

Student Dropout Analysis: A Literature Review in the Context of Higher Education

Kanika Dhingra¹, Sanjay Kumar Sharma²

^{1,2}Department of Computer Science, Banasthali Vidyapith University, Rajasthan.
kanika.dhingra.2@gmail.com¹. Skumar2.sharma@gmail.com²

Abstract

Education is a crucial requirement in everyone's life in order to make them employable and earn their living. However, attaining education with good academic achievement remains a key challenge to address. Over the time it has been observed that all students who enroll for the education do not accomplish the goal of attaining a degree. Some of the students dropout before completion of the course due to numerous reasons. Higher education dropouts have a wide range of effects on society as a whole, the job market, and the higher education system in addition to the student. These include the loss of funds intended for higher education, reduced labour competitiveness, and incomplete education. Students, their parents, universities, the economy, and society at large are all impacted by dropout rates. Therefore, determining the causes of dropout is crucial so that policies can be put in place to lessen it.

Keywords: Student dropout prediction, Machine Learning, Educational Data Mining, Learning Analytics, Higher education

1. Introduction

Now a days the data is being generated at an exponential rate. Data's growing availability in recent times has simplified the tracing, understanding, and forming opinions about activities in different fields. It has also influenced decision-making and initiatives to increase the efficacy of every process. The term "data" in data mining refers to big information sets that are processed and examined to identify trends and patterns. This factual information is analyzed to make informed decisions. In short, data is a systematic record that is majorly gathered by observations, measurements, research, or analysis. The term data mining refers to a computer-assisted process employing various tools and techniques for finding hidden links in data. Primarily the process of data mining comprises of four main stages namely data gathering, data preparation, data mining and data analysis. Applications for data mining are numerous and span a number of sectors, including marketing, banking, healthcare, finance, manufacturing engineering, education, Customer relationship management, Fraud detection, lie detection and telecommunications. These days, a lot of research is being done on data mining and its use in education, which is known as educational data mining. The process to investigate information using data generated from educational environments is the focus in the recently established field of educational research [18].

EDM can be used to collect, analyse, and present information in an organized manner and the academic institutions can use this extracted knowledge to enhance the processes of teaching and learning as well as improve performance of the learners, thus supporting students'

academic achievement [19]. It aids in the early detection of underachievers for the prompt action and lower the dropout rate. A key issue in all higher education institutions is student dropout, which is defined as the cessation of study or the inability to finish a degree. Beyond the immediate effects on the student, this phenomenon has other repercussions. First, a student's employment and professional growth prospects that might otherwise be possible with a degree are limited due to their incomplete education. A person's financial and personal well-being may suffer as a result. Furthermore, dropping out lowers a nation's workforce's overall educational attainment, which may have an impact on its ability to compete internationally. Thirdly, financial resources used for higher education are also lost when students leave out. Tuition, scholarships, and other expenses incurred by students, their families, the government, and post-secondary educational institutions are included in this. Dropouts affect society as a whole by lowering competitiveness and degrading social and economic advancement. Over the years, the field of student dropout has attracted a lot of attention in the education sector because the higher education institutions and academicians are curious to know the reasons of student dropout. Various institutions, particularly in the US, report concerningly high dropout rates, with around 30% of students ending their studies before completing their second year in colleges and universities [2]. Considering that student dropout poses a substantial issue for students, the educational community, and policymakers, research has been undertaken to assess and identify profiles of students at risk of disengagement [1]. Numerous reasons, such as academic, socioeconomic, and psychological ones, contribute to dropout. In addition to the above factors the family and social factors also influence the student dropout to a large extent. Parent's education, parent's involvement in studies, family income, influence of peer group and support from the community are categorised as family and social factors. The parent's involvement and support in their child's education are strongly linked to decreased dropout rates [17]. Numerous factors, such as loss of interest in education, relocation, absenteeism, parent deaths and separations, peer pressure, family responsibilities, gender, behaviour, early pregnancies, early marriages, and a negative perception of education in the community, are among the primary causes of student dropouts in the community, according to the findings.

By identifying and profiling students who are at risk of dropping out, educational stakeholders may help them get the support they need to stay engaged in their programs [3]. Early identification of at-risk student is necessary to address the issue of student dropout. Early detection may allow university employees to take proactive measures to assist students who require assistance. Early detection of these students is advantageous and possible because many students who ultimately leave college exhibit indications during their initial year of study [4]. Tinto [5] discussed the requirements, circumstances, and different student attributes connected to the dropout problem in the 1970s, which marked the beginning of research on student dropout prediction. The study developed a model that included these components, including course grades, learning behaviour, and socioeconomic considerations. This study aims at conducting a systematic literature review to answer the following research question:

RQ1: To identify factors that significantly impact the decision of student dropout in higher education.

