Student performance prediction in higher education: A comprehensive review

Conference Paper *in* AIP Conference Proceedings · April 2022

DOI: 10.1063/5.0080187

3 authors:

Ellysa Tjandra
Universitas Surabaya

Sri Suning Kusumawardani
Universitas Gadjah Mada

Ridi Ferdiana
Universitas Gadjah Mada

Student performance prediction in higher education: A comprehensive review

Cite as: AIP Conference Proceedings **2470**, 050005 (2022); https://doi.org/10.1063/5.0080187
Published Online: 25 April 2022

Ellysa Tjandra, Sri Suning Kusumawardani and Ridi Ferdiana

Student Performance Prediction in Higher Education: A Comprehensive Review

Ellysa Tjandra[1, 2, a)], Sri Suning Kusumawardani[2, b)], and Ridi Ferdiana[2, c)]

[1] *Department of Informatics Engineering, University of Surabaya, Surabaya 60293, Indonesia*
[2]*Department of Electrical and Information Engineering, Gadjah Mada University, Yogyakarta, Indonesia*

[a)] *Corresponding author: ellysa@staff.ubaya.ac.id*
[b)] *suning@ugm.ac.id*
[c)] *ridi@ugm.ac.id*

Abstract. Student dropout still becomes a critical problem in education. Educational Data Mining (EDM) can bring potential impact to support academic institution's goals in making academic decisions, such as regulation renewal, rule enforcement, or academic process improvement. The sooner at-risk students can be identified, the earlier institution members can provide necessary treatments, thus prevent them from dropout and increase the student retention rate. This study performs a comprehensive literature review of student performance prediction using EDM techniques, including various research from 2002 to 2021. Our study is aimed to provide a comprehensive review of recent studies based on student performance prediction tasks, predictor variables, methods, accuracy, and tools used in previous works of student performance prediction. Performing student performance prediction in an academic institution can be helpful to provide the student performance mitigation mechanism because it can be managed earlier by the management to decrease the student dropout rate.

Keywords: student performance, prediction, student dropout, Educational Data Mining, EDM review

INTRODUCTION

Student dropout and retention is a prominent issue in education at present. When the dropout rate increases, it means the institution loses the number of students, or the student retention rate decreases. According to the Indonesian academic minister of higher education (Kemenristekdikti), in 2017, there were 195.176 of 6.924.511 (28.2%) students in higher education quitted from school [1]. Various procedures or processes can be conducted to prevent student dropout: student performance monitoring, academic rule enforcement, or academic improvement. To perform these activities, higher education institution needs supporting system, which can be established by Educational Data Mining (EDM).

Educational Data Mining (EDM) is still being considered as a popular solution in education. EDM techniques provide potential impact for supporting academic institution goals to improve the quality and efficiency of learning activities and monitoring processes. Aldowah et al. (2019) summarized the previous works in the EDM field, specifically in computer-based student performance prediction, into three main objectives: evaluating learning materials (course contents, syllabus, etc.), monitoring learning activities/results (delivery methods, assignments, scoring, etc.), and preventing student dropout (performance measurement, early warning, survival index, etc.). They also categorized EDM into four domains based on its objective: Learning Analytics (LA), Predictive Analytics (PA), Behavioral Analytics (BA), and Visualization Analytics (VA) [2]. In their review, it could be concluded that predictive analytics still has the highest demand, which implies that predictive analytics is still being the most challenging subject in the EDM field.

There were plenty of EDM techniques had been conducted by many researchers [3]–[8]. The previous literature reviews mainly concentrated on using EDM methods/techniques and student performance prediction, such as

International Conference on Informatics, Technology, and Engineering 2021 (InCITE 2021)
AIP Conf. Proc. 2470, 050005-1–050005-9; https://doi.org/10.1063/5.0080187
Published by AIP Publishing. 978-0-7354-4180-4/$30.00

classification, clustering, rule-based methods, statistics, and user interfacing or visualization, but no comprehensive review of student performance prediction had been conducted. This research is aimed to provide a comprehensive review of recent studies based on student performance prediction tasks, predictor variables, predictive variables, methods, accuracy, and tools used in previous works of student dropout prediction.

