THINK INDIA JOURNAL

ISSN: 0971-1260
Vol-22-Issue-3-July-2019

Predicting Students' Future Learning Behaviour Using Educational Data Mining

Amritpal Singh
Assistant Professor
Lovely Professional University
Amritpal.17673@lpu.co.in

Varun Singla
Assistant Professor
Lovely Professional University
Varun.17705@lpu.co.in

Saloni
Student
Lovely Professional University
saloni12.deep@gmail.com

Ravinder Singh
Assistant Professor
Lovely Professional University

Ravinder.17750@lpu.co.in

***Abstract*—** This research paper focus on the significance of utilizing student information to make a change in planning the education. Furthermore, it describes the advancement of a tool that will allow faculties to classify the students in view of pastacademic performance which is measured with the help of Term Grade point Average (TGPA) reviews of the first semesterand to know in detail the factors which affects the grade point that later leads to dropout. In this paper we have outlined a Fuzzy Inference System that compute a student performance of the first semester and dropout rate as most of the students drop the college in the first semester. There are number of reasons which deviates a student from his/her academics which results in poor academic performance. A sequence of factors should be taken in consideration that is, when to identify the affecting factors which affects the academic performance of the student. The challenging job is therefore to find that affecting factor which affects the quality of academic performance and dropout.

***Index Terms* : EDM, classifier, fuzzy logic, classification.**

I. INTRODUCTION

"Educational Data Mining" is a term use to explain a research discipline that takes data from educational institutes like schools, colleges and universities to develop methods to get information and knowledge from the data [1]. The collected data is used for analysing the university and student conditions to provide additional knowledge in teaching and learning. This technique can be explained by taking different categories of students which are detained, low, below average, average, above average and high risk students. The chances of failing are for high risk students. Support vector Machine, Bayesian Networks [2], Neuro-Fuzzy, Fuzzy logic, Artificial Neural Network are the various techniques for classifying the collected data. A well-defined classification of students is very useful for the educational institutions [3]. By these classified data, institutions can categorise the weak students and can determine the factors which leads to the poor final grade. Accuracy, Speed robustness, scalability and interpretability are the few factors which depend on classification accuracy. It will help to shape the classification model by measuring the positive and negative occurrence. This research has put forward a Fuzzy Logic classification technique to allow the prediction in light of their past academic performance to predict their future academic performance

II. DATA CLASSIFICATION

There are three steps involved in data classification with respect to Educational Data Mining:

Step 1: Getting information of type of data is which is ready for use. Educational data is commonly in the form of either categorical or numeric.

Step 2: Identify the objective of classification. This is the consequences of how instructive information is driven in the order setting.

Step 3: Performing grouping is picking the order strategy or Artificial Intelligence technique. For example: Neural Network, Fuzzy rationale or Decision Tree, and so on.

ISSN: 0971-1260
Vol-22-Issue-3-July-2019

No.	Dropout value	Label
1	7-10	High
2	3-8	Medium
3	0-4	Low

Classification tools: Different methodologies exist for arrangement that is comprehensively connected in instructive information mining. As already specified, their motivation is to find information from instructive databases. Programmed characterization is conceivable because of the presence of various simulated (AI) strategies because of the existence of the gap between the tools like KNN, BayesianNetwork and Neuro Fuzzy [4]. Table 1 demonstrates a comparison of classifiers that was thought to be used as classifiers for this research. Therefore, Fuzzy logic Classification was recognized as the most appropriate method for this study.

Table 1. Comparison of Classifiers.

Tools	Advantage	Disadvantage
Neural Network	Ability to learn and model complex relationships	Difficult to deal with missing data and difficult to train
Decision Tree	Transforms the data into tree representation	Hard to manage missing information
K-Nearest Neighbor	Robust w.r.t search space. Classes need not be linear separable	It is lazy learner
Bayesian Networks	Helps to understand the problem and predicts the consequence	Computationally expensive
Fuzzy-logic	Fuzzy Logic is quite flexible in its working. Even distorted and errorneous information can be addressed	It is based on assumption so results can be inaccurate

III. RESEARCH METHODOLOGY

Understanding what sort of data accessible is a fundamental step before displaying model. Data incorporates for students' academic performance, for example, Continuous Assessment, Mid-term, Attendance (regularity), Health issue and Friend Circle and data includes for dropout rate such as financialissue, confusedvision, less self-confidence and Students Final Achievement (SFA).

Output class is going to separate into six classes as appeared in Table III.

