

Applications of Data Mining Techniques in Student Performance Analysis A Survey

Nilar Aye

University of Computer Studies (Hpa-an)



Abstract - Educational Data Mining (EDM) is one of the useful applications of data mining. In EDM, there are two key research areas, one is related with students (both high school students and university students) and the other is with faculty (both teachers and institutional responsibility). A detailed research work has already been carried out in both domains. In the present day, the enormous challenges that educational institutions are facing the explosive growth of educational data and to use this data to enhance the quality of managerial decisions. Educational institutions are taking part in an important role in our society and playing a fundamental role for growth and development of nation. It is important to study and analyze educational data especially students' performance.

To analyze the performance of the students, every institution has their own criteria. So, it is necessary to study on existing prediction techniques and hence to find the best prediction methodology for predicting the student academics progress and performance. A detailed literature survey on predicting student's performance using data mining techniques has already been carried out in analyzing educational data. The main objective of this survey is to provide a great knowledge and examining different data mining techniques that have been applied to analyze and predict students' performance.

Keywords - Educational Data Mining, Student Performance, Data Mining, Prediction, Classification Algorithm, Naïve Bayes algorithm.

I. INTRODUCTION

Education is a basic requirement to develop the country. The ability to change and to induce change and progress in society depends on education. The educational institutions are key resources for producing the good students to provide better services for the society. The main aim of higher education institutes is to give excellence education to its students and to improve the quality of managerial decisions [18]. Student Performance is measured how well the student has performed in class, and how well the student has understood the provided material. The quality of education delivered by educational institutes is measured by the performance achieved by the students. A quality education provides students with the skills they have to employ as productive, innovative and responsible individuals from society [6]. Education's quality was evaluated by rate of success student's and to what degree organization was able

of preserved students. In education sector, the capability to predict performance of student was important [13].

Data mining is a computer-aided process of digging through and analyzing enormous sets of data and then extracting the meaning of the data. Data mining techniques predict behaviors and future trends, allowing decision makers to make proactive, knowledge driven decision. Educational data mining is a recent research area that focuses on the use of data mining techniques to transform large volumes of educational data into useful and relevant knowledge that can improve the educational processes and decisions [5].

To provide quality education to learners, deep analysis of previous records of the learners can play a vital role. This paper aims to identify related factors and how predicting the performance of students can be done in educational domain by using different datamining methods. Predicting the student performance is the most essential tool to provide the

students and schools in improving their learning and teaching process.

The main objectives to this proposed work are:

- i. To analyse the different usage of various prediction techniques of data mining in education
- ii. To determine different student attributes which are mainly useful for the predicting the student performance
- iii. To review the different prediction methods which are mainly effective for predicting the student's performance.

II. COMPREHENSIVE REVIEW OF LITERATURE

This survey paper presents a detailed comprehensive literature review of major important researches in the area of Educational Data Mining for prediction of student performance. This survey will provide the teachers and learners determine which datamining algorithm is suitable with existing students' parameters and they can realize

which parameters of the students are stored to predict their performance.

The referred work is presented below in a categorized tabular form Table 1. Table 1 show the complete survey based on evaluation of faculty performance of student dataset with various parameters using different data mining techniques in prior research studies from different perspectives.

This survey performed on the basis of the following categories:

- (1) Objective of the research
- (2) Parameters used for finding the faculty performance evaluation
- (3) Methodology used in the research papers for evaluation of student performance
- (4) Techniques and tools applied for evaluation
- (5) Data Mining tools used for analysis and visualization
- (6) Finding or conclusions of the research

Table 1: A survey of Research Papers

Sn o.	(1) Year & Author	(2) Paper Name	(3) Objectives	(4) Parameters Used	(5) Methodolog y	(6) Techniq ue(s) and Tool Used	(7) Findings
1.	2017 A. K Hamoud, A.M. Humadi, W.A. Awadh, A. S. Hashim [1]	Students' Success Prediction based on Bayes Algorith ms	To analyze collective student information via a questionnaire and to classify the collected data to predict and categorize student performance	The questionnaire consists of 62 questions that cover the fields affecting students' performance the most. The questions refer to health, social activity, relationships and academic performance.	The first stage is finding the most correlated questions to the final class, and the second is applying algorithms and finding the optimal algorithm.	Two built Bayes Algorith ms (naïve Bayes and Bayes network) and Weka 3.8 were used.	The naïve Bayes algorithm is selected as an optimal choice for students' success prediction.
2.	2015 F.Ahmad, Nur Hafieza Ismail, Azwa Abdul Aziz [12]	The Prediction of Students' Academic Performan ce Using Classificat ion Data Mining	To propose a framework for predicting students' academic performance of first year bachelor students in Computer	8 year period intakes from July 2006/2007 until July 2013/2014 that contains the students' demographics, previous academic records, and family background	Decision Tree, Naïve Bayes, and Rule Based classification techniques are applied to the students' data in order to produce	Decision Tree, Naïve Bayes, and Rule Based	1. Rule Based is a best model among the other techniques. 2. The extracted knowledge from