PREVIOUS WORKS

Many literature reviews of EDM had been conducted and provided further insight into EDM fields. Each review captured the different points of view of EDM works. In 2015, [6] performed SLR to find the most frequent attributes and methods used in predicting student performance. Afterward, a review of student performance factors, clustering algorithm, and EDM tools had been conducted by [3], [4], [9], [10], followed by a literature review of student retention factors [11], EDM using big data framework [5], and domains and student success factors in the first year of higher education [12]. After that, [2] provided a broad analysis of EDM domains and applications, while [13] portrayed a comprehensive review of predictors, predictive values, and EDM techniques used to predict academic performance. [14] conducted an SLR of Student Achievement Influencing Factors. In 2021, student performance SLR is still being conducted. [15] performed an SLR of Student Performance Prediction based on Outcome-Based Education using Student Outcomes and Learning Outcomes, while [16] conducted an SLR of Student Performance Prediction Model used by each educational level. Overall, about 250 studies of student performance prediction had been established from 2002 to 2021. However, no comprehensive review of student performance prediction, especially for student dropout prevention, had been established.

These works of literature are categorized into three main objectives as shown in Table 1: categorizes student performance prediction objectives into three main categories: evaluating learning materials, such as course contents, syllabus, etc. (59 or 23.60%), monitoring learning activities/results, such as delivery methods, assignments, scoring, etc. (168 or 67.2%), and preventing student dropout, such as performance measurement, early warning, survival index, etc. (23 or 9.20%). It can be concluded that dropout and retention is still the least researched sub-field in student performance prediction, while the dropout and retention problem is still the crucial problem of education that needs to be resolved immediately.

TABLE 1. Studies of Student Performance Prediction

Objectives	Number of Studies	%
Evaluating learning materials	59	23.60%
Monitoring learning activities/results	168	67.20%
Preventing student dropout	23	9.20%
Total	**250**	**100%**

METHODS

Therefore, this study proposes five research questions:

Q1: What are the student performance prediction tasks performed by the previous studies?
Q2: What are the predictor variables (attributes) used in student performance prediction?
Q3: What are the prediction methods used for students dropout prevention?
Q4: How about the results (accuracy) in each prediction method?
Q5: What are the frequently used tools to perform prediction tasks?

To narrow the result search, this review was conducted using specific search keywords and criteria(s):

Search Keywords

This review was conducted using these keywords and combination terms:

- Educational Data Mining OR Learning Analytics
- Student Performance Prediction
- Early Warning OR Early Prediction OR Early Detection OR Early Estimation
- Higher Education OR College OR Undergraduate OR University OR Degree Program
- Student Performance Index OR Student Performance Model

- Student Retention OR Persistence OR Survival
- Student Dropout OR Student Failure OR At-Risk Student OR Student Difficulties

We limit our search for any studies performed in 2002 until December 2020. Hence, any paper published after that time is not included in our review.

Search Criteria

After a strict screening of previous studies, only strong-related papers were selected in this review, as well as duplicate papers were excluded from this review, resulting in 34 papers (each paper can contain more than one research). All articles were chosen from reputable journals and conferences published by trusted publishers.

RESULTS AND DISCUSSION

At first, this review summarizes the number of previous works in the EDM field, specifically in computer-based student performance prediction.

Student Performance Tasks

Student performance prediction is categorized into five main tasks: Student Identification & Classification, Student Modeling and Enhancement, Recommendation System, Early Warning, and Survival Indexing. Table 2 obviously describes the top three frequent tasks performed by recent studies are Student Identification & Classification (44.12%), Student Retention Modeling & Enhancement (23.53%), Early Warning (14.71%), Recommendation System (11.76%), and Survival Indexing (5.88%), while Table 3 explores details of the studies.

TABLE 2. Number of Student Performance Prediction Studies

Task	Number of Studies	%
Student Identification & Classification	15	44.12%
Student Modeling & Enhancement	8	23.53%
Early Warning	5	14.71%
Recommendation System	4	11.76%
Survival Indexing	2	5.88%
Total	**34**	**100%**

As shown in Table 3, some student performance tasks were performed in higher education. EDM is often used to predict student performance, such as identifying and grouping Students (Student Identification & Classification) [17]–[31], measuring student performance for future development (Student Modeling & Enhancement) [11], [32]–[37], providing Early Warning [20], [31], [38]–[40], providing Recommendations [30], [41]–[43], and presents a student success index (Survival Indexing) [44], [45].

