Predicting Student's Performance Using Data Mining Techniques: A Survey From 2002 To 2020

Vinod Kumar Patel, Dr. Mahesh Pawar, Dr. Sachin Goyal

Abstract: Today, many educational institutions suffer from the issue of dropping out students, failing students, recognize poor students because of the lack of a proper framework for assessing and tracking the success and performance of students. This is one of the main challenges of the educational institution, since predicting the performance of students is difficult due to vast volumes of data in educational databases. Predicting student's performance at an educational institution is mostly useful in helping the institute management to make strategy and decision making related to improving student performance. Data Mining is one of the efficient methods for predicting student's performance in large educational databases. Data Mining is applied in the field of education to predict student's performance. Different data mining methods and techniques are used for predicting student's performance. This paper present a literature research on data mining methods used to predict student's performance from 2002 to 2020. This paper reviews work done by different researcher to predict student's performance in all perspective. This paper also discusses commonly used attributes in different research for the student performance analysis.

Index Terms: Student performance, prediction, educational data mining, data mining techniques.

1 INTRODUCTION

The standard of any university institution depends on its student's academic achievement. Therefore, the performance of the students is very essential for the institute. Student performance can be assessed by previous academic record, financial, social and other factors. The educational institution has to maintain records of the students resulting in a large database. Analyzing such massive data presents significant challenges for educational institutions to predict student's performance [1]. Data mining is a popular method for analyzing large amounts of data in the present day. It is the process of uncovering hidden patterns from huge volume of data. Data Mining has the potential to predict student success from a large amount of educational databases. Recently Data Mining has been applied to student performance analysis in an educational institution. It's called educational data mining. Educational data mining techniques are the most powerful tool for forecasting student performance [2], [3], [4]. Educational data mining predicts the performance of students from vast educational datasets. Educational data mining has the potential to benefit both learners and educators. By predicting student performance academic institutions can tracked the student's progress and analyze their performance and planned policies for the future related to student performance. Educators could classify poor students and help the institute for building strategy to improve the performance of the students. Also learners can achieve better score and improve their learning skills and then increase efficiency of the teaching process.

- *Vinod Kumar patel is currently pursuing PhD in Computerscience & Engineering in UIT RGPV Bhopal,m.p.,India E-mail vinod.gwl@gmail.com*
- *Dr. Mahesh Pawar,Assistant Professor Department of Information Technology ,UIT RGPV Bhopal ,m.p.,India, E-mail: mkpawar24@gmail.com*
- *Dr.Sachin Goyal, Assistant Professor Department of Information Technology ,UIT RGPV Bhopal ,m.p.,India, E-mail:sachingoyal@rgtu.net*

2 MODEL FOR PREDICTING STUDENT PERFORMANCE

The system for predicting student performance using data mining is shown in fig.1.

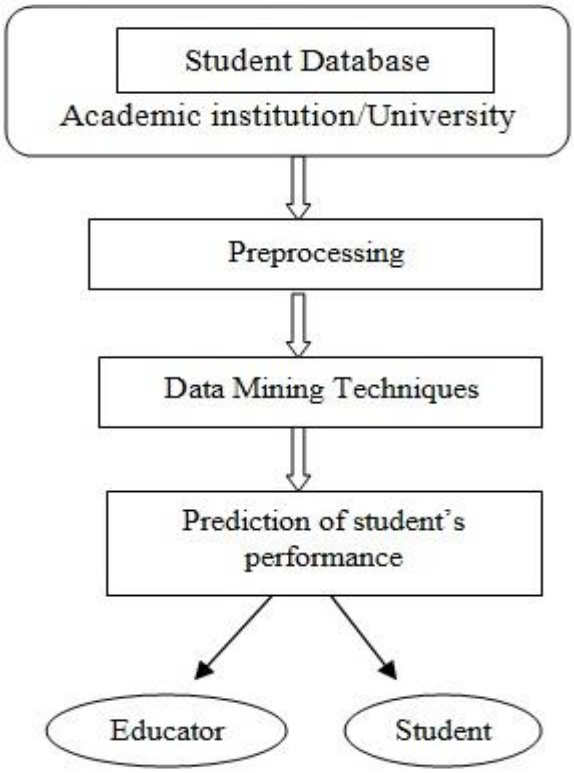

fig1. *Model for predicting student's performance using data mining*

Throughout this system, the academic institutions have students record as course information, co-curriculum information and other data for their students. Such academic data set has to be pre-processed before data mining is implemented and then data mining methods used to forecast student performance. And this predicted information is provided to educators so they can make performance improvement techniques and decision-making relevant to

the students. And also present this information in the form of advice to students to enhance future results.