		Techniques	Science course	information	the best students' academic performance prediction model.		prediction model will be used to identify and profile the student to determine the students' level of success in the first semester.
3.	2019 D. Buenaño-Fernández, David Gil and Sergio Luján-Mora [7]	Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study	To predict the performance of students based on their historical performance of grades by using the effectiveness of machine learning techniques	Student dataset consist of the academic records of 335 students. The total number of historical records of students' grades was 6358. The periods analyzed were from the semester 2016-1 to the semester 2018-2 in the Computer Systems Engineering Degree of a university in Ecuador. In addition, the dataset comprises a total of 68 subjects organized into seven knowledge areas (Programming and Software Development, Mathematics and Physics, Information Network Infrastructure, Electronics, Databases, Economy—Administration,	1. The collection and data cleansing of historical datasets of student grades takes place. 2. The methods of machine learning and data mining are selected. 3. The model for predicting student grades is generated from previously processed data. 4. The results obtained are analyzed and visualized	Machine learning tools, the decision tree algorithm, the visualization	To establish the best correlations between the input variables and the result, which is the prediction of whether the student will pass a certain subject or not.

				General Education— Languages)			
4.	A. Daud,N. Radi Aljohani, R. Ayaz Abbasi [2]	Predicting Student Performance using Advanced Learning Analytics	To predict whether a student will be able to complete his degree or not To investigate two new feature sets family expenditures and student personal information To construct an effective feature set of twenty-three features by combining proposed features along with exiting features.	The dataset consists of 100 student records (tuples) and 23 features such as the characteristics of family expenditure, family income, personal information and family assets of students.	1. The feature subset selection process is adapted by using information gain and gain ratio metrics. 2. Five-fold cross validation method 3. Five classifiers are used (BN, NB, SVM, C4.5 and CART) to analyze the influence of each feature for predicting the performance of students	Five-fold cross validation method and Five classifiers are used (BN, NB, SVM, C4.5 and CART)	1. The proposed method significantly outperforms existing methods due to exploitation of family expenditures and students' personal information feature sets. 2.Outcomes of this EDM/LA research can serve as policy improvement method in higher education. 3.Learning Analytics based on personalized features can improve prediction of students' performance .
5.	2017 Dr. K. Karthikeyan, P. Kavipriya [9]	On Improving Student Performance Prediction in Education Systems using	1. To design and implement a Student Performance Prediction System (SPPS) based on enhanced feature selection and ensemble	Dataset 1 is achieved from secondary education of two Portuguese schools (Number of records: 395, Number of Attributes: 33) The data attributes	Enhanced SVM classifier is used to cluster the student training set using enhanced K-Means	Clustering (Enhanced K-Means) and Classification (Support Vector	1.The proposed CESVM worked better than the conventional SVM classifier 2. the 2-step

		Enhanced Data Mining Techniques	classification algorithms using historical academic data. 2. To propose a feature selection algorithm based on swarm optimization techniques to improve the quality of features (or student attributes) and to optimize the performance of SPPS using a clustering-based Enhanced SVM classifier (CESVM) to improve its classification accuracy and speed of classification. 3. To propose an enhanced ensemble SPPS based on CESVM that uses a 2-step procedure and a pruning algorithm to further improve prediction accuracy and speed of ensemble system	include student grades, demographic, social and school related features Dataset 2 is taken from CMS College of Science and Commerce, Coimbatore .Size of the dataset (Number of Students: 650, Number of Attributes: 52) students' internal marks	algorithm. The k-means algorithm is enhanced to find the initial K value automatically .	Machine)	classifiers performance was higher when compared to CESVM and the three conventional classifiers
6.	2013 D.Kabackchieva [8]	Predicting Student Performance by Using Data Mining	To find out if there are any patterns in the available data that could be useful for predicting	The dataset contains about 10330 students, described by 20 parameters, including gender, birth year, birth	Each classifier is applied for two testing options – cross validation	WEKA classifiers are used in the experimental	The decision tree classifier (J48) performs best (with the