This paper is structured as follows: section 2 provides the methodology adopted for systematic review. Section 3 outlines the factors that have a significant impact on the decision of student

dropout in higher education. Section 4 includes the discussion and conclusion drawn from the systematic literature review.

2. Research Methodology

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines served as the foundation for the systematic review approach used to construct the current study.

The systematic review was developed in the following stages: First, the study's guiding research topic was developed. After that, a research protocol was created that detailed the systematic review's design, including the selection strategy for the studies, the databases that were searched, the search tactics, and the techniques for extracting and analysing data. The studies' inclusion and exclusion criteria were also determined.

Using the specified search terms and the inclusion and exclusion criteria, a thorough search of the scientific literature across several databases was then carried out in order to find pertinent studies. After reviewing the titles and abstracts of the papers found through the search, those that satisfied the predetermined inclusion criteria were chosen. The chosen studies were then thoroughly studied to ensure they met the inclusion and exclusion requirements. The relevant data was taken out of the chosen research and arranged in a database. Lastly, the findings were summarised and the results were analysed, pointing out any potential drawbacks.

Search Strategy

An extensive search was conducted in particular databases to find the appropriate information for our study in order to carry out this systematic review. The search string (“student” AND "predict*" AND "Dropout" AND "factors" AND “review” OR “survey”) was used to search for articles. The sample was limited to examining just those publications that offered information pertinent to the goal after the inclusion and exclusion criteria were applied.

Inclusion and exclusion criteria

The standards and guidelines set forth to decide which papers or publications will be taken into consideration for the systematic review and which will be disqualified are referred to as the inclusion and exclusion criteria in this scientific study. Table1 presents the inclusion and exclusion in detail. These standards are predicated on the goals and enquiries of the research.

Table1. Inclusion and Exclusion Criteria

Features	Inclusion Criteria	Exclusion Criteria
Participants	Students pursuing Higher education	Students pursuing post-graduation.
Time span	Studies from 2010-2025	Studies other than this interval.
Languages	English	Language other than English.
Emphasis	Qualitative approach	Quantitative approach

Sample Screening Procedure

According to the flow chart that is attached, 246 items were first found in the three databases. 67 articles were shortlisted after the inclusion and exclusion criteria were applied and irrelevant articles were removed. For a variety of reasons, more articles were excluded from this list. 40 papers in all were eventually included in the analysis (Figure 1).

