, most of which are related to schoolwork, because learning difficulties are usually identified at school, which can also mean that the child has some form of ailment. A child is unlikely to show all or most of the symptoms [3]. LD affects about fifteen percentages of children registered in school. Nearly 3000000 kids (aged 6 to 21) are very nervous and, of course, being studied; more than half of all school-age children are nervous. Children with special learning difficulties sometimes have questions about people and teachers. Pediatricians are regularly asked to diagnose specific learning disabilities in school-age children. Children with learning difficulties are academically and socially affected [5]. Specific learning disabilities were recognized in some countries for much of the 20th century, in others only in the second half of the century, but mostly elsewhere [5]. It is also discovered only after young children start school and have difficulty in gaining basic academic skills. Even when they need to be recognized, the amount of aid available ranges from non-service to universal service. This imbalance in response services is tragic because most children with learning disabilities who receive adequate and informed action can continue into the university system and find jobs ranging from professors to workers. On the other hand, if it does not help, the chances of adapting to the problems that arise are quite high. As the world becomes more complex, as mental objects and more abstract concepts increase, more

young people will experience difficulties and it will be recognized that they have succeeded in learning so that our collective life is not lost [5].

## **III. LITERATURE REVIEW**

Several studies are available to show the importance of "data mining" in education, suggesting that concepts are often a substitute for obtaining proper and precise information about practice and the effectiveness of the learning process [6]. In the field of scholastic technology, "Data Mining" has also been used for contemporary curricula and research themes and for scholars [7]. Most of the research has been published on the topic of study. For example, the Bhardwaj Naive Bayes algorithm was used to predict learning effects with 13 variables [8]. The issue has been used in modeling, generally to the detriment of educated people who are at risk of failure, thus providing active management and programming. In their research, Varghese, Tommy, and Jacob [9] supported 8,000 students using the "K" algorithm for 8 variables (average of test/exam inputs in university legion averages, number of dissertations, seminary notes, and attendance scores)). The output showed a strong relationship between student frequency and performance. In one of the paper [10] argue that the knowledge gained through mining analysis can improve the academic system in terms of orientation, student achievement, and governance norms. Ayesha Mustafa [11] led the study of assessments, taking into account the evolution of learning and analysis of tests at the beginning and at the end of the curriculum. Bresfelean [12] has conducted research that supports student achievement, and this simple method is often used. Cortez and Silva [3] conducted research on the education system in Portugal and then the results indicated that honest and accurate predictions had been made. It can be built with development means that can help improve the management of education in schools and thus the effectiveness of learning, which can be a vital payback. According to the Sun [13], the relationship between

assessment and teaching is an important tool for preserving and guiding the quality of education. Noaman and Al-Twijri[14] recently published a study that applied it to the entrance to a required university in the Asian country. They use algorithms and devices that need to be developed and models suitable for the general public and then describe those variables. A series of records for previous education input, entrance notes, and even notes describing the University's recommendations. Several studies have demonstrated the impact of using Moodle using data processing [15]. Sun [13] describes various information processing techniques that can be implemented to promote E- learning. Aslam and Ashraf [16] used a clustering algorithm to provide a student learning model. Further research focuses on how data works to change education systems and discuss scientific decisions in schools. Vince Kellen described in his case study a structured application for the notification processing tool - SAP HANA at the University of Kentucky, which estimates the "k-score" price for each student. This position will determine student engagement and subsequent ongoing leadership performance. Grafsgaard, Wiggins, Boyer, Wiebe and Lester [17] developed a system that recognizes facial expressions supported by the frustration or understanding of scholars at school. They also use an algorithm to detect unspoken behaviors and report acquired knowledge. Seong Jae Lee [18] also describes the use of the aforementioned models of human behavior.

