Asian Journal of Computer Science and Technology
ISSN: 2249-0701 (P) Vol.7 No.3, 2018, pp.53-57

DOI: https://doi.org/10.51983/ajcst-2018.7.3.1898

Recommender System for Student Performance Using EDM

S. Jothi Lakshmi[1] and M. Thangaraj[2]

[1]Research Scholar (PT), [2]Professor & Head,
Department of Computer Science, Madurai Kama Raj University, Madurai, Tamil Nadu, India
E-Mail: jothi2008@gmail.com
(Received 15 September 2018; Revised 30 September 2018; Accepted 13 October 2018; Available online 20 October 2018)

***Abstract*-Student's performance plays an important role in an educational institutions and economic growth of society by producing graduates. Educational Data mining algorithms are used to extract the hidden knowledge from the Educational institutions. The recommender system is a special type of information filtering system. This paper provides a recommender system for evaluate student performance that helps the students who need the special attentions.**
***Keywords*: Data Mining, Classification, Student Performance, Recommender, EDM**

I. INTRODUCTION

Educational Data mining refers to technique, tools for extracting knowledge from large repositories of data generated by educational environments. The variety of data is generated by higher education institutions. The data mining techniques and algorithms are applied in order to discover the patterns from educational database [14]. Now a days EDM techniques used by the institutions to guide the student learning environment, develop the course model, student performance and behaviour [20].

A recommender system is a special type of information filtering system and it is popular in E-commerce, entertainment, social networking, higher education, The recommender system[5] is used to prioritise information about items such as music,news,books,image or web page to use with respect to their interest. The recommendation is based on the knowledge of user behaviour or knowledge of all items in the database.

The main objective of this paper is to use data mining techniques to analyze student performance in distance learning system. Data mining techniques provides many tasks that could be used to study the student performance. The classification algorithm is used to evaluate student's performance and provide the recommendations to the institution.[2]

II. RELATED WORK

This section provides the detailed study of previous research work on student performance. Han and Kamber describes data mining process that allow the user to analyze data from different dimension, categorize it and summarize the relationship which are identified during mining process[12]. Amajad ,Abu conducted a research on student performance. They used ID3,C4.5 algorithm to construct a tree. Furthermore they designed a model which predicts student performance based on related personal and social factor[5].This paper[23] reviews prediction of student performance using with data mining algorithms. Parneet Kaur conducted a study on data mining algorithms to predict slow learners in education sector [18].

Brijesh Kumar designed a model using with classification model to extract knowledge that describes the student performance in the end semester exams.[9],This paper[10] reviews the model for analyze student performance in Learning management systems. They used clustering and classification technique for extract the knowledge. KalpeshAdhatrad designed a model to predict the individual student performance using with Classification algorithms in EDM[14].

III. EDUCATIONAL DATA MINING PROCESS

Data mining is the process of Knowledge Discovery in Database. Data mining techniques are used to extract hidden pattern and relationship from large amount of data which is used in decision making [14]. While data mining and knowledge discovery in data base are frequently treated as synonyms, data mining is actually part of the knowledge discovery process [10]. Now a day's student performance is determined by internal assessment and end semester examination.

The Internal assessment is carried out by the teachers based upon the student performance in various activities, and end semester exam is scored by the student in the semester exam. Each student has to get minimum pass mark to pass a semester in internal as well as external. This paper provides a recommender system named as SPEDM(student performance educational data mining)for student performance based on their internal and end semester marks in the distance education mode.

A. Data Set

In this study, The Indian University Distance learning result database from various course is obtained. Initially the size of data is 2000.In this step data is stored in different tables was joined in a single table after joining process errors were removed.

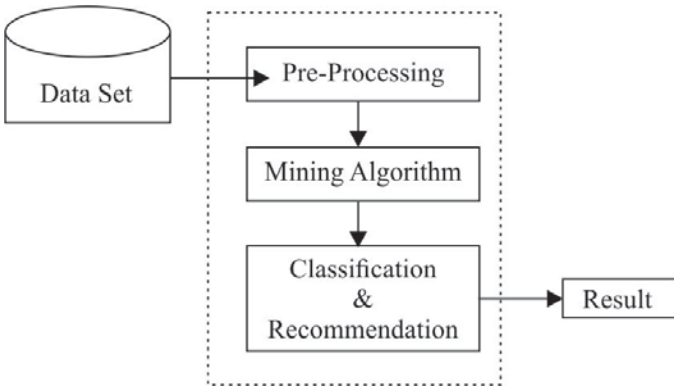

Fig. 1 Recommender system-SPEDM

B. Data Pre-Processing

Data pre-processing is data mining process that involves transforming raw data into pre-process able data format. Data is obtained from different databases. It is susceptible to noise, missing value and inconsistency, The data is pre-processed in order to get the appropriate result .Data cleaning, convert common log format, user identification, session identification, stop word removal, stemming process, white space removal and identifying user request are performed at pre-processing stage.