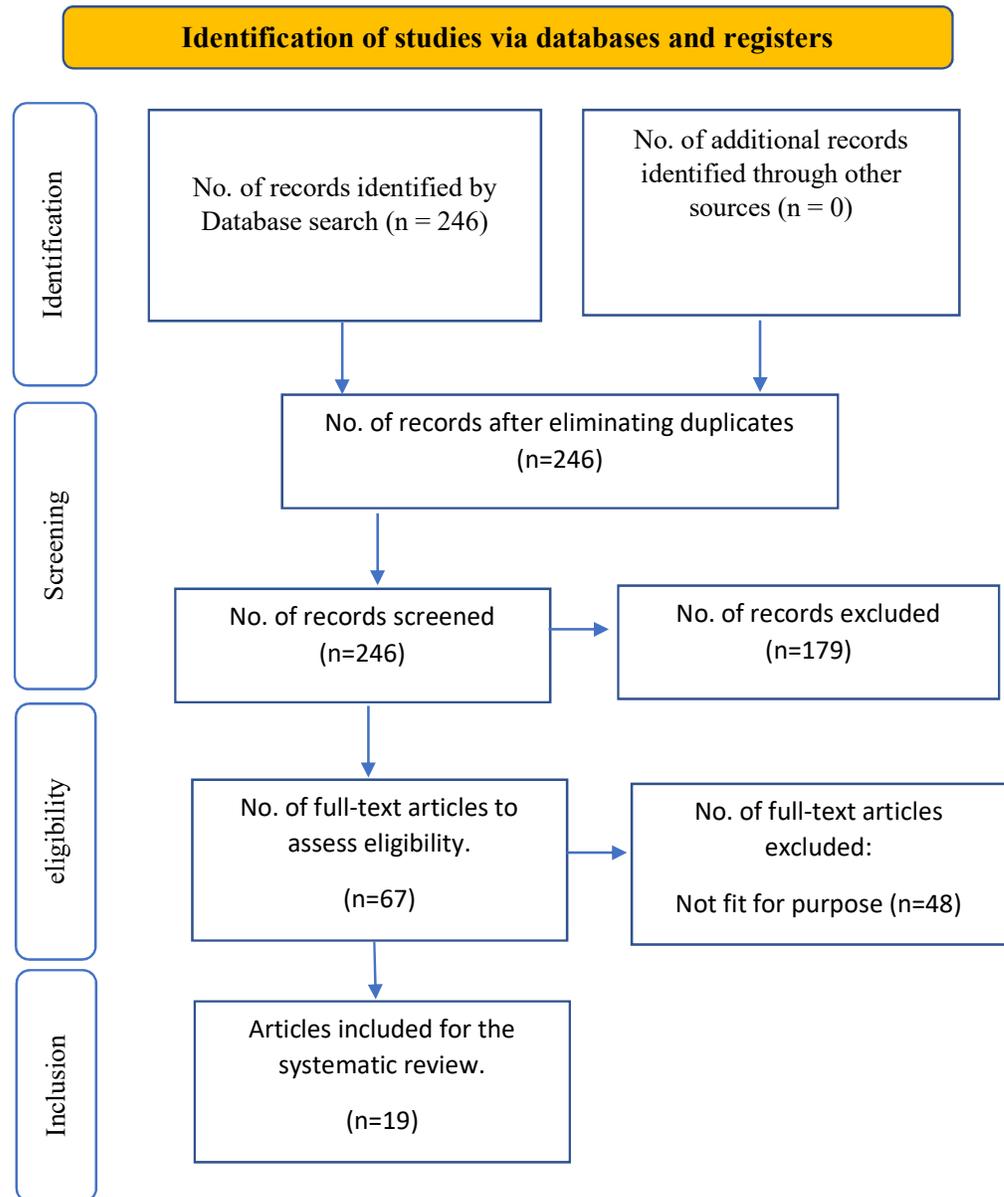


Figure 1: Flowchart of the search and selection method for the systematic review references.

A thorough report of the systematic review is written in the final stage, which includes a thorough explanation of the methodology that is used, the findings, and the conclusions drawn.

Search Methodology

The systematic review approach provides an extensive analysis of articles from databases like Scopus, IEEE Xplore, SpringerLink and ScienceDirect, focused on summarizing, synthesizing, and critically analyzing existing research to offer clear, evidence-based insights into the field of Student Dropout. "Student dropout," "machine learning techniques," and "higher education" were among the search terms used. Publications that provide empirical research on the causes of student dropout, with a special emphasis on higher education, were included. We thoroughly examined the main causes of student dropouts in higher education. In order to establish the foundation for more thorough future study, this literature analysis offers a basic overview of the dropout problem, highlighting pertinent indicators of student dropout.

3. Factors impacting Decision of Student Dropout in Higher Education

The problem of student dropout in higher education has gained significant attention over the years. Many students who eventually drop out of university display signs during their initial year of studies. This pervasive issue contributes to adverse effects on students, educational institutions, and the broader economy. The repercussions extend to taxpayers, who bear a substantial financial burden funding the education of students unable to complete their studies [6]. The issue of student dropout stems from a variety of factors, including academic, socio-economic, and psychological. The decision of dropout in higher education is highly influenced by the academic performance of the student.

According to the majority of research papers in the literature, the following factors influence students' academic performance: gender, high school grade, parental education, financial background, teaching medium, family status, living location, previous semester grades, seminar performance class, test grade, assignment performance, attendance in class and lab work, general proficiency, interest in a particular course, study behaviour, engage time and family support for study, previous school marks, admission type, accommodation type, parent's occupation, and parent's qualification. These qualities are all categorised into academic, institutional, social, family, and personal categories. The most important personal characteristics of the student taken into consideration include age, gender, interest in the subject, type of admission, and study habits [7], [8], [9], [10], [11]. Family characteristics that are thought to be significant for academic prediction include parental education, occupation, family income, family status, and study support [8], [11], [12]. From academic point of view features such as high school grades, students' grades from previous semesters, class exam grades, seminar performance, assignment performance, attendance in class, and lab work, prior school grades are taken into consideration [12], [8], [7], [13], [10], [11], [14], [15] and institutional characteristics such the mode of education, kind of lodging, infrastructure, lavatory and water facilities, teaching style, and transportation [15], [8], [16], [13], [12], [10], [7]. Hence Student academic performance can be considered to be influenced by a complex interplay of internal (individual-level) and external (environmental or institutional-level) factors as depicted in figure 2.

