A Comparative Study on Student Academic Performance Prediction using AI for Estimating Student's Success Rate

Kajal Mahawar[1]

[1]Assistant Professor (Computer Science Department),
[1]St. Aloysius' College, Jabalpur (M.P.)

Abstract

In artificial intelligence and educational data mining, student academic performance prediction is a challenging concern.Therefore, early detection and prevention measures are important in educational institutions. Several studies have been proposed by researchers in the fields of education and artificial intelligence, in which it is challenging to predict the academic performance of the students. Also, it is a major challenge for researchers to identify which machine learning techniques are very accurate in determining a student's academic success. To solve this challenge, in our study, we explored surveys and studies in this area. These surveys and studies from different countries mainly focus on trying to identify effective predictive models of students' academic performance prediction. In this study, we try finding answers to study concerns such as what techniques are available to determine the correct sample size and what sampling techniques are accessible in machine learning. This paper enclosed the critical studies based on student academic performance prediction and techniques available for determining accurate sample size and sampling process from 2018 till 2022 with research objectives, research gaps, results, and discussion on findings, and future research recommendations.

Keywords: artificial intelligence,student academic performance prediction, machine learning, sample size, sampling.

1   Introduction

In educational data mining, for developing or evaluating models, choosing effective parameters are a difficult task. According to (Nagy et al., 2013), the number of information that was generated in the college data system had expanded along with the large influx of colleges that have embraced information systems in current history. Intuitively, EDM was used to unearth facts concealed in the information that raised the standard of the entire educational system. To automate judgment for trainers and administration, EDM can be used to identify trends in unreliable data. Researchers provided a framework for students that assemble an expert system through machine learning classifier and cluster techniques. To lower the high percentage of underachievement among the students', their approach can be used to offer advice to the first-year college learners about the course subject students' select. They used real-world data (the year 2000 to 2012) for a case study.

EDM's purpose is to improve educational results and acquire an understanding of educational phenomena by better identifying the settings in which they learn. To analysisstudents performance, a range of machine learning-related algorithms have been developed (Guo et al., 2015). Scholastic analytics, teaching analytics, and educational data mining are linked and connected to suit the demands of students, teachers, administration, financiers, school officials, and community representatives, according to the education enterprise-wide viewpoint.

Figure 1 below outlines the different tiers of assistance provided by educational data mining as well as the user segments and analysis that are most attained. The five tiers of educational data mining called, course level (L1), department level (L2),

institution level (L3), region level (L4), and national and international level (L5) with six types of user attained termed, learners, instructors, administrations, financiers, education authorities and government(Khanna et al., 2016).
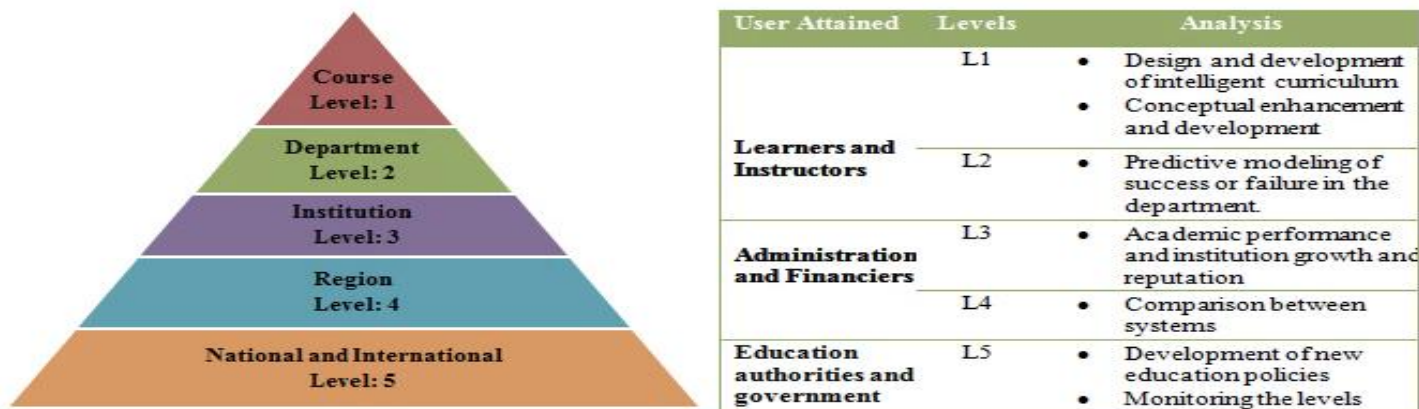

Figure 1: Educational Data Mining Levels with User Attained and Analysis

1.1   PROBLEM STATEMENT

In this study, we have focused on the major difficultiesof researchers to perform students' academic performance prediction research. To perform student's academic performance prediction research,identifying sample size techniquesand sampling techniquesis a significant challenge for researchers. To solve this problem, we conducted a methodical literature review from the year 2018-2022 to identify the effective techniques of machine learning. From the methodological overview, the following research concerns were answered:

1.1.1   What techniques are available to determine the correct sample size?

1.1.2   What sampling techniques are accessible in machine learning?

1.2   RESEARCH MOTIVATION

In student's academic performance prediction research, it is a critical task to find answers to research questions such as which sample size techniqueand sampling techniques are effective to determine a student's academic success. The motivation of this research is based on these research concerns.