C. Data Selection

In this step only the required fields were selected which were used in the data mining process. All the variables which were derived from the database are given in Table I for reference.

TABLE I ATTRIBUTES LIST

Variable	Description	Possible Values
Course	Course of the student	BCA,B.Sc(CS), B.Sc(IT)
Gender	Gender of the student	Male,female
PSM	Previous semester mark	Pass,fail
ITG	Internal test Grade	{good,avg,poor}
ASS	Assignment	{yes,no}
ATT	Attendance	{good,avg,poor}
LW	Lab work	{yes,no}
GP	General Proficiency (Like seminar)	{yes,no}

D. Data Mining Algorithm

C4.5 is a decision tree algorithm used in this work to generate decision tree since it has a high accuracy in decision making. C4.5 algorithm uses student result database obtained from Indian university distance learning system. It uses the training data as the input data for generating the decision tree. This tree is used to generate rules for recommendations to improve student performance.

TABLE II C4.5 ALGORITHM

C4.5 Algorithm
{
Input: an attribute-Valued dataset D
Output: A Decision tree.
Tree={}
If D is "pure" OR other stopping criteria met than terminate
end if
for all attribute a £ D do
Compute information-theoretic criteria if we split on a.
end for
a_{best} =Best attribute according to above computed criteria
Tree=Create a decision node that tests a_{best} in the root
Dv=Induced sub-datasets from D based on a_{best}
for all D_v do
Treev=c4.5(D_v)
Attach $Tree_v$ to the corresponding branch of Tree
end for
return Tree
}

E. Attribute Selection Measure

The information gain measure is used to select the test attribute at each node in the tree. Such a measure is referred to as an attribute selection measure or a measure of goodness of split. The attribute with the highest information gain (or greatest entropy reduction) is chosen as the test attribute for the current node. This attribute minimize the information needed to classify the samples in the resulting partitions and reflects the least randomness or "impurity" in these partitions. Such an information theoretic approach minimizes the expected number of tests needed to classify an object and guarantees that a simple (but not necessary the simplest) tree is found[12].

Let S be a set consisting of s data samples. Suppose the class label attribute has m distinct values defining m distinct classes, C_i (for i = 1, ... , m). Let s_i be the number of samples of S in class C_i. The expected information needed to classify a given sample is given by$I(s_1, s_2, \ldots, s_m) = -\sum_{i=1}^{m} p_i \log_2 (p_i)$where p_i is the probability that an arbitrary sample belongs to class c_i and is estimated by s_i/s. Note that a log function to the base 2 is used since the information is encoded in bits.

Let attribute A have v distinct values, $(a_1, a_2, \ldots, a_v)$. Attribute A can be used to partition S into v subsets, $\{S_1, S_2, \ldots, S_v\}$, where S_j contains those samples in S that have value aj of A. If A were selected as the test attribute (i.e., the best attribute for splitting), then these subsets would correspond to the branches grown from the node containing the set S. Let s_{ij} be the number of samples of class C_i ina subsets s_j. The entropy, or expected information based on the partitioning into subsets by A, is given by

$$E(A) = \sum_{j=1}^{v} \frac{s_{1j} + \ldots + s_{mj}}{s} I\left(s_{1j}, \ldots s_{mj}\right)$$

The term $\sum_{j=1}^{v} \frac{s_{1j} + \ldots + s_{mj}}{s}$ acts as the weight of the jth subset and is the number of samples in the subset (i.e. having value a_j of A) divided by the total number of samples in S. The encoding information that would be gained by branching on A isGain (A) = $I(s_1, s_2, \ldots, s_m) - E(A)$.In other words, Gain (A) is the expected reduction in entropy caused by knowing the value of attribute A.

The algorithm computes the information gain of each attribute.The highest information gain is chosen as the test attribute for the given set S. A node is created and labelled with the attribute, branches are created for each value of the attribute, and the samples are partitioned accordingly [12].