3 ATTRIBUTES USED IN PREDICTING STUDENT'S PERFORMANCE

The two main factors used to predict student performance are the attributes and prediction methods [5]. In this section we will discuss about the attributes used to predict student performance. We'll discuss about prediction techniques in the next section. Table 1 provides a list of common attributes, along with methods used to predict student's performance.

TABLE 1: *LIST OF COMMON ATTRIBUTES USED FOR PREDICTING STUDENT PERFORMANCE.*

Used attributes	Reference No.	Data Mining task
CGPA	[6],[8],[9],[10],[11],[12],[14]	Decision tree, Neural network, naïve byes
Assessments, quizzes marks, lab work, attendance, class test	[7],[8],[10],[14],[15],[16]	Decision tree
Gender, age, family background, disability	[13], [16], [11], [17], [18], [8]	Decision tree
Final exam marks	[10], [15], [19], [20], [17], [8]	Decision tree
Extra-curricular activities	[10],[16],[7],[8],[21]	Decision tree
High school background	[17],[18]	Decision tree
Social interaction network	[22],[19],[23]	Decision tree
Psychometric factor	[24], [25], [21]	Decision tree, Neural network, k-Nearest Neighbor
Pre-admission tests	[27], [26]	Neural network, decision tree, support vector machine, naïve bayes

It is shown in [5] that student CGPA and internal assessment are the attributes most commonly used in student performance prediction at institutions. CGPA has been used by the researchers in [6],[8],[9],[10],[11],[12] as the prime attributes for predicting student performance. This is because CGPA provides the true values of education and career opportunities for the students [13]. It is shown that student attendance and CGPA attributes with student performance have a strong impact [14]. Assessment, quizzes, laboratory work, attendance, class test marks all attributes are often called internal assessment and often used by researchers to predict student performance [7][8][10][14][15][16]. Student demographics include gender, age, family history, disability [13],[16], [11], [17], [18], [8], and external assessments including final examination grade are the next most frequently used attribute for student performance prediction [10][15][19][20][17][8]. It is noted that many researchers use the gender factor, since male and female students have different learning processes [13]. Next, Extra activities[10],[16],[7],[8],[20] high school background[17],[18] and social interaction networks[22],[19],[23] are the three attributes often used in predicting student performance. Several researchers also use psychometric variables such as student interest, student attitude, commitment time and family support for student performance prediction [24],[25],[21]. M.Mayilvaganan and D.Kalpanadevi [7] made study and found that performance of students can be assessed by their personal interest and behaviour. In another analysis there are also several researchers who have found pre-admission test score in universities for student performance prediction [27],[26].

4 METHOD USED FOR PREDICTING STUDENT'S PERFORMANCE

Predictive data mining technology is used to forecast student success in academic institutes. Classification, regression, and categorization are predictive tasks. Many Data Mining classification techniques are available for classification tasks. Decision tree, Neural Network, Naive Bayes, Support Vector Machine and K-Nearest Neighbor are mostly used for student performance prediction classification techniques.

4.1 NEURAL NETWORK

Neural networks are popular methods used to predict performance of students. Neural networks process data in the same way as human brain does. Neural networks are a group of interconnected nodes. This group of nodes in a neural network is similar to the way neurons are in the brain. Researcher published several papers using a neural network to predict the success of students. Wang and Mitrovic [30] predict student performance using feed-forward network, with four inputs, a single hidden layer, and a single output with high accuracy. Arsad et al. [28] predict academic success of engineering background students on CGPA attributes at semester eight based on results of fundamental subjects in the earlier semester and conclude that fundamental subjects at earlier semester have high influence in the final CGPA upon graduation. Some researchers used several methods and comparing results [7],[29],[25],[27]. The researcher used multilayer perception for neural network on pre-admission test attribute data and achieved higher accuracy compared to Decision Tree, Support Vector Machine and Naïve Bayes techniques [27].

