



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Predicting of Academic Achievement of Student Using Machine Learning Model

Moola Koushik, Koduri Bhava Priya, Kshatriya Vaishnavi, Kella Sai Ganesh Pavan Kumar

UG Scholar, Dept. of Computer Science Engineering (Artificial Intelligence and Data Science), Satya Institute of Technology and Management, Vizianagaram, India

Ramarapu Bangari

Assistant Professor, Dept. of Computer Science Engineering (Artificial Intelligence and Data Science), Satya Institute of Technology and Management, Vizianagaram, India

ABSTRACT: Predicting university student graduation is a beneficial tool for both students and institutions. With the help of this predictive capacity, students may make well-informed decisions about their academic and career paths, and institutions can proactively identify students who may not graduate and offer tailored support to ensure their success. The use of machine learning for predicting university student graduation has drawn more attention in recent years. Large datasets of student academic performance data can be used to train machine learning algorithms to identify patterns that are applicable in predicting future outcomes. In accordance with some studies, this approach predicts student graduation with an accuracy rate as high as 90%. Many systematic literature reviews (SLRs) have been conducted in this field, but there are still limitations, including not discussing the predictive models and algorithms used, a lack of coverage of the machine learning algorithms applied, small database coverage, keyword selection that does not cover all synonyms relevant to the investigation, and less specific data collection transparency.

KEYWORDS: Datamining, Educational analysis, systematic literature review, machine learning

I. INTRODUCTION

The diversity and complexity of today's world have made higher education essential, and as a result, the body of scientific literature devoted to forecasting academic achievement or the likelihood of student dropout has grown significantly [1]. Higher education institutions' traditional function of disseminating knowledge has evolved; the training of skilled professionals and innovation in new knowledge, particularly with the rise of artificial intelligence [2], have led to many of them interacting with various facets of society. Actually, the main goals of its academic and cultural activities are to raise the standard of knowledge in society through teaching, research, and the capacity to share and transmit this information. In order to produce competitive professionals in society, they play a crucial role in imparting information, skills, and values to pupils. Since HEIs must carry on the job started and further expand their engagement, relevance, and service capacity in connection to the social, cultural, and economic context, guiding students towards academic achievement is therefore transcendental [3].

As a result, educational managers are interested in using historical data on students who have successfully finished their university education to forecast academic achievement. This is because it enables them to make better judgements and develop machine learning alternatives for improvement. Therefore, developments in machine learning methods and other fields of research are forerunners of educational data mining. The graduation rate, which is the total number of students who graduate out of all the students who enrol, is a statistical indicator of students' academic performance in higher education. Indeed, by examining both endogenous and exogenous elements in the student environment, it is feasible to consider student achievement in a broader sense [4]. As a result, the ongoing desire to improve students' academic performance has prompted machine learning to be customised in order to provide particular prediction models that offer valuable data.

Numerous studies in the past ten years have concentrated on research projects that tackle university students' performance, dropout, and academic achievement issues. The authors stress that students from underprivileged social

strata who exhibit university dropout behaviour are more likely to experience failure or drop out of college, as explained in [5]. The authors are inclined to think that extracurricular activities that ensure retention should be improved in order to maintain university permanency among their findings. Early detection has so emerged as a means of resolving these issues. The findings were predicted using a variety of input components, including academic background, university surroundings (both material and intangible resources), and other data [6]. To address these issues, both qualitative and quantitative research methodologies have been applied. In more recent times, there have been several studies that have used data mining or machine learning approaches, which include the use of algorithms and two well-known methods to extract valuable information from data. The first method, supervised classification, predicts the target variable (class) by analysing the data. There have been findings pertaining to supervised categorisation in the work of [7–8].

Similarly, they employed a collection of pre-selected algorithms that categorise the data by applying the voting methodology in [9–10], which is another method based on supervised classification. Both strategies aim to accurately forecast pupils' academic performance or achievement. Finding hidden patterns in the data is the main goal of the second strategy, unsupervised classification, in which the objective variable is unknown. Association rules, which are made up of two parts—antecedent and consequent—are generally used to find facts that occur inside the data. For instance, the rule $\{A, B\} \Rightarrow \{C\}$ indicates that when A and B occur, then C happens. By concentrating on the association rules and assessing them using metrics like lift, support, and confidence, among others, they search for the recurrence of data in [11–12].