Table 2. Student Performance Classification Label

No.	Grade value	Label
1	80-101	High
2	65-85	Above average
3	50-70	average
4	35-55	Below average
5	0-40	Low
6	0-1	Detained

Table 3. Student Dropout Rate Classification Label

For our research, the system consists of five input and one output attribute. The information parameters of fuzzy rule system are Continuous Assessment, Mid-term, Attendance (regularity), Health issue, Friend circle and the output consists of Students Final Achievement (SFA). These will be sorted with respect to fuzzy labels such as detained(X1), Low(X2), below average(X3), average(X4), above average(X5) and high (X6).Similar to this, there is four information parameters and one output of dropout rate. The info – rmation parameters of dropout are financial issue, confused vision, less self-confidence and students'performance. For instance: "if Continuous Assessment of a student is Low and mid-term is Low and Attendance is medium and health is Low , and Friend circle is Low THEN student performance will be Low" . Another case for dropout is "if financial issue of student is moderate and confused vision is moderate andless self-confidence is Moderate and Result is Low THEN chances of dropout will be moderate ".

Membership Functions

For Continuous Assessment
Range: 0-100, low-[0-40], medium-[30-70], High [60-100]

Similarly, for mid term
Range: 0-100, low-[0-40], medium-[30-70], High [60-100]

For Attendance
Range: 0-100, low-[0-75], medium-[70-80], High [80-100]

For Health issue
Range: 0-5, Critical-[0-2], mild-[1-4], Fit [3-6]

For friend Circle
Range: 0-5, Bad-[0-2], Average-[1-4], Good [3-6]

Membership Functions for student performance

Type - Triangular, Range: 0-100, Detained-[0-1], Low-[0-40], below average [35-55], average [50-70], above average [65-85], High [80-100]

For financial problem
Type - Triangular, Range: 0-5, low - [0-2], medium - [1-4], High [3-5]

For Confused vision

ISSN: 0971-1260
Vol-22-Issue-3-July-2019

Range: 0-5, low-[0-2], medium-[1-4], High [3-5]

For less self-confidence
Range: 0-5, low-[0-2], medium-[1-4], High [3-5]

For student performance
Range: 0-100, Detained-[0-1], Low-[0-40], below average [35-55], average [50-70], above average [65-85], High [80-100]

Membership Function for chances of dropout rate
Type - Triangular, Range: 0-10, low - [0-4], medium-[3-8], High [7-10]

IV. FUZZY INFERENCE ENGINE

Then the degree of membership for the output value will be defuzzified by using Mamdani'sdefuzzification method. Mamdani has given many methods.

The formula for centroid method is given below.

$$X^* = \frac{\sum A\bar{X}}{\sum A}$$

X* is the desired crisp value
A is the area of the segment of aggregated fuzzy set.
X, is the corresponding centroid

After defuzzification the result will be a crisp value.

V. RESULT ANALYSIS

Interaction with the students is done to get a reasonable thought regarding the components that are influencing their academic performance .The system automatically tells the dropout rate and performance rate of the students.The UI is appeared in "Fig. 1," and "Fig. 2". The questioner filled the factors in view of answer given by the student.

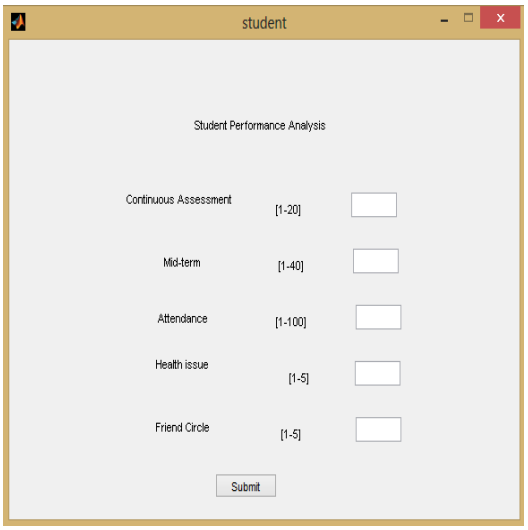

Fig.1. User interface of student performance

At the point when the user taps the Analyze catch, the values of the components are figured. The fuzzy input factors and output factors are appeared in "Fig. 3" and case fuzzification [5,6] of input variables are appeared in "Fig. 4".Similar to this way, dropout rate will also work.