IV. IMPLEMENTATION

SPEDM has been implemented with JAVA language using NetBean version 7.3 as JAVA environmentAll the Experiments were done Intel core i3 2.10GH 4GB RAM, running windows 8. The data set of Indian University Distance learning result database from various course is obtained.WEKA java API is used to implement the c4.5 algorithm.From the dataset Decision tree and recommendation rules are generated for student performance. Student performance data is as follows

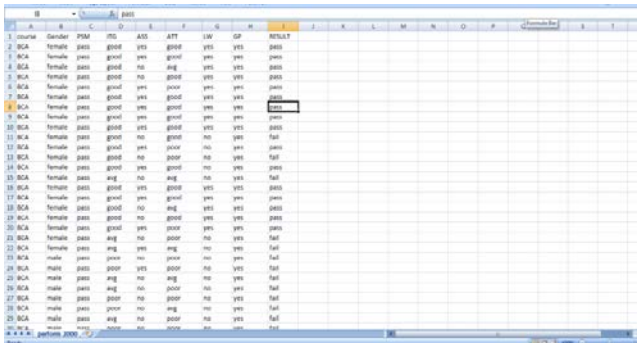

Fig. 2 Student Performance

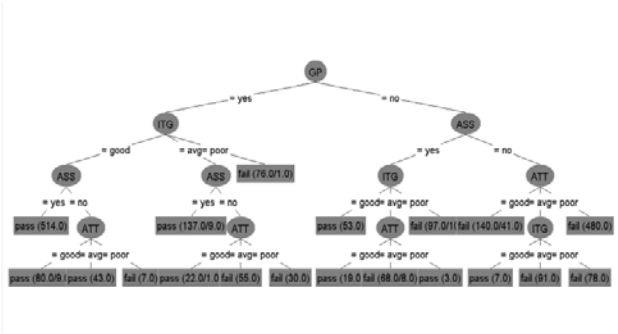

Fig. 3 Decision tree of student performance

A. Production Rules

The IF-THEN rule may be easier to understand the decision tree and generate recommendations that used by recommender system. It can be helpful to the students to improve their performance in the examinations.

TABLE III PRODUCTION RULE

If GP='yes' And ITG='good' and ASS='yes' then Result='Pass'
IF GP=;yes' and ITG='avg' and ASS='yes' then result='pass'
If GP='yes' And ITG='good' and ASS='NO'ATT='good' then Result='Pass'
IF GP=;yes' and ITG='avg' and ASS='no' and ATT='avg' then result='pass'
If Gp='no'and ASS='yes'and ITG='good' then result='pass'
IFGP='no' and ASS='yes'and ITG='avg'andATT='good'then result='pass'
IF GP='no'and ASS= 'no'and ITG='avg'and ATT='good'then result='pass'
IF GP='yes'and ITG='poor'then result='fail'
IFGP='yes'and ITG='avg'and ASS='no' andATT='avg'then result='fail'
IFGP='yes'and ITG='avg'and ASS='no' andATT='poor'then result='fail'
IF GP='no'and ASS= 'yes'and ITG='avg'and ATT='avg'then result='fail'
IF GP='no'andASS='no'andATT='poor'then result='fail'
IF GP='no'and ASS= 'no'and ATT='avg'and ITG='avg'then result='fail'
IF GP='no'and ASS= 'no'and ATT='avg'and ITG='poor'then result='fail'

V. EXPERIMENTAL EVALUATION

The recommendation system can be evaluated using various types of Quality measurement. Accuracy is the fraction of correct recommendations out of total possible recommendations.The SPEDM system is evaluated with Decision support accuracy that are popularly used are Precision, Recall and F-measure. These metrics help the user in selecting items that are very high quality out of the available set of items.

A. Experiment 1: Precision

Precision is the fraction of recommended item that is actually relevant to the user.

$$\text{Precision (p)} = \frac{True\ positive}{True\ positive + False\ positive}$$

TABLE IV PRECISION

Data set	2000	4000	6000	8000	10000
Precision	0.981	0.985	0.991	9.994	0.998

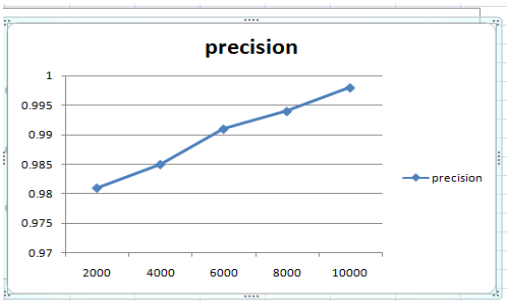

Fig. 4 Precision Graph

Fig. 4 is drawn using with Value of student result data seta and precision values. In this graph X-axis represents the various range of data set in the result and Y-axis shows the corresponding precision values. The high value of precision is 0.998 reached with the Dataset of 10,000. The precision

value highlights the correct positive predictions out of all positive predations. High precision indicates low false value.