Research in the form of a systematic literature review (SLR) was carried out in order to obtain a comprehensive grasp of how academic performance may be employed in student evaluation. An SLR is a technique for locating, assessing, and condensing the body of research on a certain subject. The findings of studies conducted in the field of EDM over the past five years (2018–2023) are used in this review of the literature. The research's data sources, methodology, variables used for the prediction phase, and software are all included in the review. The first of the five portions of the SLR is devoted to the SLR's preparation. The research methodology will be thoroughly examined in the second section, and the results and answers to the study questions will be reviewed in the third. The last section will present a summary of the study findings and, based on the findings, provide recommendations for further research. The fourth section will examine the constraints and challenges encountered. The learner module in smart learning environments has advanced significantly as a result of the current study.

II. METHODOLOGY

The phrase "context" refers to the particular setting or set of conditions in which the intervention being studied was conducted. Various characteristics, such as the organisational type, size, and development process, as well as the tools and techniques used, may be included in the contextual components covered in this study. Higher education is the context in which this review is located. The following research questions were derived from the PICOC:

- RQ1: What is the current level of research in higher education on educational data mining (EDM) and educational data analytics (EDA)?
- RQ2: What data sources, factors, and collecting methods have been employed to use EDM to forecast students' academic achievement in higher education?
- RQ3: What are the shortcomings and difficulties in EDM research?

