

# A Review Analysis Data Mining Techniques to Predict Students Performance

Vinod Kumar Rathaur<sup>1</sup>, Ankur Taneja<sup>2</sup>

<sup>1</sup>Mtech Scholar, vinodrathaur1@gmail.com, SAMCET(RGPV) Bhopal, India

<sup>2</sup>Assistant Prof., SAMCET(RGPV) Bhopal, India

---

**Abstract** – Education has evolved in this era of computerization and is no longer limited to the old lecture style. The constant search for innovative ways to make it more effective and efficient for students is ongoing. A lot of data is collected in educational databases nowadays, yet it is rarely used. Powerful tools are required to reap the benefits of such large amounts of data. Data mining is a strong tool for analysis and prediction that is just getting started. It has been used successfully in the detection of fraud, advertising, marketing, loan assessment, and prediction. However, in the realm of education, it is still in its infancy. Although a lot of work has been done in this field, there are still a lot of unexplored territories.

**Keywords:** Student performance, educational data mining, performance prediction

---

## I. Introduction

Throughout the study life of students, performance is one of the most critical aspects of their success.

This condition is because usually, performance is an essential part of learning from junior school to higher institutions of learning. Providing quality education to students requires continuous monitoring of learning and teaching activities in an education environment. However, this tends to be difficult because of large amounts of data in educational databases. Nevertheless, the new technology of big data and machine learning comes in to solve all these problems. Big data is a set of methods and techniques that necessitate a new form of combinations to find large hidden values from various complex datasets on a considerable scale. This technology is very advantageous in a way that it can determine the patterns obtained from data analysis to understand hidden information and ease decision making [1], [2] Currently, many methods have been proposed by researchers to evaluate the performance of students in the education field. However, one of the most recently popular proposed and applied procedures to predict the performance of students is data mining. This method has several advantages; for example, data mining techniques can be used in e-learning since it is easy to access students' log data, and automatic data analysis can be done [3]. In big data and machine learning technology, the application of data mining in the education domain is known as 'education data mining' (EDM). This aims at converting education data into very useful information which brings a positive impact to the education sector and promotes research practices such as predicting student performance [4]. Performance prediction can help strengthen the education sector and assist

responsible leaders in providing practical teaching approaches, thus encouraging higher achievements since student performance is monitored right from the early stages of education. Therefore, students are also equipped with the understanding and improving their learning activities to avoid risks such as school dropouts, poor performance, and failure to graduate. To ease the process of students' performance prediction, we propose a systematic review to help us achieve this study's objectives that are;

1. To identify the commonly used methods to predict student performance.
2. To find out the most suitable method for student performance prediction among the commonly used methods.

## II. LITERATURE SURVEY

### 2.1. General applications of data mining

Analyzing large amounts of data is a necessity. Presently, data mining continues to achieve significant success in many areas of human activities such as medicine, business, and robotics. It is, therefore, beneficial to all Institutions and enterprises of all kinds to collect and analyze their data. With proper analysis, administrators in health sectors can learn trends and identify issues that come up in the patients' records, find ways of managing the hospital's human resources better and enhance the performance of hospital staff [5], [6]. In addition to that, nowadays, the technology of big data gives a chance to professional health workers to screen and predict mental illness [7], derive essential health trends, and support timely preventive care [8]. With the significant increase of data in the agricultural domain,

data mining is the only capable technology employed to transform agriculture into smart agriculture [9],[10],[11].

Data mining technology can also be employed to monitor energy consumption of household appliances thorough monitoring of electricity consumption patterns (ECPs) and their relationship with household features, thereby economizing the electricity consumption in residential areas [12], [13]. The technology of machine learning and data mining can be applied in cyber-manufacturing systems to detect and effectively manage cyber security and computer network intrusions by using big data tools [14], [15]. In addition to the above, the technique of data mining can be useful for financial firms to correct their material incase of an inaccuracy. Applying algorithms such as SVM, naive bayes, and ANN on financial data gives an easy way of revising past published financial statements [16], [17]. Big data applications can be also used to spot trends in blogs and understand the similarities and differences in social media messages to make an easy way of operation for online businesses to operate [18].