TABLE 3. Student Performance Prediction Tasks

Task	Study
Student Identification & Classification	[17], [18], [27]–[31], [19]–[26]
Student Modeling & Enhancement	[11], [32]–[37], [46]
Early Warning	[20], [31], [38], [40], [47], [48]
Recommendation System	[30], [41]–[43]
Survival Indexing	[44], [45]

Student Performance Attributes (Predictor Variables)

At first, student performance factors must be analyzed to perform the best prediction. We categorize student performance predictor variables into internal and external attributes (factors). Internal factor is any factor(s) that happened inside the university (or belongs to the university), while external factors come from the outside (e.g., students' family condition or other environments outside the university). Internal factors are considered to be more flexible because the university members can change them via regulation renewal, rule enhancement, or process improvement, while external factors cannot. In this research, personal factors such as behavioral, psychological, and motivational factors also are examined as well as academic factors.

Student performance factors have been already examined in previous studies, and most of them used specific academic fields [24], [29][44], [46]. Academic and personal attributes have a significant impact on student academic performance. In 2014, [49] tried to add more factors besides academic factors: gender, high school background (secondary school grade), a chosen priority of the program (first, second, or third), and the financial condition (government-financed or self-financed). However, these studies mainly focused on academic factors. Studies of student personal approaches had been emerged to provide a better representation of a student. [50] examined social behavior of the students, resulting in a significant increase in dropout prediction accuracy. [23] analyzed psychological factors related to the personality of students (called the bio-psycho-social level of development) and external factors, consist of students' socio-economic (student demographic), cultural (ethnicity), and educational environments, and stated that psychological factors also provide a significant influence in student academic engagement. These studies also stated that personal factors also have a significant impact on student academic performance.

From Table 4 we can conclude the most-widely used predictors are: student demographic (13.75%), internal assessment (12.50%), admission (8.75%) student achievement index (GPA) (6.25%), psychological factors (6.25%), and social behavior (6.25%). Internal assessment includes test scores and assignments (internal assessment), prerequisite course grades, and course engagement or course attendance.

TABLE 4. Student Performance Prediction Attributes used in Studies

Category	Attributes/Predictor	Number of Studies	%
External	Student Demographic	11	13.75%
Internal	Internal Assessment	10	12.50%
External	Admission	7	8.75%
Internal	CGPA	5	6.25%
External	Financial Condition	5	6.25%
External	Psychological Factors	5	6.25%
External	Social Behaviour	5	6.25%
Internal	Attendance & Delivery Mode	4	5.00%
External	High School Background	4	5.00%
External	Ethnicity (Nationality)	3	3.75%
External	Parent Educational Level	3	3.75%
Internal	English Proficiency	2	2.50%
External	Gender	2	2.50%
Internal	Extra-Curricular Activities	2	2.50%
Internal	Soft Skills	2	2.50%
External	Student Habit	2	2.50%
Internal	Educational Environment	1	1.25%
External	External Assessment	1	1.25%
Internal	Institute Rank	1	1.25%
External	Job Time	1	1.25%
Internal	Number of Students	1	1.25%
External	Religion	1	1.25%
Internal	Scholarship	1	1.25%
Internal	Student Status	1	1.25%
	TOTAL	**80**	**100%**

Student Performance Prediction Methods

Student Performance Prediction Methods used in previous research can be seen in Table 5. One research may perform more than one method and vice versa. The top five frequently used methods are Decision Tree (27.27%), Naïve Bayes (18.18%), K-Nearest Neighbour (12.12%), Neural Network (12.12%), and SVM (12.12%).

TABLE 5. Student Performance Prediction Methods used in Previous Studies

Methods	Number of Studies	%
Decision Tree	9	27.27%
Naïve Bayes	6	18.18%
K-Nearest	4	12.12%
Neural Network	4	12.12%
SVM	4	12.12%
Random Forest	3	9.09%
Ruled-Based	2	6.06%
Regression	1	3.03%
TOTAL	**33**	**100%**

Methods accuracy results for each prediction method used in previous works are shown in Table 6. The top five methods with best accuracy were performed using Neural Network (97.00%) [34], Random Forest (96.01%) [40], Decision Tree (92.80%) [25], Decision Tree (90%) [35], [43], and Random Forest (88.00%) [28].