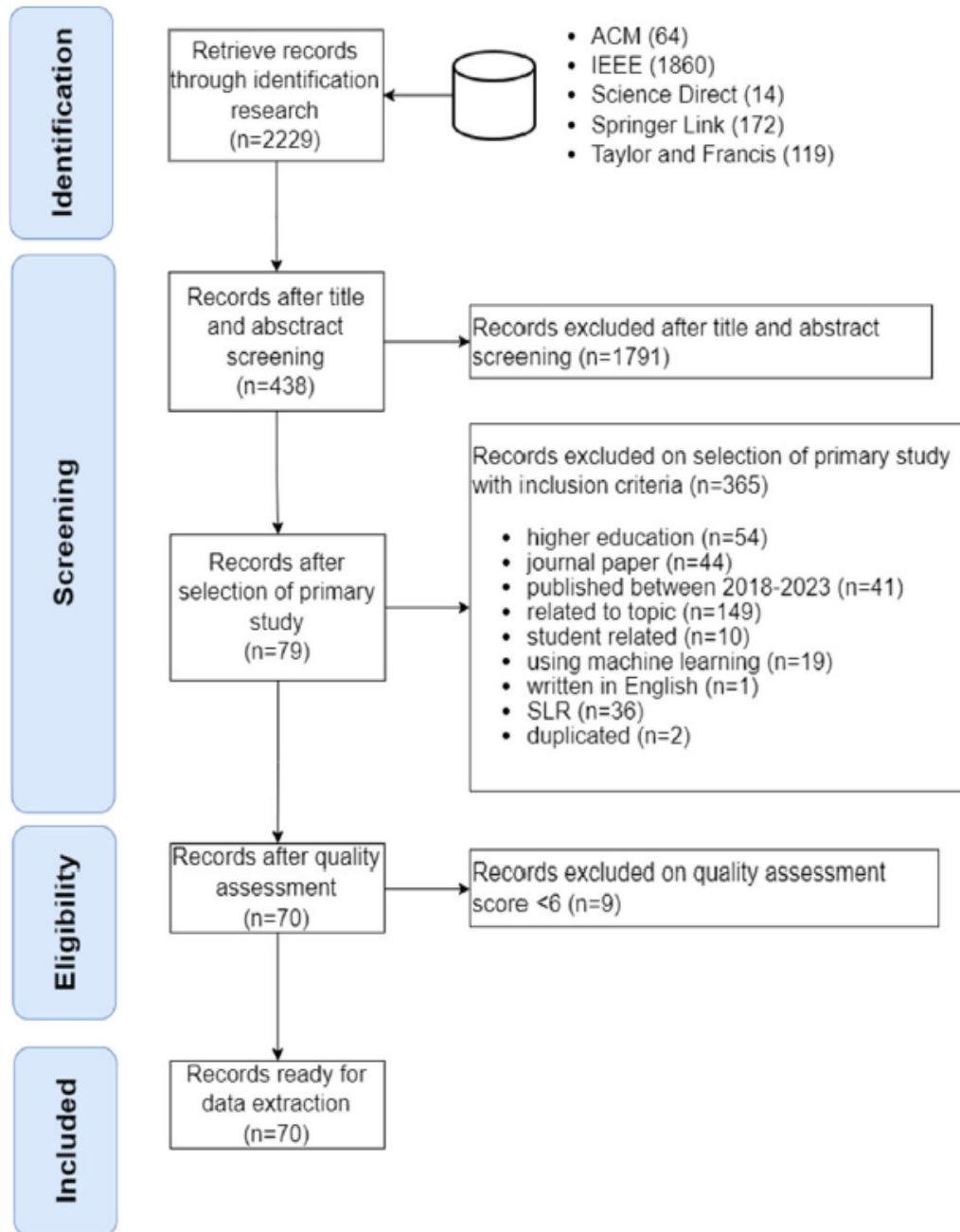


Figure 1. Flowchart for search and selection methodology

It is essential to thoroughly select and support a search strategy for answering the research questions before starting the investigation. This involves a careful assessment of the databases and other resources that will be investigated, as well as the keywords and language used in the search for scholarly material. The search strategy's formulation is tailored to the specific research question. Once the search strategy has been developed, a thorough search across several significant electronic sources must be carried out. This study includes academic resources including Taylor & Francis, ACM Digital Library, IEEE Digital Library, Science Direct, and Springer Link. Following the guidelines will increase the possibility that we will find as many primary studies as possible that are pertinent to answering the research questions.

After identifying the database, the next stage is to create a thorough inventory of synonyms, abbreviations, and other spellings based on the PICOC framework, research objectives, and important, pertinent phrases from titles and abstracts. The purpose of this procedure is to make choosing relevant keywords easier. The following keywords are used in the study: student performance, higher education, learning analytics, machine learning, educational data analytics, early warning system of student dropout, dropout prediction, and student graduation prediction. Boolean operators like AND and OR may then be used to create more complex keyword searches. "Educational data analytics," "educational data mining," and "learning analytics" are the search terms that were developed and used in this literature review. They are also combined with "machine learning," "dropout prediction," "early warning system of student dropout," and "student graduation prediction." Finally, "higher education" is included.

III. RESULTS AND DISCUSSION

A. System Environment

The proposed model is developed using said hardware system configuration and software required to implement the design is shown below.

Hardware System Configuration:

- Processor: Intel i5
- RAM : 8 GB
- Hard Disk : 256 GB
- Key Board : Standard Windows Keyboard
- Mouse : Two or Three Button Mouse
- Monitor : SVGA

Software Requirements:

- Operating system : Windows 7 Ultimate.
- Coding Language: Python.
- Front-End : Python.
- Back-End : Django-ORM
- Designing : Html, css, javascript.
- Data Base : MySQL (WAMP Server).

B. Experimental findings

LMS and SIS are the main data sources used in this study, mostly in response to RQ2. Academic, behavioural, demographic, pre-university, and university admission test data are among the variables commonly used for prediction. According to this study, the most often used data sources are SIS and LMS. Subsequent research revealed that the data produced by the two data sources was the cause of this. SIS and LMS offer a wealth of information on students' academic development. While SIS mainly keeps demographic information, enrolment details, and academic records, LMS stores data on how students interact with digital learning materials, assignments, and assessments. Both systems offer longitudinal data, which makes it possible for machine learning algorithms to examine pupils' development over time. Predicting academic results and identifying trends or patterns that may have an influence on student accomplishment are two areas in which the longitudinal method excels. Because they provide easy access to the data they contain, both data sources are crucial components of educational institutions. The implementation of machine learning models is made possible by this accessibility, which eliminates significant barriers to data acquisition.