1.3   RESEARCH OBJECTIVES

Followed by these research concerns, this paper conducted a methodical literature review from the year 2018-2022 research papers to identify the effective studies related to student academic performance prediction by usingmachine learning techniques .To conduct this survey, we devised taxonomy of research directions and classified the relevant publications accordingly. The outline of the review paper is as follows: Section 2presented arelated literature review with a comparative study table; Section 3discusses the materials and methods used, Section 4 presented the results and discussionon findings, Section 5 presented the research gaps, Section 6highlighted theresearch recommendations. At last, in Section 7, the conclusion is presented.

2   REVIEW OF LITERATURE

This study aims to address the research issues that have been raised and to identify the research gaps with a focus on "student performance analysis" and "student academic performance prediction" related literature from the year 2018-2022.

2.1 RESEARCH APPROACH

In this paper, figure 2 presents the research approach used to conduct this study. A well-planned search strategy is critical for discovering every relevant piece of work in a literature review.
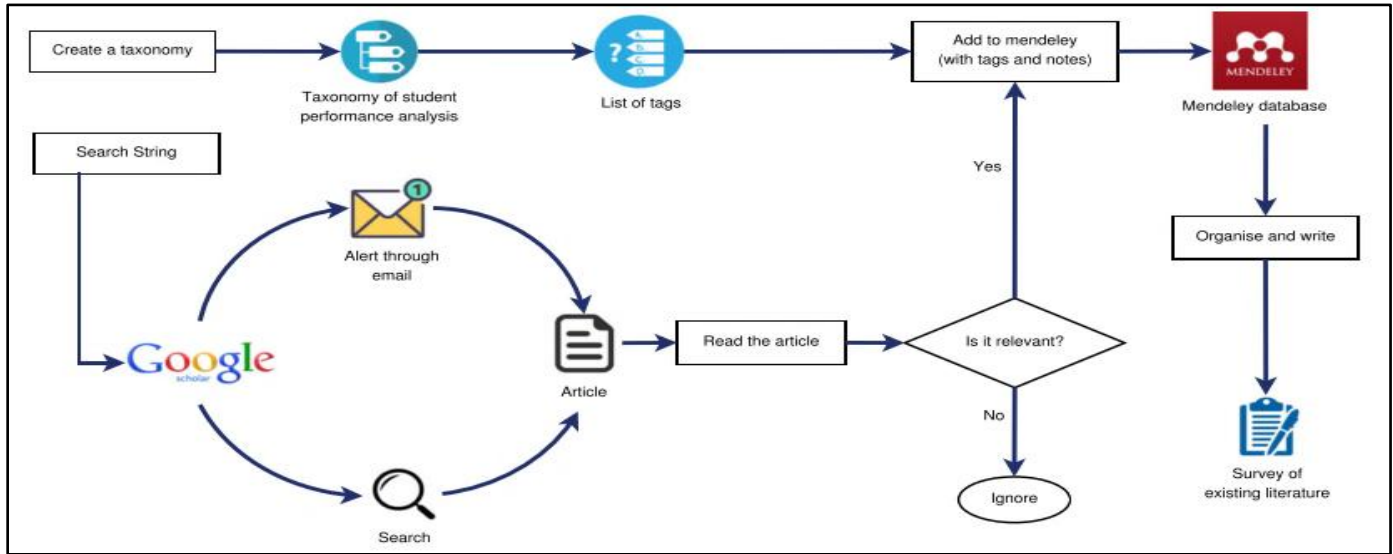

Figure 2: Methodology Used To Conduct the Study(Khan & Ghosh, 2021)

As a result, as part of the survey strategy, we first constructed taxonomy of study directions. After defining the taxonomy, we used a hybrid strategy to find relevant literature. Using the strings "student academic performance prediction", "student academic performance parameters", "sample size", "sampling",and "machine learning" in Google Scholar, we find papers. This method allows us to find the studies after January 2018. As a first step, we looked at the titles of studies to see if they were about "student performance prediction" or "student academic performance parameters". We were able to uncover 24 different studies connected to our survey using this method. We scan the abstract and keywords in each one for additionalscreening, and it filters out 5 papers that aren't relevant. Following that, we reviewed the complete text and attempted to determine whether the study's contribution is critical to the survey's research goal. This procedure aids us in weeding out research that is unrelated to academic performance prediction or parameters for students. Due to these criteria, 5 studies fail the eligibility test, leaving only 19 eligible publications for this review. The nominated papers are added to Elsevier's Mendeley software as the next step in the qualitative synthesis. Finally, we organize and prepare a literature review of pertinent papers (Khan & Ghosh, 2021).

2.2 State-of-the-art Studies

This section is divided into two sub-sections, major contribution research publications and other contribution research publications.

2.2.1 Major Contribution Research Publications

➢ **Research Studies Proposed In 2018**

(Dharmawan et al., 2018)study looked at non-academic elements that can affect student graduation and employed classification methods such as Decision Trees, K-Nearest Neighbor, and Support Vector Machine to determine the best accuracy model for their study. Researchers collected 103 samples using a basic random sampling procedure. (Ciolacu et al., 2018)focus on predicting students' final grades before they take the final exam. Researchers developed an Early Recognition System that would be based on real data collected throughout a blended learning session.

