Educational Data Mining For Student Performance Analysis: A Review

Akash Thakur[1], Sunil Kumar[2], Pushpa Devi[3], Dr. K.L. Bansal[4]
[1, 2, 3] Dept of Computer Science
[4]Professor, Dept of Computer Science
[1, 2, 3, 4] Himachal Pradesh University, Shimla

Abstract- *To extract meaningful information and hidden insight from raw data, data mining techniques are used. The information thus obtained is valuable and have a significant impact on the decision-maker and decision-making process. Educational data mining (EDM) analyses a large amount of student data in educational settings using data mining techniques. EDM's main goal is to study and solve educational system problems like bringing education and technology together, it can be difficult to keep students' interest and attention to properly teach new concepts. Technology needs to come into the classroom to keep up with the learning demands of the 21st century. And also having well-educated tutors that will help them achieve their goals and guide them throughout their education. Educational Data Mining is a new field that aims to create ways for analyzing a vast volume of data from educational settings in order to better understand students' behavior, interests, and outcomes. Various publications relating to this specialization have been published in recent years, and multiple data mining approaches introduced in these publications have been employed to address a different kinds of educational difficulties. EDM can also be used to find out students' risk to be in danger of not graduating, or being promoted. And also enhances rates of students' completing their graduation, prioritizes requirements of learning for several groups, thoroughly investigates the performance of the institution and optimizes curriculum of topic renewal, and many more things. The main aim of this research is to give a proper review of the whole work in which several techniques related to data mining have been utilized to tackle specific educational challenges, as well as a comparative study of the existing data mining techniques which are used in the prediction of students' academic performance. The comparative analysis of several techniques can help in deciding which algorithm or technique will be suitable for better prediction in future work.*

Keywords- Data Mining, Education, Educational data mining (EDM), Students

I. INTRODUCTION

Because of the significant contributions that technology has made to all disciplines today, it's understandable that all aspects of life are changing, especially the way we communicate and the barriers that have been broken down. Thanks to the Internet and the vast amounts of data being generated and stored in various information systems. Many educational studies have included Data analysis and data mining because of the importance of educational data mining. As a result, more surveys for each of the EDM applications should be undertaken to more precisely identify the most effective strategies. Furthermore, in the evaluation, broader evaluation and comparison criteria should be applied. The sector of data mining in the educational field is now expanding and has many benefitssuch as integration with modern algorithms and methods developed in several areas of machine learning and data mining. EDM can also be used to find out the risk in students, enhance rates of graduation, prioritize requirements of learning for several groups, thoroughly investigate the performance of the institution and optimize curriculum of topic renewal, and many more things.

The major goal of educational institutions is to provide students with the knowledge and skills that will help them in their carrier. The effectiveness with which educational systems achieve this goal is a fundamental predictor of both social and economic advancement [3]. The technologies utilized in educational systems produce a large amount of data which is difficult to interpret with the naked eye [24]. Educational data mining (EDM) analyses student data in educational settings using various data mining methodologies [29].EDM's main goal is to study and solve educational problems in order to improve educational processes [18]. As a result, its purpose is to study educational data in order to resolve educational challenges. Within the educational profession, EDM entails extracting relevant, interpretable, entertaining, and unique information from data.

With the recent advent of educational data mining applications in educational settings, numerous strategies for

implementing these applications have been developed. As a result, this research looks into the EDM approaches used in educational data mining applications. The most common applications are predicting student performance, recognizing unwanted student behaviors, categorizing students, and student modeling. There are a variety of EDM applications, each having its own set of goals, such as enhancing and improving the quality of the learning and comprehending process [5]. Furthermore, EDM applications are aimed at a variety of educational stakeholders, including students, researchers, administrators, and educators. Students' learning can be enhanced by providing recommendations, personalization, and feedback.

II. DATA MINING

In this digital world, each second of data is generated from diverse locations which is a reason, mining and figuring out interesting patterns from a huge volume of data is such an important and complex task. Data mining is a tool that finds out valid patterns and meaningful information and hidden insight from huge records, in this way the above-mentioned problem can be solved. It can be defined as deriving facts from large datasets or we can say that it is a method used for investigating meaningful patterns from rough facts [16]. The derived knowledge can be used for different purposes like customer retention, weather forecasting, business forecasting, scientific exploration, fraud detection, risk management, space research, sports, and market analysis. Data mining provides an opportunity for decision-makers to make better decisions and predictions.