**IV. DECISION TREE**

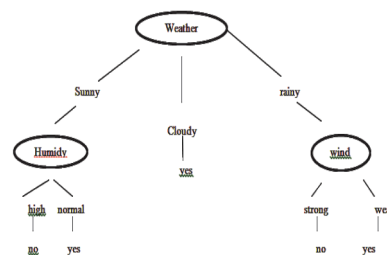
What is decision Tree? Decision trees are one amongst the algorithms to perform what's referred to as a classification task. As personally tend to discuss on top of the classification task is given and an instance an object, it has got to place in one amongst the number of categories. So, it has got a predefined set of categories, now it got to place that instance or the item into one amongst these categories and these categories are known as the class. Let us see an example. Let's assume that one would like to play court game

on a specific day .So it's time to decide whether to play or not. For that play need to stop and check if it is hot or cold, check the wind speed and humidity, regardless of the weather, i.e. it is cloudy, cloudy or raining. Take these factors into account when choosing whether or not to play. So calculate these factors for the last 10 days and write a table of actions as below. (Example table)

**TABLE I**  
*Sample observation table of 10 days*

Day	Weather	Temperature	Humidity	Wind	Play?
1	Sunny	Hot	High	Weak	No
2	Cloudy	Hot	High	Weak	Yes
3	Sunny	Mild	Normal	Strong	Yes
4	Cloudy	Mild	High	Strong	Yes
5	Rainy	Mild	High	Strong	No
6	Rainy	Cool	Normal	Strong	No
7	Rainy	Mild	High	Weak	Yes
8	Sunny	Hot	High	Strong	No
9	Cloudy	Hot	Normal	Weak	Yes
10	Rainy	Mild	High	Strong	No

Above table helps us to predict whether to play or not. But there is a chance that the row in table will not match the particular day that the player is planning to play. This can be a problem. Here the decision tree helps us illustrate details like this and consider all the paths that may have led to the final decision by following a tree structure.



**Fig.1 Decision tree**

In this all things can be seen clearly. In this each node of tree represents an attribute or characteristic, and therefore the branch of each node represents the end of that node. Finally, these are the leaves of the tree on which the last call is formed. If the parameters are still incomplete, check the value of the internal node using the drop-down threshold. In this way, the general formula of the call tree can be modeled;

- Choose the simplest attribute/characteristic.
- The best attribute is one which splits best or separate the information.
- Ask the appropriate question and then follow the solution path.
- Go to step one again and again till you found the solution.

### V. PROPOSED METHOD

According to the paper written by Julie M. David, Kannan Balakrishnan they investigated and provided the result for decision tree method. In fact, they use the same 16 points commonly used to assess the most common signs and symptoms of LD to look for the presence of agitation. This type can be a series of questions that are general indicators of learning difficulties. It is not a protection or assessment activity, but a checklist as the focus for understanding .The list of the 16 attributes used in the LD prediction is shown below in Table 2.

**TABLE II**  
*Listed Attributes*

SL. No.	Attributes	Signs and Symptoms of LD
1	DR	Difficulty with Reading
2	DS	Difficulty with Spelling
3	DH	Difficulty with Handwriting
4	DWE	Difficulty with Written Expression
5	DBA	Difficulty with Basic Arithmetic skills

6	DHA	Difficulty with Higher Arithmetic skills
7	DA	Difficulty with Attention
8	ED	Easily Distracted
9	DM	Difficulty with Memory
10	LM	Lack of motivation
11	DSS	Difficulty with study skills
12	DNS	Does not like School
13	DLL	Difficulty Learning a Language
14	DLS	Difficulty learning a Subject
15	STL	Slow To Learn
16	RG	Repeated a grade

The J48 algorithm is used in weka, a machine learning platform, which contains a framework within the Java class library type. Basically, it can be estimated the value of an attribute by measuring the knowledge gain ratio by category. The attributes are then ranked according to their individual evaluations using the gain ratio, entropy, etc. During this study, J48 in the weka algorithm, used to construct the tree, correctly classified 97.47% of the data sets. The precision speed of the decision tree is given below

TP Rate	FP Rate	Precision	Recall	F Measure	ROC Area	Class
0.984	0.030	0.981	0.984	0.983	0.968	T
0.979	0.022	0.964	0.979	0.972	0.969	F
Correctly Classified Instances					500 Nos.	97.47%
Incorrectly Classified Instances					13 Nos.	2.53%
Time taken to build a model						0.08Sec

