

# EDUCATIONAL DATA MINING

*M-TECH, 4th SEMESTER, DEPARTMENT OF CSE*

SUMAN KUMAR DAS, AMIT DEWANGAN,

COMPUTER SCIENCE AND ENGINEERING,

Dr. C.V. Raman University, Bilaspur, India.

**Abstract :** Data mining is about explaining the past and predicting the future by means of data analysis. Educational Data Mining is a promising discipline which has an imperative impact on predicting students' academic performance. Thousands of students take admissions in Universities and colleges every year, at the time of admissions they collect the students' data. In the same way while the Teachers join in the institution they collect their personal and professional data. Understand the importance of data is essential from a business point of view. Data collected at the time of admission can be used for classifying and predicting students' behavior and performance as well as teachers' performance. Therefore, in this paper, we are examining the role of Data mining in an Educational Field. , we have identified possible grade values i.e., Excellent, Good, Average and Poor or Fail. We have used K-means clustering algorithm to find the best cluster center for attributes like attendance, Sessional marks and assignment marks etc. We have also discussed a Rule-Based Classification (RBC) method; it extracts a set of rules that shows relationships between attributes of the data set and the class label. In this paper we have also addressed the evaluation of Teachers' performance by using data mining techniques at University and College level.

**Keywords-** Data Mining, Educational Data Mining, Prediction, Clustering, Relationship Mining, Classification, Discovery with models.

## I. INTRODUCTION

One of the primary goals of any educational system is to equip students with the knowledge and skills needed to transition into successful careers within a specified period. How effectively global educational systems meet this goal is a major determinant of both economic and social progress. The main functions of data mining are applying various methods and algorithms in order to discover and extract patterns of stored data . EDM provides a significant amount of relevant information and offers a clearer picture of learners and their learning processes. It uses DM techniques to analyze educational data and solve educational issues. Similar to other DM techniques extraction processes, EDM extracts interesting, interpretable, useful, and novel information from educational data. However, EDM is specifically aimed at developing methods that use unique types of data in educational systems From a practical point of view, EDM allows users to extract knowledge from student data.[4] This knowledge can be used in different ways such as to validate and evaluate an educational system, improve the quality of T& L processes, and lay the groundwork for a more effective learning process [5]. Similar ideas have been applied successfully, especially in business data, in different datasets, such as e-commerce systems, to increase sales profits [6]. Thus, the success of applying DM techniques in business data encourages its adoption in different domains of knowledge. Notably, DM has been applied to educational data for research objectives such as improving the learning process and guiding students learning or acquiring a deeper understanding of educational phenomena. However, while EDM has made comparatively less progress in this direction than other fields, this situation is changing due to increased interest in the use of DM in the educational environment [7].

## RELATED WORK

Data Mining is a powerful tool that enables educational institutions to better allocate resources and staff and proactively manage student outcomes [8]. The educational system can improve their policies, can enhance their strategies and their by can improve quality of their management system [9]. Different Data Mining techniques have the capability to provide effective, improving tools for student and teacher performance. Data Mining could be used to improve business intelligence process, including the education system to enhance the overall efficiency by optimally utilizing the resources available. The performance or success of students in the examination as well as their overall personality development could be exponentially accelerated by thoroughly utilizing Data Mining techniques to evaluate their admission academic performance and finally the Placement in an organization.

## EDM METHODS

Educational data mining methods come from different literature sources including data mining, machine learning, psychometrics, and other areas of computational modelling, statistics, and information visualization. Work in EDM can be divided into two main categories: 1) web mining and 2) statistics and visualization [11]. The category of statistics and visualization has received a prominent place in theoretical discussions and research in EDM [8], [7], [12]. Another point of view, proposed by Baker [3], classifies the work in EDM as follows:

Prediction.

(1) Classification.

Regression.

Density estimation.

(2) Clustering.

(3) Relationship mining.

Association rule mining.

Correlation mining.

Sequential pattern mining.

Causal DM.

(4) Distillation of data for human judgment.

(5) Discovery with models.