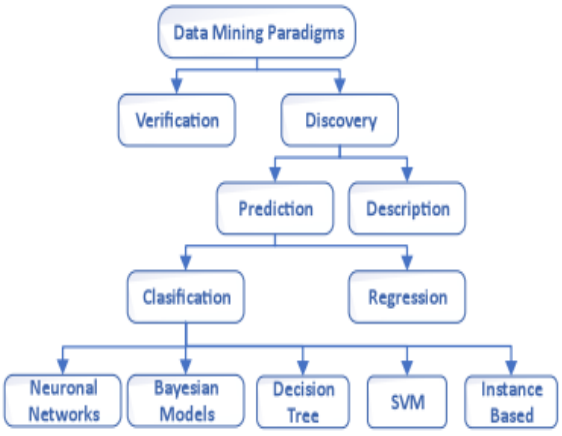

Figure 1:Data Mining Taxonomy

III. IMPLEMENTING A DATA MINING PROJECT

The implementation of data mining projects is done with the objective of discovering design and pertinent patterns with attractive information in high volumes. This is completed with the evolution of 4 stages [30],

1. Data filtration
2. Feature selection
3. Extraction of knowledge
4. understanding and measures

Basically, all the techniques ineducational setting have now been proven [25], and several cases have been studied in detail to understand the performance of the various techniques and to meet the key objective of data mining for education purposes which also includes identifying students'behavior patterns in their academic environments and classifying student types based on recorded performance.

IV. EDUCATIONAL DATA MINING

EDM educational data mining is a subs-field of DM which uses mining techniques for exploring valid and meaningful patterns by using educational dataset. Data is most crucial factor for making decisions and in educational organizations, for making decisions educational data is required. Mining educational data becomes a developing and up growing area of research and typically researchers take an interest in it[20].

The intent of educational DM is to transform raw/rough facts that are coming from academic institutions into valid information [33]. In the academic sector, the use of EDM helps the academic responsible and managers to improve and enhance their learning process.

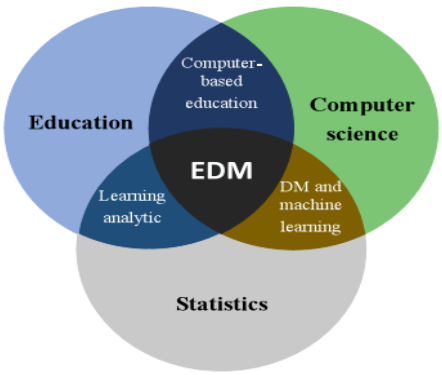

Figure 2: Main areas involved in Educational data mining. Source [19]

V. EDM METHODS

Educational Data Mining methods are the techniques that are used for the prediction of future data so that better decisions can be made. These methods can be predictive or

descriptive in nature. Various DM techniques are available that can be used for mining educational data. The most well-known techniques that are used in an academic context are sequential pattern mining, rule mining, clustering and classification [34].

VI. RELATED WORK

Several studies on Educational Data Mining (EDM) applications and methodologies in academic settings have been conducted, highlighting the usefulness of EDM for obtaining correct information about students' behavior and learning process efficacy [6]. So far, few survey studies on EDM have been published. Four areas of applications were recently counted in one of studies on EDM in higher education, namely forecastingthe performance of the students [24]. They highlighted EDM strategies that have been used to improve and analyze students' learning processes based on these applications. [5] suggested 13 types of EDM applications under student modeling, decision support systems. Many educational studies have included analysis and data mining because of the importance of educational data mining. As a result, more surveys for each of the EDM applications should be undertaken to more precisely identify the most effective strategies. Furthermore, in the evaluation, broader evaluation and comparison criteria should be applied.

[2] investigated approaches for predicting first year computer science undergraduate students' academic achievement in 2015. They applied various methodologies to student data, such as Naive Bayes, Rule-Based, and Decision Tree, to build the best prediction model for students' academic achievement. The findings of this investigation revealed that the Rule Based model is the best prediction model, with a prediction accuracy of 71.3 percent.

A study[14] looked at how different classification algorithms could be used to predict slow learners among students. Data from 152 high school students has been collected, and analyzed using WEKA tool to examine and test the students' performance. The Multilayer Perception technique achieved the best prediction accuracy of 75%.

In 2016, [12] published a study that employed data mining techniques like Nave Bayes, Decision Trees, and Neural Networks to predict student success in two courses. The Naive Bayes method, with an accuracy of 86 percent, was followed by the Decision Tree and Neural Network algorithms, with 82.7 percent and 79.2 percent, respectively.