B. Experiment 2: Recall

The ratio of correctly predicted positive values to the actual positive values is known as Recall.

$$\text{Recall(R)} = \frac{True\ positive}{True\ positive + False\ Negative}$$

TABLE V RECALL

Data set	2000	4000	6000	8000	10000
Recall	0.931	0.937	0.942	0.948	0.963

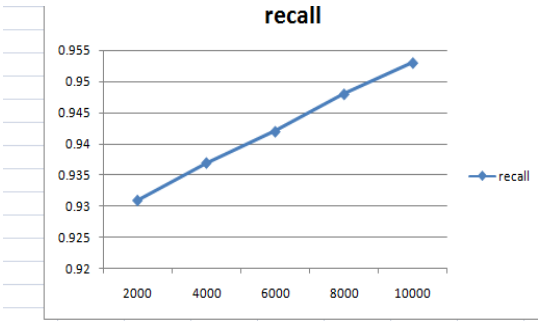

Fig.5Recall Graph

The Graph is created using with Value of Result data set. In this Recall graph, X-axis shows the various range of result centre and Y-axis shows the corresponding Recall values. The high value of recall is 0.963 reached with the Dataset of 10,000. This metrics highlights the sensitivity of the algorithm.

C. Experiment 3: F-Measure

F-measure defined below helps to simplify precisions and recall in to single metric .They are computed as

$$F - measure = \frac{2 * precision * recall}{\text{Precision} + \text{Recall}}$$

TABLE VI F-MEASURE

Data set	2000	4000	6000	8000	10000
F-.Measure	0.955	0.961	0.966	0.971	0.976

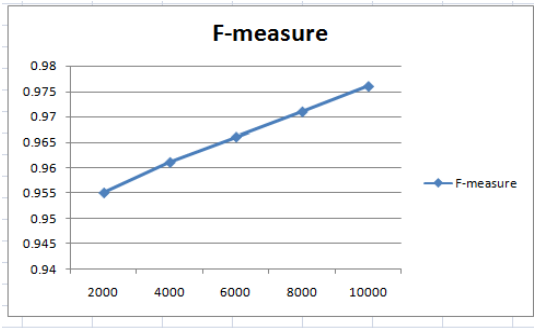

Fig. 6 F-measure

The Fig-6- is plotted against with Value of study result data set. The X-axis represents the various ranges of result data and Y-axis represents the corresponding F-measure values. The high value of F-measure is 0.976reached with the Dataset of 10,000. High value of F-measure indicates the relevant result to the data items.

V. CONCLUSION

In this paper classification technique is used by recommender system to predict student performance. There are various methods used for data classification, the decision tree method is used here. Decision trees and IF-THEN rules are generated which can be used by the recommender system to give suggestion to the higher education institutions. This study can motivate and help the universities to perform data mining task on their student data to find out patterns may be improve their performance. The Analysis of other data mining techniques and very large volume of data set may be the future work.

REFERENCES

[1] AbeerBadr El Din Ahmed and Ibrahim Sayed Elaraby, "Data Mining: A prediction for Student's Performance Using Classification Method", *World Journal of Computer Application and Technology*, pp. 43-47, 2014.

[2] B. E. D Ahemad and I.S AElarby, "Prediction for student'sperformace using classification Method World Journal of Computer Application and Technology,pp. 43-47 2014.

[3] M. A. Al-Barrak and M. Al-Razgan, "Predicting Students Final GPA Using Decision Trees: A Case Study," *Int. J. Inf. Educ. Technol.*, Vol. 6, no. 7, pp. 528–533, 2016.

[4] Amirah Mohamed Sahariri, WahidahHusian and Nuraininabdul Rashid, "A Review on predicting student's performance using Data mining Techniques",Elesevier,The Third Information Systems International Conference,pp. 414-422 (2015)

[5] Amajad Abu Saa,"Educational Data Mining and Student's Performance Prediction", *International Journal of Advanced Computer Science and Applications*, Vol. 7 No.5, pp. 201624.

[6] G. Arumugam, and S. Suguna, "Predictive Prefetching Framework Based on New Preprocessing Algorithms Towards Latency Reduction", *Asian Journal of Information Technology*, Medwell Journals.ISSN: 1682 -3915,2008.

