Impact Of Data Mining In The Era Of Electronic Data Management In Digital Examination System With Support For Knowledge-Driven Decisions

Dr. Karthik K R[1]

Assistant Professor, Management Studies, Visvesvaraya Technological University- Belagavi, Karnataka, India, vtulakshmi@gmail.com

Dr. Kamal Gulati[2]

Associate Professor, Amity University, Noida, Uttar Pradesh, India, drkamalgulati@gmail.com

Sunil Bahiru Ghane[3]

Assistant Professor, Sardar Patel Insitute of Technology, Andheri, Mumbai, Maharashtra, India, sunil_ghane@spit.ac.in

Dr. Vishal Ratansing Patil[4]

Assistant Professor, Computer Department, Sandip University, Nashik, Maharashtra, India, vp0106@gmail.com

Dr. Neeta Bhatt[5]

Associate Professor, IT, Mumbai University, Maharashtra, India, neeta.b@aimsr.edu.in

Gargi Verma[6]

University Institute of Liberal Arts and Humanities, Chandigarh University, gargi.e8825@cumail.in

ABSTRACT

Due to the availability of a vast amount of educational information that can be obtained from various sources, EDM has attracted a significant amount of attention from scholars over the past few decades. The primary purpose of EDM is to make DM models more effectively in order to protect the vast quantities of educational information and to develop an environment that is conducive to the safety of the student's educational pursuits. Within the framework of this strategy, a variety of models have been implemented for DM and its analytics. In addition, prediction models such as classification, regression, and latent factor evaluation technologies were utilised in the analysis process.

Keywords: Data Mining, KDD, Artificial Neural Networks, Sequential Pattern, Modeling building

INTRODUCTION

Machine learning (ML) is a cutting-edge research area in artificial intelligence. Machine learning has a vast area of applications, from managing search engines to safeguarding passwords. Machine learning is the study of machines learning from human behavior to solve simple to complex problems. Machine learning algorithms are revolutionizing real-world applications and research paths. Machine learning makes systems more intelligent and automated when dealing with high-dimensional and complex data. ML excels at several problems, including knowledge acquisition, learning, prediction, and problem solving. In real time, ML saves every human's time and effort. From many perspectives, it has become unavoidable for everyone to make their everyday lives run more smoothly. Offline lessons were not feasible due to widespread virus worries. The health and safety of pupils was a top focus. Educational institutions are forced to cancel actual classroom learning sessions, and delivering tests is problematic. E-learning, remote learning, and online education have all emerged as viable methods for learning education.

It was thought that the examinations or assessments needed to be done online. The online education mode came to help students finish their classes without taking a break from academia. During a lockdown scenario in academia, the exam management system is the most affected. Examinations were a big part of ML's role in getting students to the end of their schooling. Any analysis from an academic point of view needs to start with the student data. The research and data mining on each student's classes help the educational infrastructure build a support system for students and teachers to get ready for final assessments.

The student learning management data aid in the development of a decision-making model for educational prediction or analysis. The construction and process of these learning and assessment models rely heavily on machine learning technologies. The sample student data, on the other hand, should be adequate and informative. The use of machine learning models in prediction analysis to assess student performance based on data is advantageous. This feature of ML's futuristic prediction helps students improve their performance by allowing for early intervention in learning and student achievement.

EDUCATIONAL DATA MINING (EDM)

Due to the availability of a vast amount of educational information that can be obtained from a variety of sources, EDM has attracted a significant amount of attention from scholars over the course of the past few decades. The primary purpose of EDM [9] is to make DM models more effectively in order to protect the vast quantities of educational information and to develop an environment that is conducive to the safety of the student's educational pursuits. Within the

framework of this strategy, a variety of models have been implemented for DM and its analytics. In addition, prediction models such as classification, regression, and latent factor evaluation technologies were utilised in the analysis process.