(Nieto et al., 2018)employs machine learning methods to display survey findings that focus on which academic decisions they are concern about and the variables involved in them. In a real-life case study, researchers forecasted graduation rates to aid decision-making. Support vector machine and artificial neural networktechniques are contrasted and studied to see which approach suits better prediction with KNIME.

➢ Research Studies Proposed In 2019

(Akçapınar et al., 2019)investigated two questions: which algorithms and features best predict students' end-of-term academic performance by comparing various classification algorithms and pre-processing techniques, and whether academic performance can be predicted in earlier weeks using these features and the chosen algorithm. The goal of the (Ramaswami et al., 2019)article was to assess educational data mining approaches to improve the predicted accuracy of student academic performance in a university course context. The inquiry was aided by data acquired in real-time and over self-paced activities. In (Imran et al., 2019)researchers developed a supervised learning decision tree classifier-based student performance prediction model and used the ensemble method to increase the classifier's performance. Researchers used logistic regression, SVM, and Artificial Neural Network (ANN) machine learning techniques for the study. The result shows that the proposed model achieved a classification accuracy of 84% to 93%, logistic regression achieved an accuracy rate of 79.82% to 85.60% and the support vector machine achieved an accuracy rate of 79.95% to 89.14%. The limitation of the study was that the class imbalance problem was not considered (Waheed et al., 2019).

➢ Research Studies Proposed In 2020 and 2021

(Helal et al., 2020)looked at data from students' demographic and academic information upon enrolment, as well as course assessment and participation data from Moodle, to see what characteristics influence student performance. Researchers collected three types of data from first-year domestic UG students at Australian universities: enrolment data, Moodle data, and composite data encompassing both enrolment and Moodle features. Researchers employed the SNS (SNS software tool), DSSD (DSSD software tool), NMEEFSD (KEEL data mining tool), BSD (Vikamine), SD-Map (Vikamine), and APRIORI-SD subgroup discovery methods (Orange data mining tool).

(Mahdy, 2020)studied 1392 students from 92 different nations. The researcher used the questionnaires method with a spreadsheet and an online Google form. Of 1392 students, 718 were females and 674 were males, and the age of students was between 18 to 52 years. In research, about 52.7% of the students were in the age group of 18 to 22 years, 38.5% were in the age group of 23 to 32 years, 5.7% were in the age group of 33 to 42 years, and 3.1% were in the age group of 43 to 52 years. The majorities (80.7%) of the students were undergraduate students and about 19.3% were post-graduate students.

(Alsayed et al., 2021)investigated and tested various machine learning algorithms for predicting students' right undergraduate course before admission at the undergraduate level based on current job markets and experience, and based on the findings, researchers recommended intelligent recommender systems implementation in the future. For this, 216 students from CMS Business School MBA students were chosen from the Kaggle website in January 2020.

Research paper   ,

➢ Research Studies Proposed In 2022

(Solano et al., 2022)paper goal was to evaluate several machine learning models using the CRISP-DM methodology to determine the best model for predicting high school student's performance in the test in the Colombian Caribbean region, using its metrics, while also proposing a new methodology for evaluating the test results by regions, taking into account the socio-economic differences between them.

Another research was conducted to create a dataset that provides the demographic and academic records of college students. The 168 female student's data was taken from first year (first semester and second semester) computer science department students (2020-2021) of Imam Abdulrahman University (IAU), Saudi Arabia. Another aim of the study was to develop a student's prediction model for students' academic performance prediction and then compared it with state of art models. The study also helps in the identification of the important parameters that can be used in performance prediction. At last, researchers develop a web page to implement the proposed model which generates a prediction report (Alboaneen et al., 2022).

The comparative studies of each paper with factors are presented in Table 1.

Table 1: A Comparative Study of Major Contribution Research Papers

References	Source	Sample Size	Target Group	Data collection	Country	Time-period
YEAR 2018						
(Dharmawan et al., 2018)	IEEE	103	College	Several Universities of Indonesia	Indonesia	Not Specify
(Ciolacu et al., 2018)	IEEE	285	College	Deggendorf Institute of Technology (DIT)	Germany	2016-2017, 2017-2018
(Nieto et al., 2018)	Scopus	12477	College	District University Francisco Jose de Caldas	Colombia	2004-2014
YEAR 2019						
(Akçapınar et al., 2019)	Scopus	72	College	State university in Turkey	Turkey	14 weeks
(Ramaswami et al., 2019)	Scopus	240	College	Xorro- Q	New Zealand	2016-2017
(Imran et al., 2019)	Scopus	1044	School	Two schools in Portugal Alentejo region.	Portugal	2005-2006
(Waheed et al., 2019)	Scopus	32,593	College	OULA	UK	2014-2015
YEAR 2020						
(Helal et al., 2020)	Scopus	20 features	College	Australia university	Australia	Not specify
(Mahdy, 2020)	Scopus	1,392	College	Online form	92 countries	April 13th - August 5th , 2020
YEAR 2021						
(Alsayed et al., 2021)	Scopus	216	College	Kaggle	Malaysia	January, 2020
YEAR 2022						
(Solano et al., 2022)	Scopus	Not specify	School	Colombian Institute for the Evaluation of Higher Education (ICFES)	Colombia	2017-2019
(Alboaneen et al., 2022)	Scopus	168	College	Computer Science Department of Imam Abdulrahman bin Faisal University (IAU)	Saudi Arabia	2020-2021

2.2.2   Other Contribution Research Publications

In this section, sample size and samplingtechniques are discussed.