4.2 DECISION TREE

One of the most common methods in predicting student success is the decision tree. The decision tree is a system similar to a flow-chart it has a root node, branches, and leaf nodes. Internal nodes denote a test on an attribute, branch denotes a test outcome, and each leaf node carries a class label. Many researchers used decision tree as their method to evaluate student performance because of its simplicity and comprehensibility in uncovering small and broad data structures and forecasting the value [8]. Quadri and kalyankar [31] uses J4.8 algorithm on WEKA software and analyze the academic performance of graduate students on CGPA basis and estimate the list of students who are likely to drop out college because of poor performance. Mishra et al. [21] uses two algorithms J48 and Random Tree implemented on WEKA software for measures the performance of MCA third semester students. Elakia et al. [16] uses various Decision Tree algorithms implemented on Rapid Miner to predict behavioural patterns of the high school students. Researcher discovers individual student characteristics that are associated with their success [1]. Anuradha and Velmurugan [32] include works of prediction

of the performance of students in end semester university examinations and present report that state the decision tree presents rules that can be clearly understood and interpreted by the user and that work well for categorical and numerical features and does not require complicated data planning. Migueis et al. [6] predict student's overall academic performance using the information available at the end of the first year of students' academic career. Researcher predicts the success rate of students enrolled in their courses [8]. In [27] research work includes predict applicant's academic performance at university.

4.3 NAÏVE BAYES

Naïve bayes relies on Bayes' theorem of probability to forecast the class of unknown data sets assuming the presence of a particular feature in a class is unrelated to any other feature being present. Naive Bayes forecasts different class probability based on different attributes. Dorina Kabakchieva [33] said that due to simplicity, efficiency, very good performance and fast to predict many researchers chooses this technique to make a prediction. The researchers uses naïve bayes methods along with several others methods and compare results to find most efficient prediction method in prediction of student performance [7][6][29][25][27].

4.4 K-NEAREST NEIGHBOR

K-Nearest Neighbor is a classification technique based on feature similarity. It predicts the classification of a new object based on how closely out of sample features resemble training set. The research paper used K-Nearest Neighbor to predict performance of the students [25],[7],[33],[32],[29]. Marbouti et al. [29] state that K-Nearest Neighbor shown higher overall accuracy for predicting students who have passed the course compared to other methods; however, K-Nearest Neighbor performs poorly in correctly identifying students at risk.

4.5 SUPPORT VECTOR MACHINE

The purpose of this classification technique is to construct a hyperplane in an N-dimensional space (N — the number of features) that classifies the data points distinctly based on their classes. By constructing a hyperplane which has the maximum distance between data points in both classes, it separates the two classes in data points. Support Vector Machine has been used by few researchers as their tool for predicting student success. In research [24] Smooth Support Vector Machine is used which is Support Vector Machine's further development to predict the final grade of the student and state that Smooth Support Vector Machine is faster than other methods and has better generalization capability. In [6][29][25] Support Vector Machine is used for forecasting student success with other classification techniques. Authors use the above classification methods, are summarized in the table below.

TABLE 2: *CLASSIFICATION METHODS USED BY AUTHORS*

Author	Classification Methods used	Published Year
Wang ,Mitrovic [30]	Neural Network	2014
Arsad et al. [28]		2013
Natek, Zwilling [8]	Decision Tree	2014
Quadri, Kalyankar [31]		2010
Mishra et al. [21]		2014
Elakia et al. [16]		2014
Guruler et al. [1]		2010
Sembiring et al. [24]	Support Vector Machine	2011
Mayilvaganan,Kalpanadevi [7]	Naïve Bayes,K-Nearest Neighbor	2014
Marbouti et al. [29]	Naïve Bayes,Support Vector Machine,K-Nearest Neighbor	2016
Gray et al. [25]	Naïve Bayes,Decision tree,Support Vector Machine, Neural Network,K-Nearest Neighbor	2014
MENGASH [27]	Neural Network, Decision tree, support Vector Machine, Naïve Bayes	2020
Anuradha, Velmurugan [32]	Decision Tree, Naïve Bayes, K-Nearest Neighbor	2015
Migueis et al. [6]	Decision Tree, Support Vector Machine,Naïve Bayes	2018
Kabakchieva [33]	Decision Tree, Naïve Bayes, K-Nearest Neighbor	2013
Ramesh et al.[18]	Naïve Bayes, Neural etwork, Decision Tree	2013