TABLE 6. Accuracy of Prediction Methods

Method	Accuracy	Study	Year
Neural Network	81.00%	[41]	2002
K-Nearest	82.00%	[17]	2003
SVM	80.00%	[18]	2006
Regression	70.60%	[32]	2006
Decision Tree Naïve Bayes	92.80%	[25]	2007
Decision Tree Naïve Bayes Neural Network	73.00% 76.00% 71.00%	[26]	2008
Neural Network	97.00%	[34]	2013
Decision Tree K-Nearest Naïve Bayes SVM	66.00% 83.00% 73.00% 80.00%	[27]	2014
Decision Tree	90.00%	[35]	2014
Decision Tree	90.00%	[43]	2014
Decision Tree Random Forest	88.00%	[28]	2014
K-Nearest	70%	[45]	2014
Decision Tree Neural Network K-Nearest Ruled-Based Naïve Bayes Random Forest	69.23% 62.50% 74.04% 55.77% 83.65% 71.15%	[20]	2017
Naïve Bayes SVM (SMO) Decision Tree Ruled-Based	83.20% 81.00% 80.00% 79.00%	[51]	2018
Decision Tree Random Forest	85.75% 96.01%	[40]	2019

Method	Accuracy	Study	Year
SVM (SMO)	86.03%		
Naïve Bayes	85.51%		

Student Performance Prediction Tools

Many researchers use tools to conduct prediction analyses in EDM [1]. Based on its objective, the student performance prediction tool is categorized into three categories: Data Manipulation, Analysis of Algorithm, and Data Visualization [10]. The frequently used tools in this research are elaborated from EDM tools reviews by [10], [52]–[54]. Tools for Data Manipulation are Microsoft Excel (Google Sheets), EDM Workbench, Phyton & Jupyter Notebook, SQL, SAP HANA, eLAT, while analysis of algorithm tasks using Rapid Miner, WEKA, SPSS, KNIME, Orange, KEEL, Spark MLLib, EDAIME, MMT (Moodle Mining Tool). For Data Visualization, they use Tableau, D3.js, EPRules, GISMO, TADA-Ed, Synergo/CoIAT, PDinamet, and SNAPP. Frequently tools used by EDM analysts and researchers to perform student performance prediction can be seen in Table 7.

TABLE 7. Frequently Used Tools in Student Performance Prediction

Objective	Tools
Data Manipulation	Microsoft Excel (Google Sheets)
	EDM Workbench
	Phyton & Jupyter Notebook
	SQL
	SAP HANA
	eLAT
Analysis of Algorithm	Rapid Miner
	WEKA
	SPSS
	KNIME
	Orange
	KEEL
	Spark MLLib
	EDAIME
	MMT (Moodle Mining Tool)
Data Visualization	Tableau
	D3.js
	EPRules
	GISMO
	TADA-Ed
	Synergo/CoIAT
	PDinamet
	SNAPP

CONCLUSION

This review finds that student academic prediction played an essential role in providing opportunities and solutions to various academic institution problems, mainly student dropout mitigation. In general, most data mining techniques are well suited to perform student academic prediction. According to the review result, we suggest Random Forest, Neural Network, Decision Tree, and Naïve Bayes methods to perform student performance prediction because the techniques have high accuracy results. Furthermore, we find that academic factors are frequently used in EDM fields. There is still limited research using personal characteristics - such as psychological and social/behavioral factors - to conduct student performance predictions. Hence, it needs to be explored in the future so that the dropout rate can be decreased as well.