## 2.2. Data mining in education

Several researchers have been successful in doing their researches about data mining applications in education, and this can be witnessed. However, among these aspects, predicting students' performance tends to be one of the most important and useful concepts in educational data mining. Student performance prediction involves estimating an unknown value, which is the student's score or mark [34]. This review, therefore, bases itself on student performance prediction as a crucial activity involved in educational data mining. The next section shows the research methodology, research questionnaire criteria, and research strategy, and then continues the survey to determine the results that will reach the objectives of the study as well as answer the research questions to make the research study valid.

Amirah Mohamed Shahiria et. al. Predicting students performance becomes more challenging due to the large volume of data in educational databases.

Currently in Malaysia, the lack of existing system to analyze and monitor the student progress and performance is not being addressed. There are two main reasons of why this is happening. First, the study on existing prediction methods is still insufficient to identify the most suitable methods for predicting the performance of students in Malaysian institutions. Second is due to the lack of investigations on the factors

affecting students achievements in particular courses within Malaysian context. Therefore, a systematical literature review on predicting student performance by using data mining techniques is proposed to improve students achievements. The main objective of this paper is to provide an overview on the data mining techniques that have been used to predict students performance. Author also focuses on how the prediction algorithm can be used to identify the most important attributes in a students data.

Author could actually improve students achievement and success more effectively in an efficient way using educational data mining techniques. It could bring the benefits and impacts to students, educators and academic institutions.

Pooja Thakar et. al. One of the most recent and biggest challenge that higher education faces today is making students skillfully employable. Many universities/institutes are not in position to guide their students because of lack of information and assistance from their teaching-learning systems. To better administer and serve student population, the universities/institutions need better assessment, analysis, and prediction tools.

Considerable amount of work is done in analyzing and predicting academic performance, but all of these works are segregated. There is a clear need for unified approach. Other than academic attributes, there are large numbers of factors that play significant role in prediction, which includes non-cognitive factors (set of behaviors, skills, attitudes). Suitable data mining techniques are required to measure, monitor and infer these factors for prediction. Thus enriching the input vector with qualitative values may increase the accuracy rate of prediction as well.

## III. METHOD

This paper uses a 'systematic review' protocol to summarize existing information about predicting student performance clearly and systematically. Systematic reviews involve three phases, i.e., planning, conducting, and reporting the analysis .

This systematic review aims to support research questions, fulfill the gaps in previous related studies, and find out the most suitable and trusted method for predicting student performance while placing a new research work in the appropriate setting. Next, research questions were identified to lead to the intended results of the study.

### The research questions

For any researcher to carry out a systematic review, it is essential to include research questions as they are always helpful in identifying the scope of the study. To formulate research questions for this study, the researchers followed the steps suggested by Kitchenham to structure the right research questions. Kitchenham recommends considering questions from three viewpoints; that is, population, intervention, and outcomes. In this study; (1) population used was university students (2) intervention ie; data mining methods for student performance prediction (3) Outcomes measures and results ie; trusted prediction method, accuracy of performance prediction, and reduced student failure to graduate.

The research questions formulated for this study were;

Q1: What are the most commonly used methods in students' performance prediction?

Q3: Which one is the most suitable method for students' performance prediction among the commonly used methods?

### **The search strategy**

In this study, the search strategy was used to search for related studies, using search terms and resources to be searched. This systematic review used search terms suggested and developed by Kitchenham in generating a search strategy [35]. The search strings obtained for this study were as follows: (students' performance) AND (techniques OR methods OR applications OR systems OR processes OR methodologies OR procedures) AND (educational data mining) AND (prediction OR estimation).

The search strategy of this study contains resources like searched databases of Science Direct, Springer and IEEE Xplore, Google Scholar, and ResearchGate. Other sources include backtracking of relevant References from the selected review articles and primary studies, conference proceedings, and the internet. Search items were mainly Journal articles and conference papers. The authors also applied a full search on the full text to avoid excluding related studies that did not have the main search strings in their titles. The papers and articles considered in this review were of publication period between 2010-2020.