APPLICATIONS OF EDM IN HIGHER EDUCATION

Student modelling is defined as a method of representing the cognitive elements of student actions such as examining student performance and behaviour, differentiating the fundamental misconception, showing student aims and achievements, discovering the advanced as well as acquired knowledge, retaining episodic memory, as well as defining personality characteristics. The definition of classifying EDM applications is taken into account by this method. Each section of this course demonstrates a strategy for defining the objectives and targets of the students. According to the research that was done in[16] student modelling can include a variety of characteristics, including knowledge and skill (1), error and misinterpretation (2), learning style with priority (3), affective and cognitive components (4), and metacognitive factor (5). (5). EDM's taxonomy of application structures and hierarchies (Figure 1).
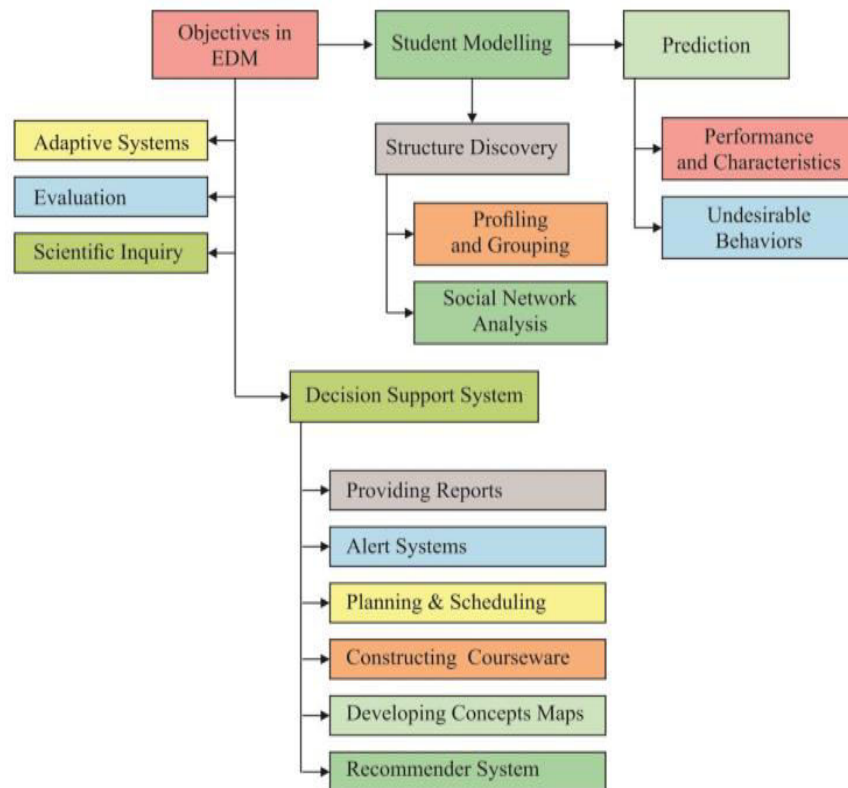

Figure 1: Classification of different EDM Applications

LITERATURE REVIEW

A literature review is a method of examining and comprehending current material that has been proposed. The fundamental goal of a literature review is to determine how well a problem has been solved and what potential alternatives for improvement exist. Every real-life problem is always resolved in stages. A researcher's solution is used as a foundation for improving it by decreasing or eliminating its restrictions. Similarly, in data mining-based research, it is necessary to apply current algorithms and data schemes in order to improve the efficiency of our system. Limitations can also be attributed to the limitations of the algorithms we've utilised. Limitations can sometimes be attributed to the data that we have used. In both circumstances, a good survey is essential before upgrading any or all of these types of constraints to determine whether such changes have already been made by someone else. There is no point in repeating research that has already been done.

Jovial Niyogisubizo et al. [2022] An ensemble of Random Forest, XGBoost, Gradient Boosting, and Feed-forward Neural Networks (FNN) is proposed to predict student dropout from university programmes. The suggested strategy outperforms the base models in terms of testing accuracy and the area under the curve (AUC) in comparison to the dataset gathered from 2016 to 2020 at Constantine the Philosopher University in Nitra. In light of the study's findings, educators and other educational stakeholders can identify students at risk of dropping out based on influential factors and use this information for early intervention in the uncontrolled behaviour that could lead to dropout and proactive precautionary measures before the issue arises [1]

Roberto Bertolini et al. [2022] For 3225 students enrolled in an undergraduate biology class, bootstrapping was used to investigate performance variability among five data mining methods (DMMs) and four filter preprocessing feature selection procedures for anticipating course grades. While the AUC values produced by bootstrapping were much lower than the AUC point estimates obtained without resampling, DMMs and feature selection strategies had varied effects on variability. The elastic net regression (GLMNET) ensemble technique greatly outperformed all other DMMs and had the least amount of fluctuation in the AUC. When this phase was removed from the data pipeline, however, all filter feature selection strategies greatly increased variability in student success forecasts. We address the advantages and disadvantages of introducing bootstrapping into prediction pipelines for tracking, monitoring, and forecasting classroom performance, as well as the dangers of merely looking at point estimates [2].