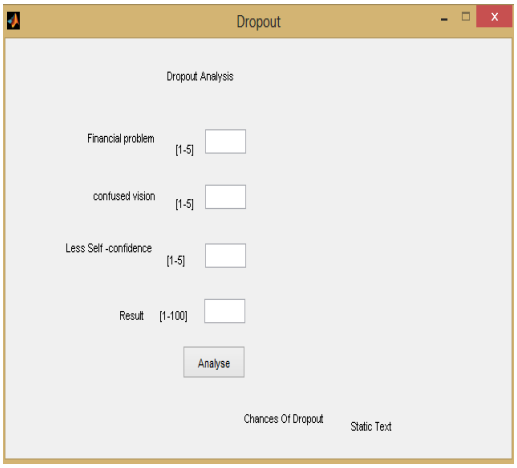

Fig.2. User interface of dropout rate

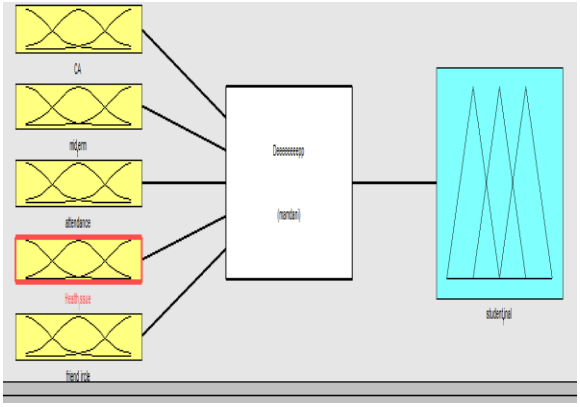

Fig.3.Fuzzy input and output variables

ISSN: 0971-1260
Vol-22-Issue-3-July-2019

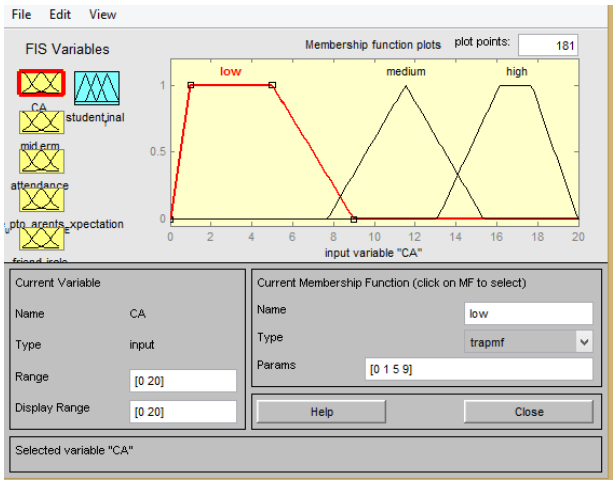

Fig.4. Fuzzification of input variables

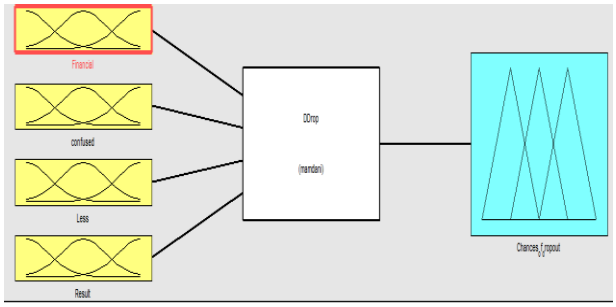

Fig.5.Fuzzy input and output variable for dropout

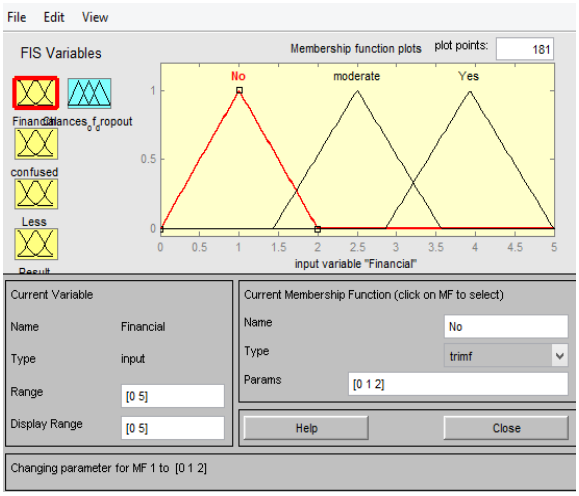

Fig.6. Fuzzification of input variable of dropout rate

The rules for fuzzy rule [7] based system have been taken into consideration for student performance and dropout rate. Some of the rules are as shown in Fig. 6 and Fig.7.