The ANN/MLP model and the k-NN technique were also used in the study. According to the study, SVM, RF, and LR are now widely used algorithms to forecast students' academic achievement. SVM has a regularisation parameter that helps control overfitting, which is crucial when working with sparse data in tasks that predict student performance [37]. All three approaches can handle both binary and multi-class classification problems and have demonstrated effectiveness in solving previous educational prediction tasks. Additionally, the parameters of these algorithms may be adjusted to maximise performance for certain datasets. Because of these models' adaptability, researchers and data scientists may tailor them to the unique characteristics of the educational data being used. The techniques may be modified to work with datasets of any size. Algorithms that can handle a variety of scales are desirable since educational datasets contain a vast range of student counts and attributes. RQ3 deals with identifying the study's problems and research needs. These include the lack of available data, incomplete data from poor data collection techniques, reliance on a single data source, limitations of some models that

make it difficult for them to evaluate particular kinds of data, a lack of attention to significant studies, possible oversimplification, and an over-reliance on features or measurements, which can introduce bias into the interpretation of results. The importance of ethical considerations is especially pertinent when using AI and machine learning techniques. Concern over the possible unintended consequences of predicting or intervention on kids is growing.

By recognising the extensive usage of LMS and SIS as important data sources, this study can help smart learning environments, especially learner's modules. The learner's module can specifically use pre-university, university entrance, behavioural, academic, and demographic data to create predictions. Furthermore, knowledge of SVM as the most popular approach, followed by RF, LR, k-NN, and ANN/MLP, can help the learner's module choose or enhance machine learning techniques that can be applied. The outcomes of learners' modules can also be used by educational policies to carry out more thorough and well-informed interventions.

The processed results and user interface results are shown in below figures.