➢ *Sample Size Techniques in Prediction Research Studies*

- ### Research Studies Proposed In 2019 and 2022

(Balki et al., 2019)review paper objective was to present a comprehensive evaluation of the most recent sample-size estimation techniques in machine learning to make suggestions for further research in the area. Researchers performed a thorough review of the literature. An article published between 1946 and 2018 was included in the search. (Kokol et al.,

2022)goal of the paper was to provide an explanation of the small dataset issue in machine learning and how it is resolved. To respond to this query synthesis knowledge was applied. According to their results, random forest, decision tree, support vector machine, transfer learning, and convolutional neural network are machine learning algorithms that are frequently employed on small datasets. To overcome the issues with limited datasets, data preprocessing techniques such as linear and nonlinear principal component analysis, discriminant analysis, data augmentation, virtual sample, feature extraction, and auto encoder are primarily used. Bioinformatics, image analysis and classification, fault diagnosing, prediction, social media analysis, and healthcare are the far more severely impacted regions. Their investigation also showed that limited datasets may be the outcome of the technology needed to accommodate gadgets like the Raspberry PI's low processor speed or minimal memory space. (Baeza-delgado et al., 2022)research objective was to determine a workable solution to sample size estimate for medical predictive models; it was the objective to evaluate the methodologies currently in use. To determine the sample size needed to create forecasting analytics and evaluate sample size fluctuation as a function of various characteristics, three distinct techniques were used. The sample size for the prediction model was subsequently calculated as well by the researchers. As a result, the sample size can be decreased by fewer variable predictors and by adding specific measures of the desired outcome.

Figure 3 depicts the sample size techniques used by the state-of-the-art research studies.
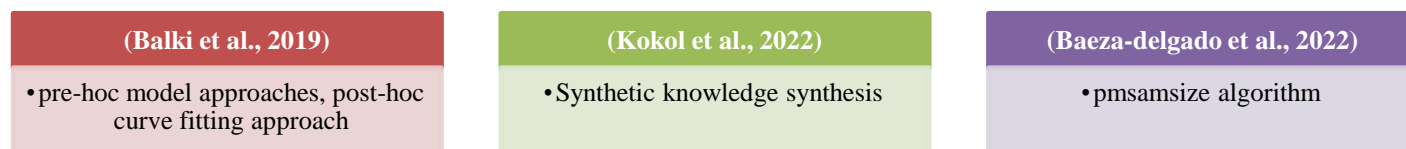
(Balki et al., 2019)	(Kokol et al., 2022)	(Baeza-delgado et al., 2022)
• pre-hoc model approaches, post-hoc curve fitting approach	• Synthetic knowledge synthesis	• pmsamsize algorithm

Figure 3: Sample Size Techniques with Reference Articles

➤ *Sampling Techniques in Prediction Research Studies*

The following sampling strategies are employed in cutting-edge research investigations and are divided into fourcategories: oversampling, under-sampling, hybrid sampling and deep learning sampling techniques.

• **Research Studies Proposed In 2018, 2020 and 2021**

(Abdulazeez & Abdulwahab, 2018)goal of the study was to create an ensembles stacked classifiers machine learning system that forecasts students'' educational performance. The reliability and effectiveness of system used to likely to occur' educational success can be increased by using stacked classification - based approaches.

In order to address the issue of class imbalance when projecting student achievement with two separate data, (Ghorbani & Ghousi, 2020)research compares alternative resampling strategies. This study used a variety of machine learning classification models to better assess how well resampling strategies solve the imbalanced situation.