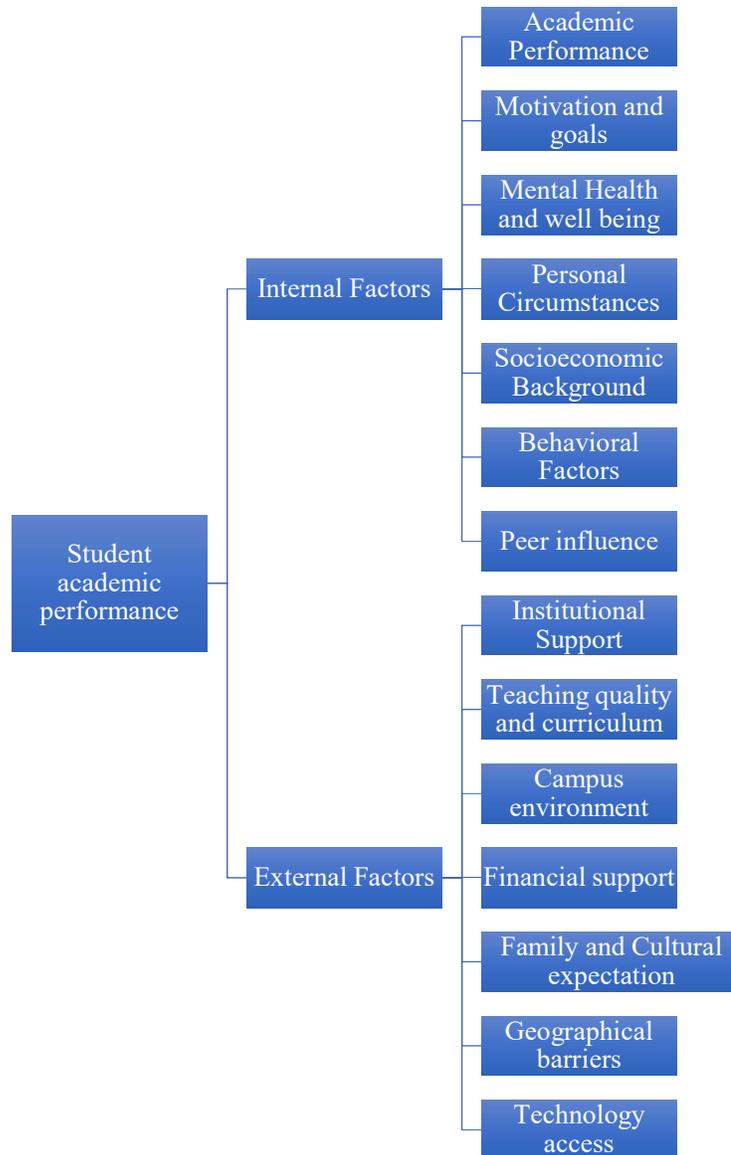


Figure 2: Factors influencing Student Academic Performance

4. Discussion and conclusion

For universities, dropout is a significant issue that must be resolved. The purpose of this study was to determine the factors affecting the decision of student dropout in Higher education. Taking into consideration the current status of research, very few papers focus on identification of factors that are responsible for student dropout decision in higher education. The prediction techniques used to forecast students' academic achievement and at-risk students are the focus of most of the studies. There are a wide range of factors that influence the decision of student dropout in Higher education and this paper has identified various factors affecting the decision of student dropout while also taking into consideration the most significant ones. According to our research, the most significant factors that influence student dropout decision are academic performance, motivation and goals, mental health and wellbeing, personal circumstances, socio

economic background, behavioural factors, peer influence, Institutional support, teaching quality and curriculum, campus environment, Financial support, family and cultural expectations, Geographical barriers, technology access. The findings of this study are important because they serve as the foundation for future investigations aimed at identifying and assisting dropout students. For educational institutions, student success is a key indicator of effectiveness and plays a critical role in their operations, reputation, and contributions to society. When current students succeed and meet workforce demands, it creates a positive experience that leads to better retention rates. From an institutional point of view, student success is not just a reflection of educational outcomes but also a driver of growth, sustainability, and societal impact. It strengthens their position as credible, effective, and impactful centres of learning.