Figure2. Login Interface

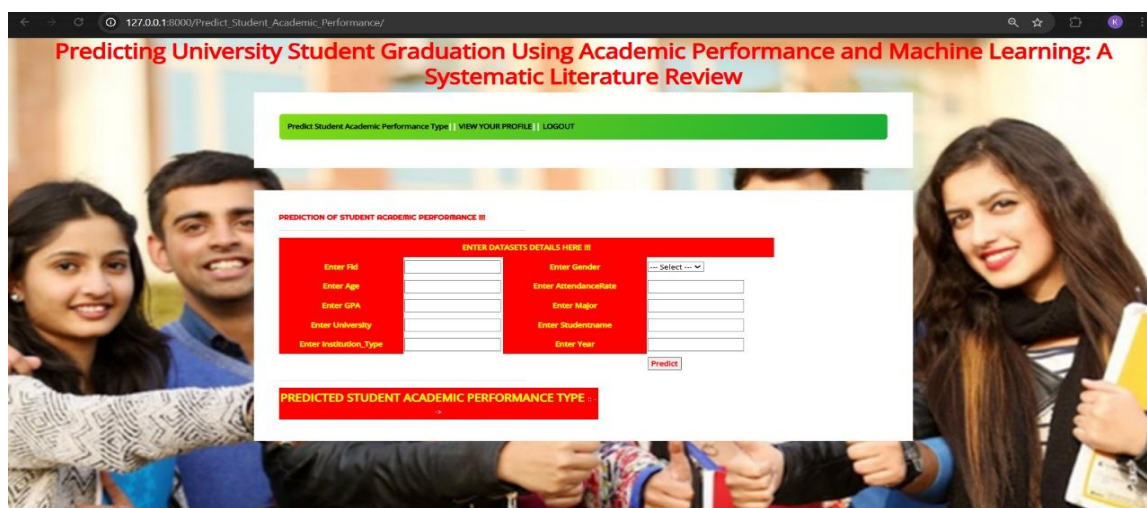


Figure 3. Interface for entering student details



Figure 4. Predicted output of student performance

IV. CONCLUSION

The utilization of predictive insights has the potential to enhance the creation of customized or adaptable learning experiences that suit the unique needs of individual students. For instance, students who are anticipated to encounter difficulties in a specific academic discipline should be offered further materials or alternate pedagogical approaches. The prediction of academic success yields vital data for the field of educational research. This study aims in future to enhance comprehension of the determinants that impact student achievement and the intricate interplay between many variables. Future research in the field of machine learning for predicting student graduation should aim beyond the mere task of outcome prediction. The goal should be to revolutionize educational methods by equipping educators with practical knowledge to customize learning experiences, detect students at risk of failure at an early stage, and apply specific interventions that foster student achievement. Researchers may develop a dynamic and transformative approach that promotes student achievement and enhances educational results for all by efficiently combining advanced methods, analyzing various data sources, and utilizing emerging machine learning approaches.

REFERENCES

1. Realinho, V.; Machado, J.; Baptista, L.; Martins, M.V. Predicting Student Dropout and Academic Success. *Data* **2022**, *7*, 146.
2. Rafik, M. Artificial Intelligence and the Changing Roles in the Field of Higher Education and Scientific Research. In *Artificial Intelligence in Higher Education and Scientific Research. Bridging Human and Machine: Future Education with Intelligence*; Springer: Singapore, 2023; pp. 35–46.
3. BOE. BOE-A-2023-7500 Ley Orgánica 2/2023, de 22 de marzo, del Sistema Universitario. 2023. Available online: <https://www.boe.es/buscar/act.php?id=BOE-A-2023-7500> (accessed on 23 March 2024).
4. Guney, Y. Exogenous and endogenous factors influencing students' performance in undergraduate accounting modules. *Account. Educ.* **2009**, *18*, 51–73.

5. Tamada, M.M.; Giusti, R.; Netto, J.F.d.M. Predicting Students at Risk of Dropout in Technical Course Using LMS Logs. *Electronics* **2022**, *11*, 468.
6. Lopes Filho JA, B.; Silveira, I.F. Early detection of students at dropout risk using administrative data and machine learning. *RISTI—Rev. Iber. De Sist. E Tecnol. De Inf.* **2021**, *40*, 480–495.
7. Fernández, A.; del Río, S.; Chawla, N.V.; Herrera, F. An insight into imbalanced Big Data classification: Outcomes and challenges. *Complex Intell. Syst.* **2017**, *3*, 105–120.
8. Rodríguez-Hernández, C.F.; Musso, M.; Kyndt, E.; Cascallar, E. Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. *Comput. Educ. Artif. Intell.* **2021**, *2*, 100018.
9. Hassan, H.; Anuar, S.; Ahmad, N.B.; Selamat, A. Improve student performance prediction using ensemble model for higher education. In *Frontiers in Artificial Intelligence and Applications*; IOS Press: Amsterdam, The Netherlands, 2019; Volume 318, pp. 217–230.
10. Mohanarajesh, Kommineni (2024). Develop New Techniques for Ensuring Fairness in Artificial Intelligence and ML Models to Promote Ethical and Unbiased Decision-Making. *International Journal of Innovations in Applied Sciences and Engineering* *10* (1):47-59.
11. Bolón-Canedo, V.; Alonso-Betanzos, A. Ensembles for feature selection: A review and future trends. *Inf. Fusion* **2018**, *52*, 1–12.
12. Meghji, A.F.; Mahoto, N.A.; Unar, M.A.; Shaikh, M.A. The role of knowledge management and data mining in improving educational practices and the learning infrastructure. *Mehran Univ. Res. J. Eng. Technol.* **2020**, *39*, 310–323.
13. Crivei, L.; Czibula, G.; Ciubotariu, G.; Dindelegan, M. Unsupervised learning-based mining of academic data sets for students' performance analysis. In *Proceedings of the SACI 2020—IEEE 14th International Symposium on Applied Computational Intelligence and Informatics*, Proceedings, Timisoara, Romania, 21–23 May 2020; Volume 17, pp. 11–16.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details