In their work [1], With the help of KNN and Naïve Bayes algorithms, the gathered information from 500 secondary school pupils who had eight characteristics. According to their findings, the Naive Bayes classifier had 94.1 percent accuracy, whereas the KNN classifier had a prediction accuracy of 62.9 percent.

[4] used the WEKA tool to apply 5 main characteristics approaches of data mining for the prediction of the performance of students in another study. He gathered information from 225 university students.

In their work [13,11] used three data mining classification approaches to predict student performance using academic attributes: Naive Bayes, Decision Tree, and K-Nearest Neighbor (K-NN). They gathered information from 1100 male students at a preparatory school in the Gaza Strip.

As per [8], There are mainly 6 data mining techniques available such as Naive Bayes, Random Forest, Decision Tree, IBk, K-star, and Naive Bayes Multiple Nominal, to predictthe marks of students. They utilized the WEKA tool to collect a dataset of 480 records with 16 attributes. Random Forest has the highest prediction accuracy of 76.667 percent, according to their findings.

In addition, [28] used Bayesian Network and Decision Tree approaches to forecast student performance in order to prevent students from failing. They gathered information from students at Islam Indonesia University studying Industrial Engineering.

VII. COMPARATIVE STUDY AND DISCUSSION

Table 1. show the performance prediction of students and accuracy achieved by the different researchers using data mining classification techniques.

Reference	Year	Technique	Dataset	Accuracy	Tool
[35]DorinaKabakchieva	2012	OneR,J48,NN,K-NN	University ofnationalandworld economy Sofia,Bulgaria	NN(73.5904%)	WEKA
[36] SaurabhPaletal.	2013	OneR,J48,MLP,IB1	Degree colleges sndInstitutionsaffiliatedwithDr. R.M.LAwadh University Faizabad,India	IB1(82.00%)	WEKA
[37] Dr.SangeetaGuptaetal.	2014	J48,RF	MCAstudentsfromvarious Institutionsaffiliated to GuruGobindSingh Indraprastha	RF(94.4186%)	WEKA
[2]Ahmad etal.	2015	NB,RB,DT	UCI MachineLearningRepository	RB(71.3%)	WEKA
[14]Parneet Kauretal.	2015	MP,NB,SMO,J48,REPTree	Highschool(real worlddataset)	MP(75%)	WEKA
[12]Mueenetal.	2016	NB,DT, NN	Collegestudents(Undergraduatestudents)	NB(86%)	WEKA
[8]Kapuret al.	2017	NB,RF,DT,IBk,Kstar, NBMN	UCIMachine LearningRepository	RF(76.667%)	WEKA
[1]Amrsetal.	2017	KNN,NB	Secondary Schoolsfrom the Ministryofeducationinthe Gazastrip	NB(94.1%)	WEKA
[4]Almarsbehetal.	2017	NB,BN,ID3,J48,NN	Student'scollegedatabase	BN(92%)	WEKA
[38] VikasChaurasiaetal.	2017	SMO,Bagging,REPTREE,DT	MCAdepartmentofVBSPurvanchal UniversityJaipur, India	Bagging(80.253)	WEKA

[11]HafezMoussaetal.	2017	NB,DT,KNN	Student datasetfrom PreparatorymaleschoolinGaza strip	DT(92.96%)	WEKA
[39]AnchalGarg etal.	2017	NB,MLP,J48	UCI Machinelearningrepository	J48(73.9241%)	WEKA
[40]Kaunangetal.	2018	DT,RF	Students ofComputer ScienceDepartments UniversityKabat	DT(66.85%)	WEKA
[41]Ketui etal.	2019	DT,ID3,RT,GBT,DTW	AgriculturalTechnologyRajmangalaUniversity ofTechnologyLanna Nan	GBT(92.41%)	RapidMiner
[42]MukeshKumareal.	2019	NB,J48,RF,REPTree, jRip, OneR, SL,ZeroR	Secondary schoolPortugues PereirasandMusinhoda Silveria	OneR(76.7334%)	WEKA

Table 1: Comparison of Data mining techniques used on student datasets in educational settings

VIII. CONCLUSIONS AND FUTURE SCOPE

Conclusion

As the use of technology increases in the system of education, a large amount of data generates every day. The sector of data mining in the educational field is now expanding and has many benefits such as integration with modern algorithms and methods developed in several areas of machine learning and data mining. EDM can also be used to find out the risk in students, enhance rates of graduation, prioritize requirements of learning for several groups, thoroughly investigate the performance of the institution and optimize curriculum of topic renewal, and many more things.We have observed from the related review work that the Random Forest algorithm has come up with the highest accuracy which is 94. 4186% that has been used in Mining Students' Data for Prediction Performance using WEKA Tool.