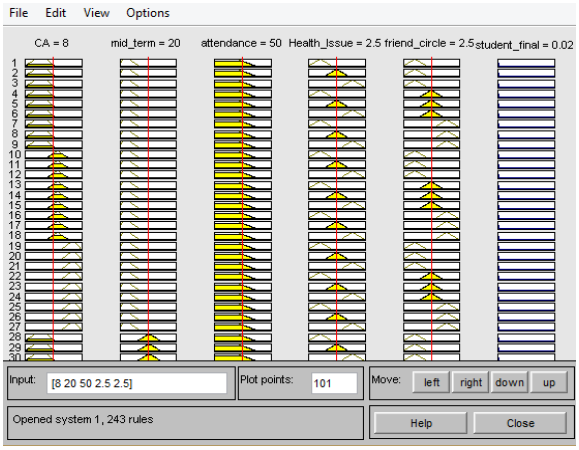

Fig.7. Fuzzy Rule Base

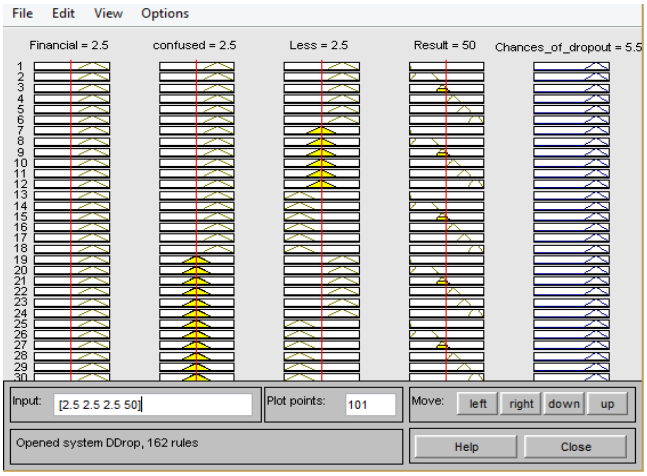

Fig.8. View of fuzzy rule base for dropout

Defuzzification: It is the way to convert fuzzy output variable into crisp value.It is done after the evaluation of the fuzzy inputs but the final output is a single number [8,9,10].There are in built five defuzzification methods. These are centriod, bisector, mom, lom, som. Centriod method is the most popular defuzification method [11].

VI. CONCLUSION

In this paper, fuzzy logic is applied corresponding to the student's future prediction and dropout rate and to improve the students' achievement and dropout rate and success more effectively in an efficient way using fuzzy logic. This will help universities to initially figure out students with low academic achievement and find ways to support them.

ISSN: 0971-1260
Vol-22-Issue-3-July-2019

VII. FUTURE SCOPE

We have worked on fuzzy logic to predict students' performance and dropout. In near future we will be testing our work using neural network as neural network proved to be good option as optimizer which can improve the results.

REFERENCES

[1] Baker, R. S. J. D., KalinaYacef (2009). "The State of Educational Data Mining in 2009: A Review and Future Visions." Journal of Educational Data Mining Vol 1(1): 3-16.

[2] Rawat, S. S., S. S. Reddy, et al. (2015). "Ascertaining the Factors Influencing Students' Performance for Engineering Pedagogy." Journal of Engineering Education Transformations: 30-33.

[3] Zhang Y, e. a. (2010). Using data mining to improve student retention in HE: a case study. Proc.12th Int. Conf. on Enterprise Information Systems, Volume 1: Databases and Information Systems Integration., Portugal.

[4] Jindal, R., Malaya Dutta Borah (2013). "A Survey on Educational Data Mining And Research Trends." International Journal of Database Management Systems (IJDMS) Vol.5(3).

[5] Gobert, J. D., Y. J. Kim, et al. (2015). "Using educational data mining to assess students' skills at designing and conducting experiments within a complex systems microworld."Thinking Skills and Creativity.

[6] Huang, S. and N. Fang (2013). "Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models." Computers & Education 61: 133-145.

[7] P. Kaur, S. K. Singh, P. Agrawal, and L. Jain, "Fuzzy Rule Based Students' Performance Analysis Expert System."

[8]G. Kesavaraj and S. Sukumaran, "A study on classification techniques in data mining," *2013 Fourth Int. Conf. Comput. Commun. Netw. Technol.*, pp. 1–7, 2013.

[9]V. K. Deepa, "Rapid Development of Applications in Data Mining," pp. 1–4, 2013.

[10]Y. Kurniawan and E. Halim, "Use Data Warehouse and Data Mining to Predict Student Academic Performance in Schools : A Case Study (Perspective Application and Benefits)," no. August, pp. 98–103, 2013.

[11]K. V. Krishna Kishore, S. Venkatramaphanikumar, and S. Alekhya, "Prediction of student academic progression: A case study on Vignan University," *2014 Int. Conf. Comput. Commun. Informatics Ushering Technol. Tomorrow, Today, ICCCI 2014*, no. 2, pp. 3–8, 2014.

[12]M.Mayilvaganan and D. Kalpanadevi, "Comparison of classification techniques for predicting the performance of students academic environment," *Commun. Netw. Technol. (ICCNT), 2014 Int. Conf. Comput. Intell. Comput. Res.*, pp. 113–118, 2014.

[13]N.T. Nghe, P. Janecek, and P. Haddawy, "A Comparative Analysis of Techniques for Predicting Academic Performance," pp. 7–12, 2007.

[14]A.S. Subaira and P. G. Scholar, "Efficient Classification Mechanism for Network Intrusion Detection System Based on Data Mining Techniques : a Survey," pp. 274–280, 2014.

[15]K.Ashokkumar and V. Raghavendran, " A semantic Feedback on student performance using Data Mining Techniques" pp. 1605–1609, 2016.