In the(Hassan et al., 2020)study, three types of sampling were used; oversampling, under-sampling, and hybrid approaches. Seven sampling strategies and five different ensemble classification types are analyzed in the study. According to the researcher's observation, ROS with Adaboost performed best when compared to the other benchmarking methodologies. When used with classification models, the SMOTE-ENN approach consistently yields amazing outcomes. In contrast, SMOTE-ENN consistently outperforms all other sampling strategies. Sequential Conditional Generative Adversarial Network (SC-GAN), which captures every participant's prior behavior for its prior sequencing and created artificial information about participants for the following time, is a deep learning method(Waheed et al., 2021)employed in their study.
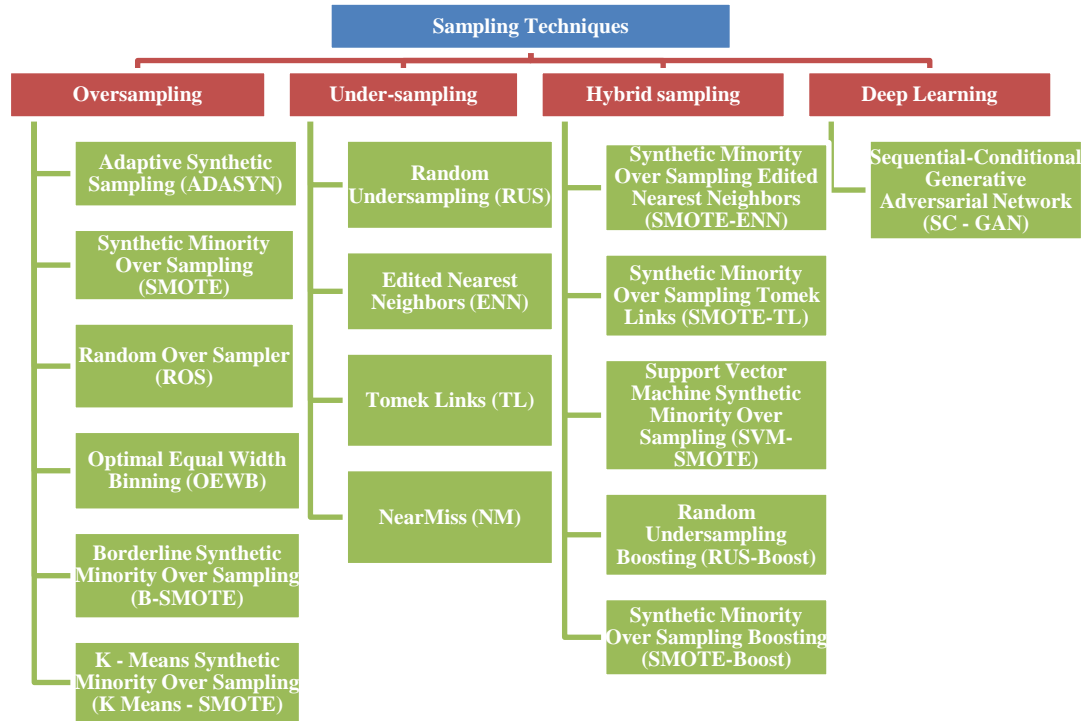
Figure 4: Sampling Techniques

Figure 5 depicts the sampling techniques used by the state-of-the-art research studies.

Figure 5: Sampling Techniques with Reference Articles

3   MATERIALS AND METHODS

In this section, we have discussed the machine learning techniques used by theresearchers from state-of-the-art research publications.Figure 6 depicts the categorization of materials and methods.

Figure 6: Materials and Methods Categorization

3.1   *Machine Learning Techniques*

The researchers have applied a variety of machine learning classification and regression techniques, according to the most recent literature. The techniques are employed for their low complexity and high accuracy rate in comparison to others. According to (Alsariera et al., 2022), the majority of researchersemployed ANN followed by SVM and LR.

This section examines research publications to determine the methods that researchers most frequently employ. Figure 7 depicts the percentage of state-of- the-art publicationsthat used various machine learningtechniques. Ten machine learning techniques such as, support vector machine (SVM) 16%, logistic regression (LR) 16%, random forest (RF) 9%, naïve Bayes (NB) 10%, k- nearest neighbors (KNN) 13%, iterative dichotomiser 3(ID3) 5%, artificial neural network (ANN) 14%, Java 48 (J48) 8%, C4.5 6% and PART 3% are most of the researcher's choice. It is evident from the study that the majority of studies utilize SVM and LR to determine the accuracy of predictions of students' academic achievement and also their average accuracy rate is higher than other techniques, as shown in Table 1.
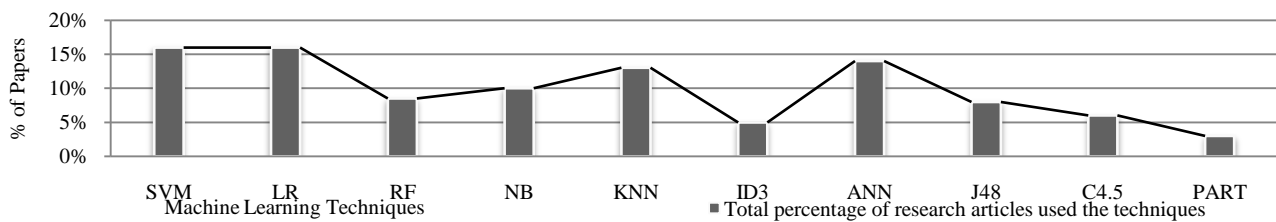

Figure 7: Comparison ofMachine Learning Techniques

4   RESULTS AND DISCUSSION

In this paper, the study enclosed the very important research studies based on student academic performance prediction, sample size and sampling techniques used for student academic performance predictions. Additional conclusions from this study include:

4.1   *RELEVANT LITERATURE*: The following graph shows the percentage of relevant literature.
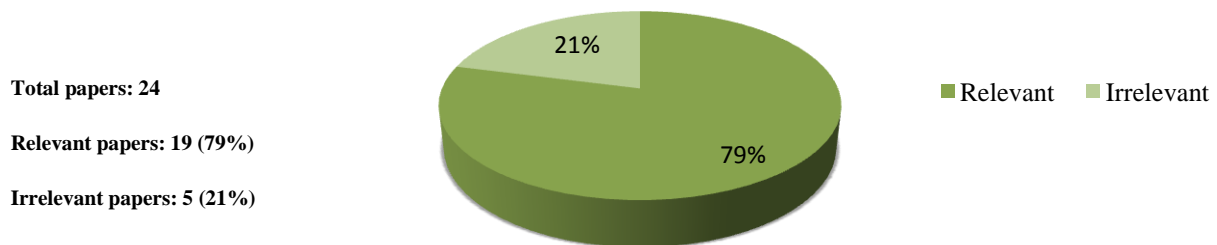
Total papers: 24

Relevant papers: 19 (79%)

Irrelevant papers: 5 (21%)

Figure 2: % Relevant Papers

4.2   *RELATIONSHIP BETWEEN THE ARTICLES*: The following figure shows that the research objective was common in both of the categories. Their research objective was the same to analyze and predicts the students' academic performances and provide a machine learning model with the highest accuracy rates.
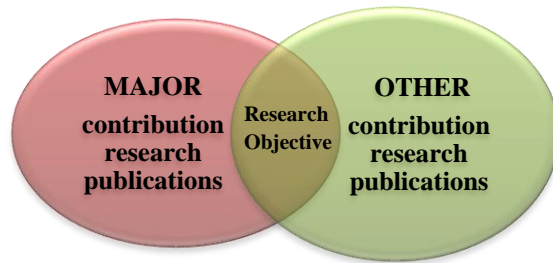
Figure 3: Relationship between the Articles

4.3   *TARGET GROUP*: The target group that the researchers chose is depicted in the following graph. Instead of school students, most researchers focused on college students.
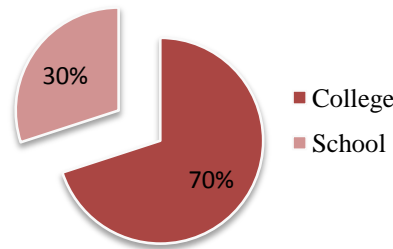

Figure 40: Major research publications based on a target group

4.4   *SAMPLING TECHNIQUES USED IN THE STATE OF THE ART PUBLICATIONS*: The number of sampling technique research articles using multiple sampling techniques is depicted in the graph below.
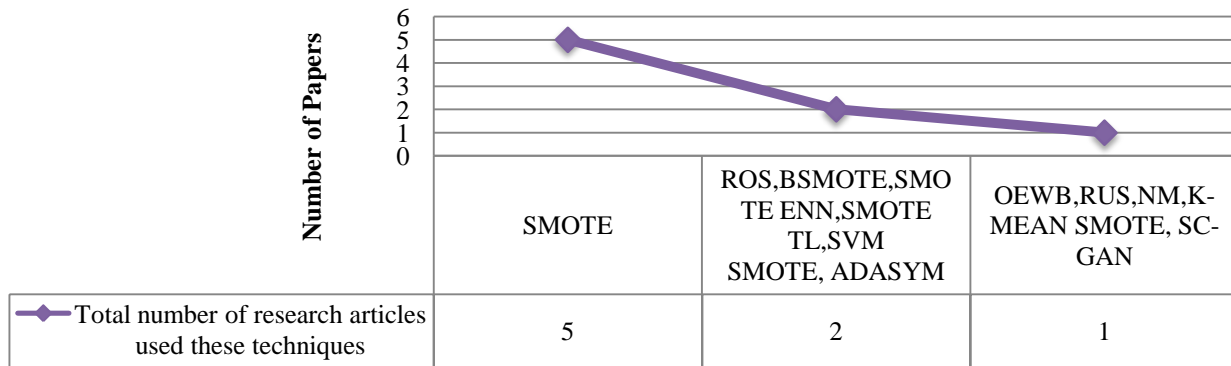

	SMOTE	ROS,BSMOTE,SMOTE ENN,SMOTE TL,SVM SMOTE, ADASYM	OEWB,RUS,NM,K-MEAN SMOTE, SC-GAN
Total number of research articles used these techniques	5	2	1

Figure 11: Research Papers Based On the Sampling Techniques

4.5NUMBER OF MAJOR RESEARCH ARTICLES BASED ON THE COUNTRIES: Researchers from different countries such as India, UK, Saudi Arabia, and Indonesia, the US, Colombia,Spain, Turkey, New Zealand, Pakistan, Southern Sweden, Korea, Australia, Ethiopia, Taiwan, Portugal, Germany, and Malaysia had conducted studies in presented research concerns. The number of the major contributed research articles, segmented by country is shown in the graph below.
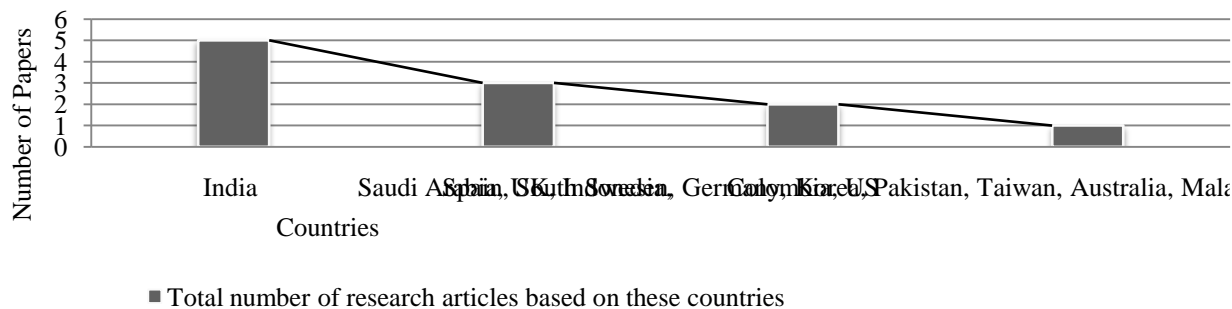
Figure 12: Research Articles Based On the Countries