Future Scope

In the future, we will compare the performance of other classification and algorithms on educational datasets for predicting student's behavior, Grades using data mining tools, and will also find the best classification algorithm for a particular dataset having minimum errors and maximum accuracy rate. Also, we can use other Data mining tools for comparing data mining tools performance on different datasets as less research work is done on tools other than WEKA.

REFERENCES

[1] A. Amra and A. Y. Maghari, "Students performance prediction using knn and na¨ıve bayesian," in 2017 8th International Conference on Information Technology (ICIT), pp. 909–913, IEEE, 2017.

[2] Ahmad, N. H. Ismail, and A. A. Aziz, "The prediction of students' academic performance using classification data mining techniques," Applied Mathematical Sciences, vol. 9, no. 129, pp. 6415–6426, 2015.

[3] Algarni, "Data mining in education," International Journal of Advanced Computer Science and Applications, vol. 7, no. 6, pp. 456– 461, 2016.

[4] Almarabeh, "Analysis of students' performance by using different data mining classifiers," International Journal of Modern Education and Computer Science, vol. 9, no. 8, p. 9, 2017.

[5] Bakhshinategh, O. R. Zaiane, S. ElAtia, and D. Ipperciel, "Educational data mining applications and tasks: A survey of the last 10 years," Education and Information Technologies, vol. 23, no. 1, pp. 537–553, 2018.

[6] Jeong and G. Biswas, "Mining student behavior models in learning-by-teaching environments.," in EDM, pp. 127–136, Citeseer, 2008.

[7] Jiawei, H., & Kamber, M. (2001). Data mining: concepts and techniques. San Francisco, CA, Itd: Morgan Kaufmann, 377–385. https://doi.org/10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C.

[8] Kapur, N. Ahluwalia, and R. Sathyaraj, "Comparative study on marks prediction using data mining and classification algorithms," International Journal of Advanced Research in Computer Science, vol. 8, no. 3, 2017.

[9] Maimon, O., & Rokach, L. (2005). Data Mining and Knowledge Discovery Handbook. Data Mining and Knowledge Discovery Handbook. https://doi.org/10.1007/0-387-25465-x_2. f

[10] Mostow and J. Beck, "Some useful tactics to modify, map and minedata from intelligent tutors," Natural Language Engineering, vol. 12,no. 02, pp. 195–208, 2006.

[11] Mousa and A. Y. Maghari, "School students' performance predication using data mining classification," School Students' Performance Predication Using Data Mining Classification, vol. 6, no. 8, 2017.

[12] Mueen, B. Zafar, and U. Manzoor, "Modeling and predicting students' academic performance using data mining techniques," International Journal of Modern Education and Computer Science, vol. 8, no. 11, p. 36, 2016.

[13] Muehlenbrock, "Automatic action analysis in an interactive learningenvironment," in Proceedings of the 12 th International Conference onArtificial Intelligence in Education, pp. 73–80, 2005.

[14] P. Kaur, M. Singh, and G. S. Josan, "Classification and prediction based data mining algorithms to predict slow learners in education sector," Procedia Computer Science, vol. 57, pp. 500–508, 2015.

[15] P. Reyes and P. Tchounikine, "Mining learning groups' activities inforum-type tools," in Proceedings of th 2005 conference on Computersupport for collaborative

learning: learning 2005: the next 10 years!,pp. 509–513, International Society of the Learning Sciences, 2005.

[16] R, B., Sánchez-Guzmán A, & García, R. (2013). Minería de datos educativa: Una herramienta para la investigación de patroness de aprendizaje sobre un contexto educativo. (Spanish). LatinAmerican Journal of Physics Education, 7(4), p662-6.

[17] R. Baker et al., "Data mining for education," International encyclopediaof education, vol. 7, pp. 112–118, 2010.

[18] Romero and S. Ventura, "Educational data mining: a review of thestate of the art," Systems, Man, and Cybernetics, Part C: Applicationsand Reviews, IEEE Transactions on, vol. 40, no. 6, pp. 601–618, 2010.

[19] Romero, C., & Ventura, S. (2013). Data mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3(1), 12–27. https://doi.org/10.1002/widm.107.