5 CONCLUSION

Student performance prediction helps educational institute strengthen our educational system. In this paper an effort is made to present a systematical review of literature on predicting student performance with the help of data mining. This paper also reviewed attributes which influences the student performance. This paper carried out our review work from 2001 to 2020. Most of the researchers used CGPA and internal assessment marks as their key attribute to predict student performance. Neural Network, Decision tree, Naïve Bayes, K-Nearest Neighbor and Support Vector Machine are the classification methods used for prediction of student performance. The Majority of research paper used Decision tree and Neural Network classification methods because of their high accuracy for student performance prediction.

REFERENCES

[1] H. Guruler, A.Istanbullu, M. Karahasan, "A new student performance analysing system using knowledge discovery in higher educational databases," Comput. Edu., vol. 55, no. 1, pp. 247–254, Aug. 2010.

[2] C. Romero and S. Ventura, ''Educational data mining: A review of the state of the art,'' IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 40, no. 6, pp. 601–618, Nov. 2010.

[3] S. K. Mohamad and Z. Tasir, ''Educational data mining: A review,'' Procedia Social Behav. Sci., vol. 97, pp. 320–324, Nov. 2013.

[4] A. Peæa-Ayala, ''Educational data mining: A survey and a data mining based analysis of recent works,'' Expert Syst. Appl., vol. 41, no.4,pp. 1432–1462, Mar. 2014.