References

- [1] L. Aulck, N. Velagapudi, J. Blumenstock, and J. West, "Predicting student dropout in higher education," arXiv.org, Jun. 20, 2016. Available: <https://arxiv.org/abs/1606.06364>
- [2] L. Aulck et al., "Mining University registrar records to predict First-Year Undergraduate attrition," in Proceedings of The 12th International Conference on Educational Data Mining (EDM 2019), 2019, pp. 9–18. Available: <https://files.eric.ed.gov/fulltext/ED599235.pdf>
- [3] A. O. Ameen, M. A. Alarape, and K. S. Adewole, "STUDENTS' ACADEMIC PERFORMANCE AND DROPOUT PREDICTION," MALAYSIAN JOURNAL OF COMPUTING, vol. 4, no. 2, p. 278, Nov. 2019, doi: 10.24191/mjoc.v4i2.6701. Available: <https://doi.org/10.24191/mjoc.v4i2.6701>
- [4] G. J. A. Baars and I. J. M. Arnold, "Early identification and characterization of students who drop out in the first year at university," Journal of College Student Retention Research Theory & Practice, vol. 16, no. 1, pp. 95–109, May 2014, doi: 10.2190/cs.16.1.e. Available: <https://doi.org/10.2190/cs.16.1.e>
- [5] V. Tinto, "Dropout from Higher Education: A Theoretical Synthesis of Recent Research," Review of Educational Research, vol. 45, no. 1, p. 89, Jan. 1975, doi: 10.2307/1170024. Available: <https://doi.org/10.2307/1170024>
- [6] M. Jadrić, Z. Garać, and M. Čukurić. Student dropout analysis with application of data mining methods. Management: journal of contemporary management issues, 15(1):31–46, 2010.
- [7] Sharma, H., Kumar, P., & Sharma, K. (2025). Smart Waste Management with IoT: An Optimized Triple Memristor Hopfield Neural Network Approach. International Journal on Smart & Sustainable Intelligent Computing, 2(1), 52-64.
- [8] F. Ahmad, N. H. Ismail, and A. A. Aziz, "The prediction of students' academic performance using classification data mining techniques," Appl. Math. Sci., vol. 9, no. 129, pp. 6415–6426, 2015, doi: 10.12988/ams.2015.53289.
- [9] K. David Kolo, S. A. Adepoju, and J. Kolo Alhassan, "A Decision Tree Approach for Predicting Students Academic Performance," Int. J. Educ. Manag. Eng., vol. 5, no. 5, pp. 12–19, Oct. 2015, doi: 10.5815/ijeme.2015.05.02.
- [10] Gusain, N., & Sharma, H. (2025). Communication-Efficient Federated Learning in Industrial IoT—A Framework for Real-Time Threat Detection and Secure Device Coordination. International Journal on Computational Modelling Applications, 2(2), 18-29.
- [11] M. P. G. Martins, V. L. Migueis, and D. S. B. Fonseca, "Uma Metodologia de Data Mining para Prever o Desempenho de Estudantes de Licenciatura," Iber. Conf. Inf. Syst. Technol. Cist., vol. 2018-June, pp. 1–7, 2018, doi: 10.23919/CISTI.2018.8399175.

- [12] K. Nahar, B. I. Shova, T. Ria, H. B. Rashid, and A. H. M. S. Islam, "Mining educational data to predict students performance: A comparative study of data mining techniques," *Educ. Inf. Technol.*, vol. 26, no. 5, pp. 6051–6067, 2021, doi: 10.1007/s10639-021-10575-3.
- [13] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," *Comput. Educ.*, vol. 113, pp. 177–194, Oct. 2017, doi: 10.1016/j.compedu.2017.05.007.
- [14] Nandan, A., Pradhan, L., Sharma, H., Bhola, A., & Srivastava, A. (2025, December). Multi-Task Learning Model for Fine-Grained Categorization of Emergency Tweets. In 2025 IEEE 17th International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 2045-2050). IEEE.
- [15] A. B. E. D. Ahmed and I. S. Elaraby, "Data Mining: A prediction for Student's Performance Using Classification Method," *World J. Comput. Appl. Technol.*, vol. 2, no. 2, pp. 43–47, 2014, doi: 10.13189/wjcat.2014.020203.
- [16] Sharma, V. K., Arnold, K. S., Mishra, P., & Sharma, H. (2021). Focaltheorem–A portfolio Web Application. *Turkish Journal of Computer and Mathematics Education*, 12(3), 3809-3814.
- [17] D. H. Ouma, Z. Ting, and J. C. Pesha, "Analysis of the Socio-Economic factors that contribute to children school dropout in artisanal Small-Scale gold mining communities of Tanzania." Available: <https://eric.ed.gov/?id=EJ1143926>
- [18] P. Peterson, E. Baker, and B. McGaw, "International Encyclopedia of Education | ScienceDirect," *Sciencedirect.com*, 2010. Available: <https://www.sciencedirect.com/referencework/9780080448947/international-encyclopedia-of-education>
- [19] S. Shrestha and M. Pokharel, "Machine Learning algorithm in educational data," 2019 Artificial Intelligence for Transforming Business and Society (AITB), Nov. 2019, doi: <https://doi.org/10.1109/aitb48515.2019.8947443>.