[20] S. Ryan and Paul S.I , Educational Data Mining and Learning Analytics, DOI:10.1007/978-1-4614-3305-7_4[online]Available at: Semantic Scholar site.

[21] Romero, S. Ventura, and P. De Bra, "Knowledge discovery withgenetic programming for providing feedback to courseware authors,"User Modeling and User-Adapted Interaction, vol. 14, no. 5, pp. 425–464, 2004.

[22] Romero, S. Ventura, P. G. Espejo, and C. Herv´as, "Data miningalgorithms to classify students.," in EDM, pp. 8–17, 2008.

[23] S. L. Tanimoto, "Improving the prospects for educational data mining,"in Track on Educational Data Mining, at the Workshop on Data Miningfor User Modeling, at the 11th International Conference on UserModeling, pp. 1–6, 2007.

[24] S. Raghavan, "Data mining in e-commerce: A survey," Sadhana,vol. 30, no. 2-3, pp. 275–289, 2005.

[25] S. Ray and M. Saeed, "Applications of educational data mining and learning analytics tools in handling big data in higher education," in Applications of Big Data Analytics, pp. 135–160, Springer, 2018.

[26] Siemens, R. S. j. d. B., & G. Siemens, R. S. j. d. B. (2012). Learning analytics and educational data mining: towards communication and collaboration. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge.

[27] Society, I. (2014). Informe Global de Internet 2014, 1.

[28] Tang, R. W. Lau, Q. Li, H. Yin, T. Li, and D. Kilis, "Personalizedcourseware construction based on web data mining," in Web Informa-tion Systems Engineering, 2000. Proceedings of the First InternationalConference on, vol. 2, pp. 204–211, IEEE, 2000.

[29] U. Khasanah et al., "A comparative study to predict student's performance using educational data mining techniques," in IOP Conference Series: Materials Science and Engineering, vol. 215, p. 012036, IOP Publishing, 2017.

[30] V. Manjarres, L. G. M. Sandoval, and M. S. Suarez, "Data mining ´ techniques applied in educational environments: Literature review," Digital Education Review, pp. 235–266, 2018.

[31] Virseda Benito, F., & Carrillo, J. (2008). Minería de datos y aplicaciones (p. 8).

[32] Y. Salal, S. Abdullaev, and M. Kumar, "Educational data mining: Student performance prediction in academic," IJ of Engineering and Advanced Tech, vol. 8, no. 4C, pp. 54–59, 2019.

[33] Elaf Abu Amrieh, Thair Hamtini, and Ibrahim Aljarah, "Mining Educational Data to Predict Students's academic Performance using Ensemble Methods", International Journal of Database Theory and Application, Volumr-9,no-8 pp-119-136.

[34] Kalpesh P., Riya A. Sharma, Shreya, Rajeshwari J.B(2017), Student Performance Prediction System using DM Approach, International Journal of Advanced Research in Computer and Communication Engineering Volume-6, Issue-3,March.

[35] Dorina Kabakchieva, "Student Performance Prediction by Using Data Mining Classification Algorithms", International Journal of Computer Science and Management Research, Vol 1 Issue 4 November 2012 ISSN 2278-733X.

[36] Ajay Kumar Pal, Saurabh Pal "Data Mining Techniques in EDM for Predicting the Performance of Students", International Journal of Computer and Information Technology (ISSN: 2279 – 0764), Volume 02– Issue 06, November 2013.

[37] Sangeeta Gupta, Dharminder Kumar, "Mining Students' Data for Prediction Performance", Conference Paper · February 2014 DOI: 10.1109/ACCT.2014.105.

[38] Saurabh Pal, Vikas Chaurasia, "PERFORMANCE ANALYSIS OF STUDENTS CONSUMING ALCOHOL USING DATA MINING TECHNIQUES", International Journal of Advance Research in Science and Engineering, vol. no. 6, issue no. 2, February 2017.

[39] Anchal Garg, Sagardeep Roy, "Predicting Academic Performance of Student Using Classification Techniques ",2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON) GLA University, Mathura, Oct 26-28, 2017.

[40] Reymon Rotikan, Fergie Joanda Kaunang, "Students' Academic Performance Prediction using Data Mining".

[41] Nongnuch Ketui, Warawut Wisomka, Kanitha Homjun "Using Classification Data Mining Techniques for Students Performance Prediction" 978-1-5386-8072-8/19/$31.00 ©2019 IEEE.

[42] Mukesh Kumar, Abdoulaev S., "Educational Data Mining: Student Performance Prediction in Academic", See discussions, stats, and author profiles for this publication at.