### **Student performance prediction**

Student performance prediction consists of identifying an unknown mark of a student. However, some factors can affect students' performance, making this task challenging to accomplish. The elements may include the economic status of a student, demographic characteristics, students' psychological profiles, past scholar experiences, different cultural backgrounds, and interactions between fellow students. Despite the above, for any institution to carry out a right performance prediction, majorly, they must first identify the risk factors that may affect prediction results. All educational institutions should always answer this critical question

before taking prediction steps; "What are the important risk factors or variables for predicting student performance?"

These risk factors may include employed student performance prediction methods, and the data sets considered to obtain prediction results. The data sets may include attributes of the previous semester grade, gender, attendance, GPA, parents' education, parents' income, scholarship, first child, etc. Different Institutions consider different data sets to predict student performance. However, diverse data sets work well with varying methods of prediction.

## **IV. Conclusion**

If the purpose of education in higher educational institutions and at all levels of education is to be realised, student performance is a critical aspect that must be closely monitored. This is due to the fact that predicting student achievement aids institutional leaders in enhancing their educational systems. The goal of this research was to look at some of the most regularly utilised categorization strategies for predicting student performance.

## **REFERENCES**

- [1] C. Wilkin, A. Ferreira, K. Rotaru, and L. R. Gaerlan, "Big data prioritization in SCM decision-making: Its role and performance implications," *Int. J. Account. Inf. Syst.*, no. xxxx, p. 100470, 2020, doi: 10.1016/j.accinf.2020.100470.
- [2] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. Ullah Khan, "The rise of 'big data' on cloud computing: Review and open research issues," *Inf. Syst.*, vol. 47, pp. 98–115, 2015, doi: 10.1016/j.is.2014.07.006.
- [3] C. Angeli, S. K. Howard, J. Ma, J. Yang, and P. A. Kirschner, "Data mining in educational technology classroom research: Can it make a contribution?," *Comput. Educ.*, vol. 113, pp. 226–242, 2017, doi: 10.1016/j.compedu.2017.05.021.
- [4] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," *Telemat. Informatics*, vol. 37, no. April 2018, pp. 13–49, 2019, doi: 10.1016/j.tele.2019.01.007.
- [5] P. Galetsi, K. Katsaliaki, and S. Kumar, "Big data analytics in health sector: Theoretical framework, techniques and prospects," *Int. J. Inf. Manage.*, vol. 50, no. March