*Fig.2 Accuracy rate by using decision Tree*

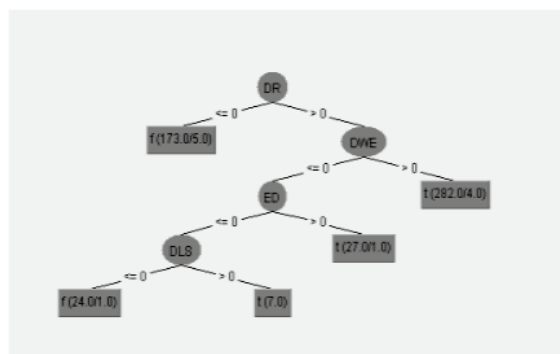


Fig.3 Tree that obtained

### VI. CONCLUSION

Using various classification methods, it can be predicted precisely and more perfectly the LD in any child. The decision tree method is very efficient for forecasting. The calculational time for a decision tree is usually short, and the interpretation of the rules derived from the decision tree will be facilitated by different graphical representations of the tree. The decision tree model shows its ability to find knowledge after the LD identification procedure.

### REFERENCES

[1] Frawley, Piatetsky. "Shaping knowledge discovery in database; an overview. The AAAI." (1996).

[2] Chen, Hsinchun, et al. "Knowledge Discovery in Data Mining and Text Mining in Medical Informatics." Chapter 1 (2005): 3-34

[3] Cortez, Paulo, and Alice Maria Gonçalves Silva. "Using data mining to predict secondary school student performance." (2008).

[4] Paige, Rod. "Secretary." US Department of Education, Twenty-fourth Annual Report to Congress on the Implementation of the Individuals with disabilities Education Act-To Assure the Free Appropriate Public Education of all Children with Disabilities. 2002.

[5] David, Julie M., and Kannan Balakrishnan. "Learning disability prediction tool using ANN and ANFIS." *Soft Computing* 18.6 (2014): 1093-1112.

[6] Ramaswami, M., and R. Bhaskaran. "A study on feature selection techniques in educational data mining." *arXiv preprint arXiv:0912.3924* (2009).

[7] Alfiani, Ardita Permata, and Febriana Ayu Wulandari. "Mapping student's performance based on data mining approach (a case study)." *Agriculture and Agricultural Science Procedia* 3 (2015): 173-177.

[8] Kumar, Varun, and Anupama Chadha. "An empirical study of the applications of data mining techniques in higher education." *International Journal of Advanced Computer Science and Applications* 2.3 (2011).

[9] Varghese, Bindiya M., et al. "Clustering student data to characterize performance patterns." *Int. J. Adv. Comput. Sci. Appl* 2 (2010): 138-140.

[10] Gulati, Pooja, and Archana Sharma. "Educational data mining for improving educational quality." *International Journal of Computer Science and Information Technology & Security* 2.3 (2012): 648-650.

[11] Ayesha, Shaeela, et al. "Data mining model for higher education system." *European Journal of Scientific Research* 43.1 (2010): 24-29.

[12] Bresfelean, Vasile Paul. "Analysis and predictions on students' behavior using decision trees in Weka environment." *2007 29th International Conference on Information Technology Interfaces. IEEE, 2007*



- [13] Sun, Hongjie. "Research on student learning result system based on data mining." *IJCSNS* 10.4 (2010): 203.
- [14] Al-Twijri, Mohammed I., and Amin Y. Noaman. "A new data mining model adopted for higher institutions." *Procedia Computer Science* 65 (2015): 836-844.
- [15] Romero, Cristóbal, Sebastián Ventura, and Enrique García. "Data mining in course management systems: Moodle case study and tutorial." *Computers & Education* 51.1 (2008): 368-384.
- [16] Aslam, Shafiq, and Imran Ashraf. "Data mining algorithms and their applications in education data mining." *International Journal* 2.7 (2014).
- [17] Grafsgaard, Joseph, et al. "Predicting learning and affect from multimodal data streams in task-oriented tutorial dialogue." *Educational Data Mining* 2014. 2014.
- [18] Lee, Seong Jae, Yun-En Liu, and Zoran Popovic. "Learning individual behavior in an educational game: a data-driven approach." *Educational Data Mining* 2014. 2014.