5   RESEARCH GAPS

Several restrictions must be acknowledged following an assessment of the literature.

➢ In (Sugiyarti et al., 2018)study**,** the authorsworked on scholarship features in students' academic performance prediction. Researchers focused on problems students faced in getting the scholarship. Inthisstudy**,** scholarship considerations features such as religion, gender, single girl child, area they lived (urban or rural), category (EWS, UR, SC, ST, and OBC), etc. was not considered.

➢ In (Mahdy, 2020) academic prediction paper, the authors worked on online classes parameter during the Covid-19 lockdown. In this research, researchers used covid-19 and online classes feature but the active hour rate on social media platforms during the online classes was not considered.

➢ In (Middleton et al., 2022)student-faculty interaction study,the authors used before COVID-19 pandemic data to test faculty-to-student and faculty-to-faculty interactions. Researchers' study findings may affect because of the pandemic.

➢ In (Alboaneen et al., 2022) paper, the authors proposed an academic prediction model and concentrated to predict the academic performance of female students. In this study, the dataset contains unmarried female student records only. Other students such as married female students, married male students, and single male students were not considered.

➢ The size of the dataset considered by many researchers is varying in nature.

6   RECOMMENDATIONS

➢ Some more important features such as the number of hours spent playing video games, child life trauma, transportation facility, career awareness, and course satisfaction should be used to predict the student's academic performance.

➢ (Zaffar et al., 2021)hybrid feature selection technique can be used to extract the important features from a large set of features. The choice of this framework not only improves prediction accuracy but also aids in the development of strategic strategies for improving students' academic achievement.

➢ To generate prediction models, additional regression and classification algorithms should be explored in the future.

➢ We suggested researchers incorporate the hybrid approach to obtain the most accurate results.

➢ The size of the dataset should follow a definite approach.

7 CONCLUSION

Predicting students' academic success is one of the difficult study subjects in educational data mining. Experts in the field of educational data mining have put out several studies, in which it is difficult to predict the student's academic achievement. Additionally, it is quite difficult for researchers to pinpoint which machine learning methods are extremely precise in determining a student's academic success. This paper enclosed the very important research studies based on student academic performance prediction and techniques available for determining accurate sample size and sampling process. In this paper, we addressed two research questions based on sample size and sampling process in students' academic prediction. This study also uncovered several obscure features in recommendations that previous studies have overlooked. Researchers will receive a brief overview of several crucial and obscure features and methods that can be applied in this field from this work. To provide meaningful study results, this paper advises researchers to adopt the proposed recommendations in future.