Most of the above mentioned items are considered DM categories. However, the distillation of data for human judgment is not universally regarded as DM. Historically, relationship mining approaches of various types have been the most noticeable category in EDM research. Discovery with models is perhaps the most unusual category in Bakers EDM taxonomy, from a classical DM perspective. It has been used widely to model a phenomenon through any process that can be validated in some way. That model is then used as a component in another model such as relationship mining or prediction. This category (discovery with models) has become one of the lesser-known methods in the research area of educational data mining. It seeks to determine which learning material subcategories provide students with the most benefits [13], how specific student's behavior affects students learning in different ways [14], and how tutorial design affects students learning [15]. Historically, relationship mining methods have been the most used in educational data mining research in the last few years

#### A. CLASSIFICATION

Machine learning Classification [20] is a typical data mining technique. Normally classification is used to classify each item in a set of data into one of predefined set of groups or classes. A Rule-Based Classification extracts a set of rules that show relationships between attributes of the data set and the class label.

A set of IF-THEN rules are used for classification

If Final grade-good and Sessional marks-excellent an assignment-good and attendance=excellent, then excellent.

If Final grade-excellent and Sessional marks-good and Assignment-good and Attendance=excellent, then excellent.

If Final grade-good and Sessional marks-good and Assignment-good and Attendance-average then excellent.

#### B. B. CLUSTERING

Clustering is a division of data into groups of similar objects. Clustering plays an outstanding role in data mining applications such as information retrieval and text mining, scientific data exploration, web analysis, spatial database applications, medical diagnostics, marketing and many more

.The K-means is one of the best clustering algorithms in data mining[21].K-Means Clustering proposes to partition n objects into k clusters in which each object belongs to the cluster with the nearest mean. Exactly k different clusters have been produced by this method with greatest possible characteristic.

#### C. PREDICTION

Aims to predict unknown variables based on history data for the same variable. However, the input variables (predictor variables) can be classified or continue as variables. The effectiveness of the prediction model depends on the type of input variables. The prediction model is required to have limited labelled data for the output variable. The labelled data offers some prior knowledge regarding the variables that we need to predict. However, it is important to consider

The effects of quality of the training data in order to achieve the prediction model.[22]

#### D. RELATIONSHIP MINING

Relationship mining aims to find relationships between different variables in data sets with a large number of variables. This entails finding out which variables are most strongly associated with a specific variable of particular interest. Relationship mining also measures the strength of the relationships between different variables [23]

#### DATA COLLECTION

In our case study, we have collected student data of 4th year M tech. (CSE) of CV RAMAN Engineering College (Affiliated to AISECT, BILASPUR). In this process, we have used a questionnaire to collect the real data from the students that describes the relationship between learning behavior and their academic performance. And also we collected the Faculty feedback from the students.

The variables for judging the learning and academic behavior of students used in the questionnaire are Assignment, Attendance, Sessional marks, (grade point average for general performance in the lab or extracurricular) and Final\_ grade in last semester. We grouped all grades into five possible values they are Excellent, Good, Average, Poor and fail which is shown in the table 1.[25]

TABLE .1

Attribute	Description	Possible values
Attendance	Attendance in one semester	Good, Average, Poor
Assignment	Work given by the teacher	(Yes, No)
GPA	Grade point average for General performance in lab or extra-curricular	Good, Poor
TA	Teacher's Assessment	{Poor , Average, Good}
Sessional marks	Percentage of marks obtained in internal exam	Excellent, Good, Average, Poor, Fail
Final Grade	Percentage of marks obtained in current semester exam	Excellent, Good, Average, Poor, Fail
ESM	End Semester Marks	{First > 60% Second >45 &<60% Third >36 &<45% Fail < 36% }
LW	Lab Work	(Yes, No)
SEM	Seminar Performance	{Poor , Average, Good}
CT	Class Test	{Poor , Average, Good}
PSM	Previous Semester Marks	{First > 60% Second >45 &<60% Third >36 &<45% Fail < 36% }

**DATA SET**

The data set of 75 students used in this study was obtained from CSVTU Bhilai (C.G.) course M-tech (Master of Technology) from session 2014 to 2017.