Michal Berkowitz et al. [2022] N = 317 students were studied to see how working memory (WM) and mathematics performance are linked in mathematics-intensive undergraduate STEM programmes. We asked two questions to mechanical engineering and math-physics students: (1) Does the relationship between verbal and visuospatial WM and three measures of mathematics performance differ: numerical reasoning ability, prior knowledge in mathematics, and successes

in mathematics-intensive courses? (2) To what extent are numerical reasoning abilities and prior knowledge in mathematics mediated by the effects of WM on achievement in mathematics-intensive courses? Verbal WM was found to be at least as significantly linked with the three mathematics measures as visuospatial WM in a latent correlational analysis. Numerical reasoning fully mediated the effects of WM on achievement in math-intensive courses, both directly and in a doubly mediated impact via prior knowledge in mathematics, according to a latent mediation model. We conclude that WM contributes significantly to the mathematics performance of mathematically competent pupils across modalities. The influence of verbal WM appears to be stronger than previously thought in the literature [3].

Yuzhi Lai et al. [2022] examined whether intention and actual use were mediated by self-regulation skills The survey data of 676 language learners from Chinese universities from various disciplines was collected and analysed using structural equation modelling. According to the findings, 37.1 percent of respondents never utilise mobile devices for self-directed language study. The bulk of the other 425 respondents who said they used mobile technology for this reason appeared to be extrinsically motivated. Learning activities involving vocabulary acquisition and translation were reported considerably more frequently than learning activities involving listening, speaking, reading, and writing. Furthermore, students' intention to utilise mobile technology was significantly described by attitude and subjective norm, but self-efficacy had no direct effect on students' intention. Furthermore, students' self-control abilities and intentions were linked to their actual use of mobile devices. The results of moderation analysis revealed that any development in self-regulation skills will strengthen the link between intention and actual conduct. The findings are discussed, and conclusions are drawn.[4].

Carlos Felipe et al. [2022] Artificial intelligence applications in education have grown in popularity in recent years. To further the systematic implementation of these ideas, however, more conceptual and methodological understanding is required. The study's initial goal is to put to the test a method for using artificial neural networks to predict academic achievement in higher education. The second goal is to investigate the significance of a number of well-known determinants of academic performance in higher education. A total of 162,030 students from both private and governmental universities in Colombia were included in the study. The findings imply that artificial neural networks can be used to categorise students' academic achievement as either high (accuracy of 82%) or low (accuracy of 82%). (accuracy of 71 percent). In evaluation measures like as recall and F1 score, artificial neural networks beat other machine-learning methods. Prior academic accomplishment, socioeconomic status, and high school features are also revealed to be significant determinants of students' academic performance in higher education. Finally, this paper addresses artificial neural network implementation guidelines as well as numerous factors to consider when analysing academic achievement in higher education.[5]

Ahajjam Tarik et al. [2021] enhances user experiences in a number of ways We discuss an intelligent solution for predicting the performance of a system in this study. Artificial intelligence relies on algorithms to allow machines to make judgments in place of humans. This technology improves user experiences in a variety of areas for Moroccan students in the GuelmimOued Noun region using a recommendation system based on artificial intelligence. During the COVID-19, we explore an intelligent approach for predicting the performance of Moroccan pupils in the GuelmimOued Noun region using a recommendation system based on artificial intelligence techniques. during COVID-19 methods.[6]