[5] A. M. Shahiri, W. Husain, and N. A. Rashid, ‘‘A review on predicting Student’s performance using data mining techniques,’’ Procedia Comput.Sci., vol. 72, pp. 414–422, Jan. 2015.
[6] V. L. Migueis, A. Freitas, P. J. V. Garcia, and A. Silva, ‘‘Early segmentation of students according to their academic performance: A predictive modelling approach,’’ Decis. Support Syst., vol. 115, pp. 36–51,Nov. 2018.
[7] M. Mayilvaganan and D. Kalpanadevi, ‘‘Comparison of classification techniques for predicting the performance of students academic environment,’’ in Proc. Int. Conf. Commun. Netw. Technol., Sivakasi, India, Dec. 2014,pp. 113–118.
[8] S. Natek and M. Zwilling, ‘‘Student data mining solution–knowledge management system related to higher education institutions,’’ Expert Syst.Appl., vol. 41, no. 14, pp. 6400–6407, Oct. 2014.
[9] K. F. Li, D. Rusk, F. Song, “Predicting student academic performance,” Complex, Intelligent, and Software Intensive Systems (CISIS),Seventh International Conference on, IEEE, 2013, pp. 27–33.
[10] D. M. D. Angeline, “Association rule generation for student performance analysis using apriori algorithm,” The SIJ Transactions on Computer Science Engineering & its Applications (CSEA) vol.1, no.1, 2013, p12–16.
[11] Z. Ibrahim and D. Rusli, “Predicting students academic performance: comparing artificial neural network, decision tree and linear regression,”21st Annual SAS Malaysia Forum, 5th September, 2007.
[12] N. Thi Ngoc Hien and P. Haddawy, ‘‘A decision support system for evaluating international student applications,’’ in Proc. 37th Annu. Frontiers Edu.Conf. Global Eng., Knowl. Borders, Opportunities Passports, Milwaukee,WI, USA, Oct. 2007, pp. F2A-1–F2A-6.
[13] U. bin Mat, N. Buniyamin, P. M. Arsad and R. Kassim, "An overview of using academic analytics to predict and improve students’ achievement:A proposed proactive intelligent intervention,”Engineering Education (ICEED), 2013 IEEE 5th Conference on, IEEE, 2013, pp. 126–130.
[14] Annisa Uswatun Khasanah ,Harwati , “A comparative study to predict student’s performance using Educational data mining techniques”in IOP Conf. series: Material Science and Engineering 215(2017)012036.
[15] S. Parack, Z. Zahid and F. Merchant, “Application of data mining in educational databases for predicting academic trends and patterns,”Technology Enhanced Education (ICTEE), 2012 IEEE International Conference on, IEEE, 2012, pp. 1–4.
[16] G. Elakia and N. J. Aarthi, “Application of data mining in educational database for predicting behavioural patterns of the students,” Elakia et al,/(IJCSIT) International Journal of Computer Science and Information Technologies vol.5,no.3, 2014,pp. 4649–4652.
[17] V. Oladokun, A. Adebanjo and O. Charles-Owaba, “Predicting students academic performance using artificial neural network: A case study of an engineering course,” The Pacific Journal of Science and Technology, vol.9, no.1, 2008, pp. 72–79.
[18] V. Ramesh, P. Parkavi and K. Ramar, “Predicting student performance: a statistical and data mining approach,” International Journal of Computer Applications, vol. 63, no.8, 2013, 35–39.
[19] A. Bogar´ın, C. Romero, R. Cerezo and M. Sanchez-Santill ´ an, “Clustering for improving educational process mining,” Proceedings of the Fourth International Conference on Learning Analytics And Knowledge, ACM, 2014, pp. 11–15.
[20] C. Coffrin, L. Corrin, P. de Barba and G.Kennedy, “Visualizing patterns of student engagement and performance in moocs,” Proceedings of the fourth international conference on learning analytics and knowledge, ACM, 2014, pp. 83–92.
[21] T. Mishra, D. Kumar and S. Gupta, “Mining students’ data for prediction performance,” Proceedings of the 2014 Fourth International Conference on Advanced Computing & Communication Technologies, ACCT ’14, IEEE Computer Society, Washington, DC, USA, 2014, pp. 255–262.
[22] C. Romero, M.-I. Lopez, ´ J.-M. Luna and S. Ventura, “Predicting students’ final performance from participation in on-line discussion forums,”Computers & Education 68 (2013) 458–472.
[23] N. Thai-Nghe, T. Horvath, ´ L. Schmidt-Thieme, “Factorization models for forecasting student performance,” Educational Data Mining 2011, 2010.
[24] S. Sembiring, M. Zarlis, D. Hartama, S. Ramliana, E. Wani, “Prediction of student academic performance by an application of data mining techniques,” International Conference on Management and Artificial Intelligence IPEDR, Vol. 6, 2011, pp. 110–114.
[25] G. Gray, C. McGuinness, P. Owende, “An application of classification models to predict learner progression in tertiary education,” Advance Computing Conference (IACC), 2014 IEEE International, IEEE, 2014, pp. 549–554.
[26] A. M. Alhadlaq, O. F. Alshammari, S. M. Alsager, K. A. F. Neel, and A. G. Mohamed, ‘‘Ability of admissions criteria to predict early academic performance among students of health science colleges at King Saud University, Saudi Arabia,’’ J. Dental Educ., vol. 79, pp. 665–670,Jan. 2015.
[27] H.A.Mengash, "Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems," in IEEE Access, vol. 8, pp. 55462-55470, 2020.
[28] P. M. Arsad, N. Buniyamin, and J.-L.-A. Manan, ‘‘A neural network students’ performance prediction model (NNSPPM),’’ in Proc. IEEE Int.Conf. Smart Instrum., Meas. Appl. (ICSIMA), Kuala Lumpur, Malaysia, Nov. 2013, pp. 1–5
[29] F. Marbouti, H. A. Diefes-Dux, and K. Madhavan, ‘‘Models for early prediction of at-risk students in a course using standards-based grading,’’Comput. Edu., vol. 103, pp. 1–15, Dec. 2016.
[30] T. Wang and A. Mitrovic, “Using neural networks to predict student’s performance,” Computers in Education, 2002. Proceedings. International Conference on, IEEE, 2002, pp. 969–973.
[31] M. M. Quadri, N. Kalyankar, “Drop out feature of student data for academic performance using decision

tree techniques," Global Journal of Computer Science and Technology 10 (2).
[32] C. Anuradha and T. Velmurugan, ''A comparative analysis on the evaluation of classification algorithms in the prediction of students performance,''Indian J. Sci. Technol., vol. 8, no. 15, pp. 974–6846, Jan. 2015.
[33] D. Kabakchieva, ''Predicting student performance by using data mining methods for classification,'' Cybern. Inf. Technol., vol. 13, no. 1, pp. 61–72, Mar. 2013.