TABLE .2

No.	ATT	ASS	GPA	SEC	ESE	LW	SEM	TA	PSM	CLASS
2014	60%	yes	9.000	Excellent	69.60	Yes	86.46	91.08	Good	First
2015	60%	yes	9.000	Excellent	73.43	Yes	82.18	55.5	Good	First
2016	60%	Yes	8.875	Excellent	83.20	Yes	85.2	88.18	Good	First
2017	60%	Yes	8.714	Good	54.68	Yes	81.73	84.31	Good	Second
Total			8.466		79.07		79.25	78.05		

To work out the information gain for relative to , we first need to calculate the entropy Here is a set of 75 examples are 60% "First", 45% "Second", <36% "Third" and >36 "Fail"..

**RESEARCH WORK AND EXPECTED RESULTS**

The main goal of the proposed .to create an enrolment recommendedsystem that will help students with selecting courses. There commendations will be based on EDM and SNA methods and collaborative filtering principles.[16][17][18] The recommendations will be personalized for each student with respect to the ir need and interests[19]

1. Selecting Attributes Defining the Student Potential
2. Defining Similarity of Students
3. Identifying Weak Student
4. Identifying Problematic Courses for Students
5. Detecting Relations between Courses
6. Finding Interesting Courses
7. Identifying Easy Courses for Students
8. Searching for Similar Courses

**DATA MINING PROCES**

In present day's educational system, a student's" performance is determined by the internal assessment and end semester examination. The internal assessment is carried out by the teacher based upon students" performance in educational activities such as class test, seminar, assignments, general proficiency, attendance and lab work. The end semester examination is one that is scored by the student in semester examination. Each student has to get minimum marks to pass a semester in internal as well as end semester examination

**1 DATA SELECTION AND TRANSFORMATION**

In this step only those fields were selected which were required for data mining. A few derived variables were selected. While some of the information for the variables was extracted from the database. All the predictor and response variables which were derived from the database are given in Table I for reference.

**2 DECISION TREE**

A decision tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decision.

**IMPLEMENTATION****A STUDENTS DATASET**

There is a work methodology which governs our work. The methodology starts from the problem definition, then data collection from Students Database.

**B TEACHER DATASET**

Teachers' performance [24] is evaluated using data mining techniques in this paper. Initially a survey of the teachers' requirements and students' requirements is made. Then we interacted with the teachers and got some knowledge about their methods. We met different teachers that have been given some ideas about the finding of the teacher's performance.

**CONCLUSION**

In this paper, we have discussed the various data mining techniques which can support education system for decision making. Initially, we discussed that how Data mining can be used to evaluate engineering students' performance. We applied data mining techniques to discover knowledge, association rules, classification rules to predict the students' performance as well as we essential part of any Data Warehouse

Teachers' regular assessment is suggested to maintain the quality in higher education. As a conclusion, we have reached our objective which is to evaluate and investigate teacher's performance by the four selected classification algorithms based on Weak. In this paper Students and Teachers performances are evaluated using different Data mining methods. This can be useful in the educational system like Universities and colleges; they can improve their standards and reputations by introducing the new courses or branches. As a result the quality of education can be improved.

**ACKNOWLEDGMENT**

We would like to thank everyone who has motivated and supported us for preparing this manuscript.