Shinichi Oeda et al.[2021] suggested a convex factorization machine-based strategy for predicting student performance. Traditional factorization machines, which predict student performance using factorization machines, produced better results than those that combine the benefits of support vector machines and factorization models like matrix factorization. The previous outcomes. Factorization machines, on the other hand, are not convex optimizations and hence acquire local minima, which is a drawback. Traditional methods that use factorization machines to predict student performance produced better outcomes than before. Factorization machines, on the other hand, are not convex optimizations and hence acquire local minima, which is a drawback. In order to increase the performance of student modelling predictions, we deployed convex factorization machines. In order to increase the performance of student modelling predictions, we deployed convex factorization machines.[7]

Shaikh Rezwan Rahman et al. [2021] studied "Effects of Co-Curricular Activities on Student's Academic Performance Through Machine Learning" is a study that investigates the impact of co-curricular activities on a student's academic performance. The goal of this research is to figure out how extracurricular activities affect student achievement. Extracurricular activities that support and enrich the academic or core curriculum are known as co-curricular activities. They are an important part of educational institutions' efforts to help students develop their personalities and learn better in the classroom. However, a large percentage of Bangladeshi students do not participate in such activities. One of the main reasons is that many people assume these activities will negatively impact a student's academic achievement. The goal of this study is to determine the true impact of co-curricular activities on students. Using Logistic Regression and Google Colab, it was discovered that there is a favourable link between co-curricular activities and academic success.[8]

FatemaAlnassar et al. [2021] Predicting student performance using a variety of relevant data has shown to be an effective technique in educational institutions for enhancing curriculum and teaching methods. Automated educational data analysis employing cutting-edge Machine Learning (ML) or Artificial Intelligence (AI) algorithms is a hot topic in academia. On the Open University (OU) dataset, the research uses three machine learning methods (SVC, k-Nearest

Neighbour (k-NN), and Artificial Neural Network (ANN)) to solve the problem of student performance prediction. Three primary indicators are examined in educational data: demographics, engagement, and performance. When comparing applied and existing research, the k-NN strategy appeared as the best for OU experiments from the experimental study. Changes in how missing values are handled and data standardisation procedures are credited with improved results.[9]

V.Uday Kumar et al. [2020] Machine learning is a branch of computer science that allows a computer to learn without the assistance of external applications. These machine learning approaches can be used to forecast the outcome of specific inputs. Machine learning techniques can be divided into two categories: supervised learning and unsupervised learning. K-means and Hierarchical clustering are used in unsupervised learning, while Naive Bayes and Decision Trees are used in supervised learning. In today's world, analysing a company's student performance is critical to training its employees. All of the above algorithms were merged and used in the recruitment process to evaluate students. All of the following algorithms are used to evaluate the performance of K L University students.[10]

Sana Bhutto et al. [2020] Introduces a model for predicting students' academic achievement that employs supervised machine learning algorithms such as support vector machine and logistic regression. The results of several trials utilising various technologies are compared, and it is discovered that the sequential minimal optimization technique exceeds logistic regression in terms of accuracy. And the knowledge gained from this study can assist educational institutions in predicting students' future conduct and categorising their performance as excellent or bad. The goal is to not only predict future student performance, but also to provide the best technique for identifying the most significant elements to work on, such as instructor performance and student motivation, in order to reduce the student dropout rate.[11]

Raghad Alshabandar et al. [2020] Massive open online courses (MOOCs) used in remote learning environments have grown in popularity thanks to advances in information and communication technology (ICT). Various tools, such as drawings, figures, and videos, have been used to convey interactive content that can drive learners to develop new cognitive skills. MOOCs have been utilised by top colleges as an effective dashboard platform where learners from all over the world can participate in such courses. Set computer-marked assessments are used to assess the pupils' learning progress. The computer, in particular, provides quick feedback to the learner when he or she has completed the online assessments. According to the study, student achievement in an online course is linked to their prior session performance as well as their level of interest. Literature has paid little attention to determining whether student performance and involvement in previous tests may influence student accomplishment in subsequent examinations. Two prediction models, namely student evaluation grades and final

student performance, have been developed in this research. The models can be used to identify the elements that influence MOOC students' learning outcomes. The outcome demonstrates that both models produce practical and accurate outcomes. The lowest RSME gain was produced by RF, with an average value of 8.131 for student assessments grades model, while GBM yielded the highest accuracy in final student performance, with an average value of 0.086.[12]