Research paper　　　© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

References

1. Abdulazeez, Y., & Abdulwahab, L. (2018). Application of classification models to predict students' academic performance using classifiers ensemble and synthetic minority over sampling techniques. *Bayero Journal of Pure and Applied Sciences*, *11*(2), 142. https://doi.org/10.4314/bajopas.v11i2.17
2. Akçapınar, G., Altun, A., & Petek, A. (2019). Using learning analytics to develop early- warning system for at-risk students. *International Journal of Educational Technology in Higher Education*, 1–20.
3. Alboaneen, D., Almelihi, M., Alsubaie, R., Alghamdi, R., Alsehri, L., & Alharthi, R. (2022). Development of a Web-Based Prediction System for Students' Academic Performance. *Data*, *7*(21), 1–19.
4. Alsariera, Y. A., Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., & Ali, N. (2022). Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance. *Computational Intelligence and Neuroscience*, *2022*, 1–11. https://doi.org/10.1155/2022/4151487
5. Alsayed, A. O., Shafry, M., Rahim, M., Albidewi, I., Hussain, M., Jabeen, S. H., Alromema, N., Hussain, S., & Jibril, M. L. (2021). Selection of the Right Undergraduate Major by Students Using Supervised Learning Techniques. *Applied Sciences*, *11*.
6. Baeza-delgado, C., Alberich, L. C., Carot-sierra, J. M., Veiga-canuto, D., Martínez, B., Heras, D., Raza, B., & Martí-bonmatí, L. (2022). A practical solution to estimate the sample size required for clinical prediction models generated from observational research on data. *European Radiology Experimental*, *6*(22), 1–10.
7. Balki, I., Amirabadi, A., Levman, J., Martel, A. L., Emersic, Z., Meden, B., Garcia-Pedrero, A., Ramirez, S. C., Kong, D., Moody, A. R., & Tyrrell, P. N. (2019). Sample-Size Determination Methodologies for Machine Learning in Medical Imaging Research: A Systematic Review. *Canadian Association of Radiologists Journal*, *70*(4), 344–353. https://doi.org/10.1016/j.carj.2019.06.002
8. Ciolacu, M., Tehrani, A. F., Binder, L., & Svasta, P. M. (2018). Education 4 . 0 - Artificial Intelligence assisted Higher Education : Early recognition System with Machine Learning to support Students ' Success. *2018 IEEE 24th International Symposium for Design and Technology in Electronic Packaging (SIITME)*, *Ml*, 23–30.
9. Dharmawan, T., Ginardi, H., & Munif, A. (2018). Dropout Detection Using Non-Academic Data. *4TH Conference on Science and Technology*.
10. Ghorbani, R., & Ghousi, R. (2020). Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques. *IEEE Access*, *8*, 67899–67911. https://doi.org/10.1109/ACCESS.2020.2986809
11. Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L. (2015). Predicting Students Performance in Educational Data Mining. *Proceedings - 2015 International Symposium on Educational Technology, ISET 2015*, 125–128. https://doi.org/10.1109/ISET.2015.33
12. Hassan, H., Ahmad, N. B., & Anuar, S. (2020). Improved students' performance prediction for multi-class imbalanced problems using hybrid and ensemble approach in educational data mining. *Journal of Physics: Conference Series*, *1529*(5). https://doi.org/10.1088/1742-6596/1529/5/052041
13. Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., Murray, D. J., Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., & Murray, D. J. (2020). Identifying key factors of student academic performance by subgroup. *International Journal of Data Science and Analytics*, *7*(April), 227–245.
14. Imran, M., Latif, S., Mehmood, D., & Shah, M. S. (2019). Student academic performance prediction using supervised learning techniques. *International Journal of Emerging Technologies in Learning*, *14*(14), 92–104. https://doi.org/10.3991/ijet.v14i14.10310
15. Khan, A., & Ghosh, S. K. (2021). Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *Education and Information Technologies*, *26*(1), 205–240. https://doi.org/10.1007/s10639-020-10230-3
16. Khanna, L., Singh, S. N., & Alam, M. (2016). Educational data mining and its role in determining factors affecting students academic performance: A systematic review. *India International Conference on Information Processing, IICIP 2016 - Proceedings*. https://doi.org/10.1109/IICIP.2016.7975354
17. Kokol, P., Kokol, M., & Zagoranski, S. (2022). Machine learning on small size samples: A synthetic knowledge synthesis. *Science Progress*, *105*(1), 1–16. https://doi.org/10.1177/00368504211029777
18. Mahdy, M. A. A. (2020). The Impact of COVID-19 Pandemic on the Academic Performance of Veterinary Medical Students. *Frontiers in Veterinary Science*, *7*(October), 1–8. https://doi.org/10.3389/fvets.2020.594261
19. Nagy, H. M., Aly, W. M., & Hegazy, O. F. (2013). An Educational Data Mining System for Advising Higher Education Students. *International Journal of Computer, Electrical, Automation, Control and Information*

Engineering, *7*(10), 1266–1270.

20. Nieto, Y., Carlos, V. G., Rubén, M., & Crespo, G. (2018). Supporting academic decision making at higher educational institutions using machine learning-based algorithms. *Soft Computing*. https://doi.org/10.1007/s00500-018-3064-6

21. Ramaswami, G., Susnjak, T., Mathrani, A., Lim, J., & Garcia, P. (2019). Using educational data mining techniques to increase the prediction accuracy of student academic performance. *Information and Learning Science*, *120*(7-8), 451–467. https://doi.org/10.1108/ILS-03-2019-0017

22. Solano, J. A., Lancheros Cuesta, D. J., Umaña Ibáñez, S. F., & Coronado-Hernández, J. R. (2022). Predictive models assessment based on CRISP-DM methodology for students performance in Colombia - Saber 11 Test. *Procedia Computer Science*, *198*(2020), 512–517. https://doi.org/10.1016/j.procs.2021.12.278

23. Sugiyarti, E., Jasmi, K. A., Basiron, B., & Huda, M. (2018). DECISION SUPPORT SYSTEM OF SCHOLARSHIP GRANTEE SELECTION USING DATA MINING. *International Journal of Pure and Applied Mathematics*, *119*(15), 2239–2249.

24. Waheed, H., Anas, M., Hassan, S. U., Aljohani, N. R., Alelyani, S., Edifor, E. E., & Nawaz, R. (2021). Balancing sequential data to predict students at-risk using adversarial networks. *Computers and Electrical Engineering*, *93*(September 2020), 107274. https://doi.org/10.1016/j.compeleceng.2021.107274

25. Waheed, H., Hassan, S.-U., Aljohani, N. R., Hardman, J., & Nawaz, R. (2019). Predicting Academic Performance of Students from VLE Big Data using Deep Learning Models. *Computers in Human Behavior*, 106189.

26. Zaffar, M., Hashmani, M. A., Habib, R., Quraishi, K. S., Irfan, M., Alqhtani, S., & Hamdi, M. (2021). A hybrid feature selection framework for predicting students performance. *Computers, Materials and Continua*, *70*(1), 1893–1920. https://doi.org/10.32604/cmc.2022.018295