		Methods for Classification	students' performance at the university based on their personal and pre-university characteristics	place, living place and country, type of previous education, profile and place of previous education, total score from previous education, university admittance year, admittance exam and achieved score, university specialty/direction, current semester, total university score, etc.	(using 10 folds and applying the algorithm 10 times – each time 9 of the folds are used for training and 1 fold is used for testing) and percentage split (2/3 of the dataset used for training and 1/3 – for testing). The WEKA J48 classification filter is applied on the dataset during the experimental study.	study, including a common decision tree algorithm C4.5 (J48), two Bayesian classifiers (NaiveBayes and BayesNet), a Nearest Neighbor algorithm (IBk) and two rule learners (OneR and JRip).	highest overall accuracy), followed by the rule learner (JRip) and the k-NN classifier. The Bayes classifiers are less accurate than the others.
7.	2014 R. Singh Yadav, A. K. Soni, Saurabh Pal, Ph.D [17]	Implementation of Data Mining Techniques to Classify New Students into Their Classes: A Bayesian Approach	To solve the allocation problem of new students by using Bayesian classification.	The training data sets contain 50 students whose are described by attributes: end semester marks, class test grade, seminar performance, assignment, general proficiency, attendance, and lab work. The class label attribute and class type, have four distinct values namely first, second, third and fail.	Bayesian, the statistical classifiers, can predict class membership probabilities such as the probability that a given tuple belongs to a particular class. Bayesian classifiers assume an attribute value on a	Bayesian classifier	The performance of Bayesian classification technique is more suitable compared to rest of techniques such as genetic algorithm method.

					given class which is independent of the values of the other attributes.		
8.	2014 Elakia, Gayathri, Aarthi, Naren J [11]	Application of Data Mining in Educational Database for Predicting Behavioural Patterns of the Students	To suggest career options for the high school students and also to predict the potentially violent behaviour among the students by including extra parameters other than academic details	High school students based on every student's interests, skills, likes, hobbies etc.	ID3 algorithm could be applied on them to learn the rules found on the training set and their corresponding labels.	1.classifiers like ID3 (Iterative Dichotomizer) and C4.5 algorithm 2.Rapid Miner (Data mining tool)	It was evident that ID3 outperforms every other decision trees. Hence ID3 algorithm was considered to be an applicable algorithm.
9.	2015 N. Puri, Deepali Khot , Pratiksha Shinde, Kishori Bhoite , Prof. Deepali Maste [16]	Student Placement Prediction Using ID3 Algorithm	To identify relevant attributes based on academics, skills and curricular of final year student and design a model which can predict placement of the student using a classification technique based on decision tree	10th %, 12th % or Diploma, Degree Aggregate, Sem Performance, Communication Skill, Work on Projects, Internship, Education Gap, Backlog, Placement	Information Gain and entropy is calculated.	ID3 algorithm, C4.5, and CHAID	ID3 decision tree algorithm is best classifier with 95% accuracy
10.	S.Jayaprakash, Balamurugan E.,Vibin Chandar [19]	Predicting Students Academic Performance using Naive Bayes Algorithm	To predict student's academic performance at the end semester exams by analyzing students feedback and their	The data was collected from 700 students in various departments of Blue Crest College, Accra, Ghana in the academic year 2014 with the internal examination score. The data are	Supervised learning approach through Naives Bayes algorithm	Naive Bayes algorithm	The comparison between feedback and internal examination marks. Navie Bayes algorithm gives the better

			performance in the mid-semester exams and also help the educational institutions to identify the weaker students in advance and arrange necessary training before they are going to appear for their final exams.	collected by means of feedback rating-scale questionnaire related to teaching and learning process of an institute.			prediction result.
--	--	--	---	---	--	--	--------------------

III. CONCLUSION

Data mining is an extremely vast area that includes applying different techniques and algorithms for pattern finding. Because of the data growth progresses, there were some difficulties for people to analyse and process large amount of data but with the help of data mining and machine learning techniques the analysis of such a huge amount of data has become easier. This paper has surveyed previous studies on predicting students performance with various analytical methods. According to the survey of applications on various Datamining methods, ID3 algorithm is the best predictor to predict the student performance. This survey showed how useful data mining can be applied in education particularly to improve students' performance.