- 2019, pp. 206– 216, 2020, doi: 10.1016/j.ijinfomgt.2019.05.003.
- [6] L. Yang, H. Xu, and Y. Jiang, “Applied research of data mining technology in hospital staff appraisal,” *Procedia Comput. Sci.*, vol. 131, pp. 1282–1288, 2018, doi: 10.1016/j.procs.2018.04.340.
- [7] Y. Liang, X. Zheng, and D. D. Zeng, “A survey on big data-driven digital phenotyping of mental health,” *Inf. Fusion*, vol. 52, no. March, pp. 290–307, 2019, doi: 10.1016/j.inffus.2019.04.001.
- [8] P. T. Chen, C. L. Lin, and W. N. Wu, “Big data management in healthcare: Adoption challenges and implications,” *Int. J. Inf. Manage.*, no. December 2019, p. 102078, 2020, doi: 10.1016/j.ijinfomgt.2020.102078.
- [9] H. Ait Issad, R. Aoudjit, and J. J. P. C. Rodrigues, “A comprehensive review of Data Mining techniques in smart agriculture,” *Eng. Agric. Environ. Food*, vol. 12, no. 4, pp. 511–525, 2019, doi: 10.1016/j.eaef.2019.11.003.
- [10] R. Aarthi and D. Sivakumar, “An Enhanced Agricultural Data Mining Technique for Dynamic Soil Texture Prediction,” *Procedia Comput. Sci.*, vol. 171, no. 2019, pp. 2770–2778, 2020, doi: 10.1016/j.procs.2020.04.301.
- [11] X. E. Pantazi, D. Moshou, and D. Bochtis, “General overview of the proposed data mining and fusion techniques in agriculture,” *Intell. Data Min. Fusion Syst. Agric.*, no. iii, pp. 287–313, 2020, doi: 10.1016/b978-0-12-814391-9.00010-8.
- [12] F. Wang et al., “Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns,” *Energy Convers. Manag.*, vol. 171, no. June, pp. 839–854, 2018, doi: 10.1016/j.enconman.2018.06.017.
- [13] Z. Guo, K. Zhou, X. Zhang, S. Yang, and Z. Shao, “Data mining based framework for exploring household electricity consumption patterns: A case study in China context,” *J. Clean. Prod.*, vol. 195, pp. 773–785, 2018, doi: 10.1016/j.jclepro.2018.05.254.
- [14] M. Wu and Y. Moon, “Alert Correlation for Cyber-Manufacturing Intrusion Detection,” *Procedia Manuf.*, vol. 34, pp. 820–831, 2019, doi: 10.1016/j.promfg.2019.06.197.
- [15] G. P. Gupta and M. Kulariya, “A Framework for Fast and Efficient Cyber Security Network Intrusion Detection Using Apache Spark,” *Procedia Comput. Sci.*, vol. 93, no. September, pp. 824–831, 2016, doi: 10.1016/j.procs.2016.07.238.
- [16] I. Dutta, S. Dutta, and B. Raahemi, “Detecting financial restatements using data mining techniques,” *Expert Syst. Appl.*, vol. 90, pp. 374–393, 2017, doi: 10.1016/j.eswa.2017.08.030.
- [17] S. Cai and J. Zhang, “Exploration of credit risk of P2P platform based on data mining technology,” *J. Comput. Appl. Math.*, vol. 372, p. 112718, 2020, doi: 10.1016/j.cam.2020.112718.
- [18] C. wen Shen, Min Chen, and C. chen Wang, “Analyzing the trend of O2O commerce by bilingual text mining on social media,” *Comput. Human Behav.*, vol. 101, pp. 474–483, 2019, doi: 10.1016/j.chb.2018.09.031.
- [19] A. Peña-Ayala, “Educational data mining: A survey and a data mining-based analysis of recent works,” *Expert Syst. Appl.*, vol. 41, no. 4 PART 1, pp. 1432–1462, 2014, doi: 10.1016/j.eswa.2013.08.042.
- [20] M. Ashraf, M. Zaman, and M. Ahmed, “An Intelligent Prediction System for Educational Data Mining Based on Ensemble and Filtering approaches,” *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 1471–1483, 2020, doi: 10.1016/j.procs.2020.03.358.
- [21] S. K. Mohamad and Z. Tasir, “Educational Data Mining: A Review,” *Procedia - Soc. Behav. Sci.*, vol. 97, pp. 320–324, 2013, doi: 10.1016/j.sbspro.2013.10.240.
- [22] M. W. Rodrigues, S. Isotani, and L. E. Zárate, “Educational Data Mining: A review of evaluation process in the e-learning,” *Telemat. Informatics*, vol. 35, no. 6, pp. 1701–1717, 2018, doi: 10.1016/j.tele.2018.04.015.
- [23] 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, 2015, doi: 10.1016/j.procs.2015.12.157.
- [24] A. Viloría, J. G. Padilla, C. Vargas-Mercado, H. Hernández-Palma, N. O. Llinas, and M. A. David, “Integration of data technology for analyzing university dropout,” *Procedia Comput. Sci.*, vol. 155, no. 2018, pp. 569–574, 2019, doi: 10.1016/j.procs.2019.08.079.
- [25] M. Mayilvaganan and D. Kalpanadevi, “Comparison of classification techniques for predicting the cognitive skill of students in education environment,” *2014 IEEE Int. Conf. Comput. Intell. Comput. Res. IEEE ICCIC 2014*, pp. 113–118, 2015, doi: 10.1109/ICCIC.2014.7238346.
- [26] K. T. Chui, D. C. L. Fung, M. D. Lytras, and T. M. Lam, “Predicting at-risk university students in a virtual learning environment via a machine learning algorithm,” *Comput. Human Behav.*, p. 105584, 2018, doi: 10.1016/j.chb.2018.06.032.
- [27] P. Kaur, M. Singh, and G. S. Josan, “Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector,” *Procedia Comput. Sci.*, vol. 57, pp. 500–508, 2015, doi: 10.1016/j.procs.2015.07.372.