CONCLUSION

Exams and assessments are crucial in a student's life since they decide their future and employment opportunities. During these trying times, technological advances came to the rescue, allowing students to continue their studies uninterrupted. Machine learning is required for schools and universities to migrate from real-time to online mode. During the lockdown, machine learning algorithms facilitated online learning and examination. This study provides a thorough examination of the role of machine learning in lockdown. Administration of Exams The significance of machine learning was reviewed and argued throughout the test cycle, from pre-exam preparation to exam examination and evaluation. The unsupervised and supervised Machine learning algorithms were identified and categorised in each operation. Machine learning perspectives are used to investigate main test components such as authentication, scheduling, proctoring, and cheat or fraud detection.

REFERENCES

[1] J. Niyogisubizo, L. Liao, E. Nziyumva, E. Murwanashyaka, and P. C. Nshimyumukiza, "Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization," *Comput. Educ. Artif. Intell.*, vol. 3, no. November 2021, p. 100066, 2022, doi: 10.1016/j.caeai.2022.100066.

[2] S. L. Priestnall *et al.*, "No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title," *Endocrine*, vol. 9, no. May, p. 6, 2020.

[3] M. Berkowitz, P. Edelsbrunner, and E. Stern, "The relation between working memory and mathematics performance among students in math-intensive STEM programs," *Intelligence*, vol. 92, no. April 2021, p. 101649, 2022, doi: 10.1016/j.intell.2022.101649.

[4] Y. Lai, N. Saab, and W. Admiraal, "University students' use of mobile technology in self-directed language learning: Using the integrative model of behavior prediction," *Comput. Educ.*, vol. 179, no. November 2021, p. 104413, 2022, doi: 10.1016/j.compedu.2021.104413.

[5] C. F. Rodríguez-Hernández, M. Musso, E. Kyndt, and E. Cascallar, "Artificial neural

networks in academic performance prediction: Systematic implementation and predictor evaluation," *Comput. Educ. Artif. Intell.*, vol. 2, no. March, 2021, doi: 10.1016/j.caeai.2021.100018.

[6]   A. Tarik, H. Aissa, and F. Yousef, "Artificial intelligence and machine learning to predict student performance during the COVID-19," *Procedia Comput. Sci.*, vol. 184, pp. 835–840, 2021, doi: 10.1016/j.procs.2021.03.104.

[7]   S. Oeda and D. Shimizu, "Verification of usefulness of student modeling with real educational data using convex factorization machines," *Procedia Comput. Sci.*, vol. 192, pp. 804–811, 2021, doi: 10.1016/j.procs.2021.08.083.

[8] S. R. Rahman, M. A. Islam, P. P. Akash, M. Parvin, N. N. Moon, and F. N. Nur, "Effects of co-curricular activities on student's academic performance by machine learning," *Curr. Res. Behav. Sci.*, vol. 2, no. September, p. 100057, 2021, doi: 10.1016/j.crbeha.2021.100057.

[9] F. Alnassar, T. Blackwell, E. Homayounvala, and M. Yee-King, "How Well a Student Performed? A Machine Learning Approach to Classify Students' Performance on Virtual Learning Environment," *Proc. 2021 2nd Int. Conf. Intell. Eng. Manag. ICIEM 2021*, pp. 1–6, 2021, doi: 10.1109/ICIEM51511.2021.9445286.

[10] V. U. Kumar and A. Krishna, "Proceedings of the International Conference on Electronics and Sustainable Communication Systems, ICESC 2020," *Proc. Int. Conf. Electron. Sustain. Commun. Syst. ICESC 2020*, no. Icesc, pp. 121–126, 2020.

[11] D. I. F. S. Engr. Sana Bhutto, Dr. Qasim Ali Arain, Maleeha Anwar, "Through Supervised Machine Learning," *Predict. Students' Acad. Perform. Through Supervised Mach. Learn.*, 2020.

[12] R. Alshabandar, A. Hussain, R. Keight, and W. Khan, "Students Performance Prediction in Online Courses Using Machine Learning Algorithms," *Proc. Int. Jt. Conf. Neural Networks*, 2020, doi: 10.1109/IJCNN48605.2020.9207196.