REFERENCES

- [1] A. K Hamoud, A. M. Humadi, W. A. Awadh, A. S. Hashim, "Students' Success Prediction based on Bayes Algorithms", International Journal of Computer Applications (0975 – 8887) Volume 178 – No.7, November 2017
- [2] A. Daud,N. Radi Aljohani, R. Ayaz Abbasi, "Predicting Student Performance using Advanced Learning Analytics", Conference Paper · April 2017 DOI: 10.1145/3041021.3054164
- [3] A. Abu Saa, "Educational Data Mining & Students' Performance Prediction", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 5, 2016
- [4] A. Ashrafa, S. Anwerb, M. Gufran Khanc, "A Comparative Study of Predicting Student's Performance by use of Data Mining Techniques", American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS) ISSN (Print) 2313-4410, ISSN (Online) 2313-4402
- [5] C.Anuradha, T.Velmurugan, R. Anandavally, "Clustering Algorithms in Educational Data Mining: A Review", International Journal of Power Control and Computation(IJPCSC)Vol 7. No.1 – 2015 Pp.47-52©gopalax Journals, Singapore
- [6] C. Simon , Y. Bugusa, "Survey on data mining approach for analysis and prediction of student performance", International Journal of Engineering & Technology, 7 (4.5) (2018) 467-470 International Journal of Engineering & Technology Website: www.sciencepubco.com/index.php/IJET
- [7] D. Buenaño-Fernández , David Gil and Sergio Luján-Mora, "Application of Machine Learning in Predicting Performance for Computer Engineering Students:A Case Study", Sustainability 2019, 11, 2833; doi:10.3390/su11102833
- [8] D. Kabakchieva, "Predicting Student Performance by Using Data Mining Methods for Classification", ISSN: 1311-9702; Online ISSN: 1314-4081 ,DOI: 10.2478/cait-2013-0006
- [9] Dr. K. Karthikeyan,P. Kavipriya, "On Improving Student Performance Prediction in Education Systems using Enhanced Data Mining Techniques".
- [10] Dr. P. Nithya, B. Umamaheswari, A. Umadevi, "A Survey on Educational Data Mining in Field of Education", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 1, January 2016
- [11] Elakia, Gayathri, Aarthi, Naren J, "Application of Data Mining in Educational Database for Predicting

- Behavioural Patterns of the Students”, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 4649-4652, ISSN-09759646
- [12] F. Ahmad, Nur Hafieza Ismail and Azwa Abdul Aziz, "The Prediction of Students' Academic Performance Using Classification Data Mining Techniques", Applied Mathematical Sciences, Vol. 9, 2015, no. 129, 6415 – 6426 HIKARI Ltd, www.m-hikari.com <http://dx.doi.org/10.12988/ams.2015.53289>
- [13] F. Razaque, Safeeullah Soomro, "Using Naïve Bayes Algorithm to Students' bachelor Academic Performances Analysis", Conference Paper . November 2017
DOI: 10.1109/ICETAS.2017.8277884
- [14] M. Durairaj , C. Vijitha, "Educational Data mining for Prediction of Student Performance Using Clustering Algorithms", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5(4), 2014, 59875991, ISSN: 09759646
- [15] M. Kumar, Prof. A.J. Singh, Dr. Disha Handa, "Literature Survey on Student's Performance Prediction in Education using Data Mining Techniques", I.J. Education and Management Engineering, 2017, 6, 40-49 Published Online November 2017 in MECS (<http://www.mecspress.net>)
DOI: 10.5815/ijeme.2017.06.05
- [16] N. Puri , Deepali Khot , Pratiksha Shinde , Kishori Bhoite , Prof. Deepali Maste, “Student Placement Prediction Using ID3 Algorithm”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), ISSN: 2321-9653, Volume 3 Issue III, March 2015
- [17] R. Singh Yadav, A.K. Soni, Saurabh Pal, Ph.D, “Implementation of Data Mining Techniques to Classify New Students into Their Classes: A Bayesian Approach”, International Journal of Computer Applications (0975 – 8887) Volume 85 – No 11, January 2014
- [18] Sana, Isma Farrah Siddiqui, Qasim Ali Arain, "Analyzing Students' Academic Performance Through Educational Data Mining", Tecnología. Glosas de innovación aplicadas a la pyme. Special Issue, May 2019, pp.402–421. doi: <http://dx.doi.org/10.17993/3ctecno.2019.specialissue2.402-421>
- [19] S. Jayaprakash, Balamurugan E., Vibin Chandar, “Predicting Students Academic Performance using Naïve Bays Algorithm”.