[7] C. Anuradha1 and T. Velmurugan, "A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance".

[8] Brijesh Kumar and Saurabh, *Indian Journal of Science and Technology*,Vol. 8, No. 15,2015.

[9] Brijesh Kumar, Saurabh, Pal, "Mining Educational Data to Analyze student's performance", *International Journal of Advanced Computer Science and Application*,Vol. 2,No.6,2011.

[10] Dorina Kabakchieva, "Student performance by using Data mining Classification Algorithms", *International journal of Computer Science and Management Research*, Vol. 1, 2012.

[11] Ellenita Red, T. Tesaionica,Briags"Classification of Students performance in a Learning Management System Using their eLearning Readiness Attributes", *Research Gate*,2015.

[12] Han and M. Kamber, "Data mining concepts and Technologies", Morgan Kaufmann, 2000.

[13] Home, *International Educational Data Mining Society*. [Online]. Available: http://www.educationaldatamining.org/.

[14] S.Jothilakshmi,Dr.M.Thangaraj, "Design and Development of Recommender System for Target Marketing of Higher Education Institution Using EDM", *International Journal of Applied Engineering Research* ISSN 0973-4562, Vol.13, 2018.

[15] Kalpesdh Adhatrao and AdityaGaykar, "Predicting Studnet's Performance using ID3 and C4.5 classification Algorithms",

International Journal of Data mining &knowledge Management process, Vol. 3, 2013.

[16] T. Mishra, D. Kumar, and Sangeeta Gupta, "Students' Employability Prediction Model through Data Mining," *International Journal of Applied Engineering Research*, Vol. 11, No. 4, pp. 2275-2282, 2016.

[17] B. Minaei-Bidgoli, D. A. Kashy, G. Kortemeyer, and W. F. Punch, "Predicting student performance: an application of data mining methods with an educational Web-based system," in *33rd Annual Frontiers in Education (FIE 2003)*, Westminster, CO, 2003.

[18] O. K. Oyedotun, S.N. Tackie, and Ebenezer O. Olaniyi, "Data Mining of Student'sPerformance: Turkish Students as a Case Study," *International Journal of Intelligent Systems and Applications*, Vol. 7, No. 9, pp. 20-27, 2015.

[19] ParneetKaur, Manpreetsingh and GurpreetsinghJosan,"Classification and Prediction based data mining algorithms to predict slow learnears in education sector",Elsevier,Procedia Computer Science, pp. 500-508,2015.

[20] M. Ramaswami and R. Bhaskaran, "A CHAID based performance prediction model in educational data mining," *International Journal of Computer Science*, Vol. 7, No. 1, pp. 10-18, 2010.

[21] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert systems with applications*, Vol. 33, No. 1, pp. 135-146, 2007.

[22] S. Slater, S. Joksimovic´, V. Kovanovic, R. S. Baker, and D. Gasevic, "Tools for Educational Data Mining: A Review," *Journal of Educational and Behavioral Statistics*, 2016.

[23] P. Strecht, L. Cruz, C. Soares, J. Mendes-Moreira, and R. Abreu, "A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance," in *International Educational Data Mining Society*, pp. 392-395, 2015.

[24] P. Sharma, D. Singh, and A. Singh, "Classification Algorithms on a Large Continuous Random Dataset Using Rapid," *IeeeSpons. 2Nd Int. Conf. Electron. Commun. Syst.*, no. Icecs, pp. 704–709, 2015.

[25] G.Sujatha, Sindhu,"Predicting student performance using personalized analytics", International Journal of Pure and Applied Mathematics, Vol. 119, No. 12, pp. 229-238,2018.

[26] E.Venkatesan, S.Selvaragini,"A Study on the result based analysis of student performance using Data mining",*International Journal of Pure and Applied Mathematics*, Vol. 116, No. 16,pp. 375-379.2017.

[27] S.K YadavBBharadwaj, and S. Pal, "A comparative study for predicting student's performance", *International Journal of Innovative Technology &Creative Engineering*, Vol. 1, No 12,2012.

[28] S.K.Yadav and S Pal, "A Prediction for Performance improvement of engineering students using classification", *World of Computer Science and Information Technology*, Vol. 2, No. 2, pp. 51-56,2012.

[29] B. J. Zimmerman and A. Kitsantas, "Comparing student's self-discipline and self-regulation measures and their prediction of academic achievement", *Contemporary Educational Psychology*, Vol. 39, No. 2, pp. 145-155, 2014.