**REFERENCES**

- [1] S.-T. Wu, "Knowledge discovery using pattern taxonomy model in text mining," 2007.
- [2] U. Fayyad, Piatetsky, G. Shapiro, and P. Smyth, *From data mining to knowledge discovery in databases*, AAAI Press / The MIT Press, Massachusetts Institute Of Technology. ISBN 0-262 56097-6, 1996.
- [3] S. K. Mohammad and Z. Tasir, "Educational data mining: A review," *Procedia-Social and Behavioral Sciences*, vol. 97, pp. 320-324, 2013.
- [4] R. Baker et al., "Data mining for education," *International encyclopedia of education*, vol. 7, pp. 112-118, 2010.
- [5] C. Romero, S. Ventura, and P. De Bra, "Knowledge discovery with genetic programming for providing feedback to courseware authors," *User Modeling and User-Adapted Interaction*, vol. 14, no. 5, pp. 425-464, 2004.
- [6] N. S. Raghavan, "Data mining in e-commerce: A survey," *Sadhana*, vol. 30, no. 2-3, pp. 275-289, 2005.
- [7] C. Romero, S. Ventura, M. Pechenizkiy, and R. S. Baker, *Handbook of educational data mining*. CRC Press, 2010.
- [8] Richard A. Huebner, Norwich University, "A Survey on Educational Data Mining", *Research in Higher Education Journal*. (www.aabri.com)
- [8] Dr. MohdMaqsood Ali, Jazan University, "Role Of Data Mining In Education Sector", *IJCSCMC*, Vol. 2, Issue. 4, April 2013.
- [9] *An Introduction to the WEKA Data Mining System* Zdravko Markov Central Connecticut State University
- [10] R. S. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *JEDM-Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3-17, 2009.
- [11] T. Barnes, M. Desmarais, C. Romero, and S. Ventura, in *Proc. Educational Data Mining 2009:2nd International Conf*, 2009.
- [12] S. L. Tanimoto, "Improving the prospects for educational data mining," in *Track on Educational Data Mining*, at the Workshop on Data Mining for User Modeling, at the 11th International Conference on User Modeling, pp. 1-6, 2007.
- [13] J. E. Beck and J. Mostow, "How who should practice: Using learning decomposition to evaluate the efficacy of different types of practice for different types of students," in *Intelligent tutoring systems*, pp. 353-362, Springer, 2008.
- [14] M. Cocea, A. Hershkovitz, and R. S. Baker, "The impact of off-task and gaming behaviors on learning: immediate or aggregate?," 2009.
- [15] H. Jeong and G. Biswas, "Mining student behavior models in learning-by-teaching environments.," in *EDM*, pp. 127-136, Citeseer, 2008
- [16] PavelZezula, Giuseppe Amato, VlastislavDohnal, and Michal Batko. *Similarity Search: The Metric Space Approach*, volume 32 of *Advances in Database Systems*. Springer,2006. ISBN0-387-29146-6.
- [17]JaroslavBayer,HanaBydžovská,JanGéryk,TomášObšíváč,andLubomírPopelínský. Predictingdrop-outfromsocialbehaviourofstudents. InProceedingsofthe5th InternationalConferenceonEducationalDataMining-EDM2012,pages103-109, 2012. ISBN978-1-74210-276-4.
- [18] G.W. Dekker, M. Pechenizkiy, and J.M. Vleeshouwers. Predicting students drop out: a case study. In T. Barnes, M. Desmarais, C. Romero, and S. Ventura, editors, *Proceedings of the 2nd International Conference on Educational Data Mining (EDM'09)*,pag
- [19] JoaoGamaandPetrKosina. Learningdecisionulesfromdatastreams. InProceedingsoftheTwenty-SecondInternationalJointConferenceonArtificialIntelligence, pages1255-1260,2011. ISBN978-1-57735-512-0.es41-50,2009.
- [20] SuchitaBorkar, K. Rajeswari, "Predicting Students Academic Performance Using Education Data Mining", *IJCSCMC*, Vol. 2, Issue. 7, July 2013
- [21] M. Sukany et al,DataMining:performance improvement in education sector using classification & clustering algorithm,ICCCE ,April 2012
- [22] C. Romero, S. Ventura, P. G. Espejo, and C. Hervás, "Data mining algorithms to classify students.," in *EDM*, pp. 8-17, 2008.
- [23] C. Wallace, K. B. Korb, and H. Dai, "Causal discovery via mml," in *ICML*, vol. 96, pp. 516-524, Citeseer, 1996.
- [24] Ajay Kumar Pall, Saurabh Pal, "Evaluation of Teacher's Performance: A Data Mining Approach", *IJCSCMC*, Vol. 2, Issue. 12, December 2013,
- [25] J. R. Quinlan, "Introduction of decision tree: Machine learn", 1: pp. 86106, 1986.