REFERENCES

1. I. Nirmala and A. N. R. Attamimi, *Buku Statistik Pendidikan Tinggi Tahun 2017 (Higher Education Statistical Year Book 2017)*, 1st ed., vol. PT 17, no. 1. Jakarta: Pusat Data dan Informasi Iptek Dikti, 2017.
2. 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. January, pp. 13–49, 2019, doi: 10.1016/j.tele.2019.01.007.
3. A. Dutt, M. A. Ismail, and T. Herawan, "A Systematic Review on Educational Data Mining," *IEEE Access*, vol. 5, no. c, pp. 15991–16005, 2017, doi: 10.1109/ACCESS.2017.2654247.
4. L. Khanna, S. N. Singh, and M. Alam, "Educational data mining and its role in determining factors affecting students academic performance: A systematic review," *India Int. Conf. Inf. Process. IICIP 2016 - Proc.*, 2017, doi: 10.1109/IICIP.2016.7975354.
5. S. M. Muthukrishnan, N. B. M. Yasin, and M. Govindasamy, "Big data framework for students' academic performance prediction: A systematic literature review," *ISCAIE 2018 - 2018 IEEE Symp. Comput. Appl. Ind. Electron.*, pp. 376–382, 2018, doi: 10.1109/ISCAIE.2018.8405502.
6. 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.
7. M. A. Ghazal, O. Ibrahim, and M. A. Salama, "Educational process mining: A systematic literature review," *Proc. - 2017 Eur. Conf. Electr. Eng. Comput. Sci. EECS 2017*, pp. 198–203, 2018, doi: 10.1109/EECS.2017.45.
8. A. Almasri, E. Celebi, and R. S. Alkhawaldeh, "EMT: Ensemble meta-based tree model for predicting student performance," *Sci. Program.*, vol. 2019, 2019, doi: 10.1155/2019/3610248.
9. M. A. Jayanthi, R. L. Kumar, A. Surendran, and K. Prathap, "Research contemplate on educational data mining," *2016 IEEE Int. Conf. Adv. Comput. Appl. ICACA 2016*, pp. 110–114, 2017, doi: 10.1109/ICACA.2016.7887933.
10. S. Slater, S. Joksimović, V. Kovanovic, R. S. Baker, and D. Gasevic, "Tools for Educational Data Mining: A Review," *J. Educ. Behav. Stat.*, vol. 42, no. 1, pp. 85–106, 2017, doi: 10.3102/1076998616666808.
11. C. Dewberry and D. J. R. Jackson, "An application of the theory of planned behavior to student retention," *J. Vocat. Behav.*, vol. 107, no. August 2017, pp. 100–110, 2018, doi: 10.1016/j.jvb.2018.03.005.
12. P. J. A. C. van der Zanden, E. Denessen, A. H. N. Cillessen, and P. C. Meijer, "Domains and predictors of first-year student success: A systematic review," *Educ. Res. Rev.*, vol. 23, no. December 2017, pp. 57–77, 2018, doi: 10.1016/j.edurev.2018.01.001.
13. A. Hellas *et al.*, "Predicting academic performance: a systematic literature review," pp. 175–199, 2019, doi: 10.1145/3293881.3295783.
14. M. Suhaini, "Factors Influencing Student Achievement: A Systematic Review," *Int. J. Psychosoc. Rehabil.*, vol. 24, no. 5, pp. 550–560, 2020, doi: 10.37200/ijpr/v24i5/pr201720.
15. A. Namoun and A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques: A systematic literature review," *Appl. Sci.*, vol. 11, no. 1, pp. 1–28, 2021, doi: 10.3390/app11010237.
16. R. Alamri and B. Alharbi, "Explainable Student Performance Prediction Models: A Systematic Review," *IEEE Access*, vol. 9, pp. 33132–33143, 2021, doi: 10.1109/ACCESS.2021.3061368.
17. B. Minaei-Bidgoli, D. A. Kashy, G. Kortemeyer, and W. F. Punch, "Predicting student performance: An application of data mining methods with an educational web-based system," *Proc. - Front. Educ. Conf. FIE*, vol. 1, p. T2A13-T2A18, 2003, doi: 10.1109/FIE.2003.1263284.
18. W. Hamalainen and M. Vinni, "Comparison of machine learning methods for intelligent tutoring systems," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4053 LNCS, pp. 525–534, 2006.
19. N. Iam-On and T. Boongoen, "Generating descriptive model for student dropout: a review of clustering approach," *Human-centric Comput. Inf. Sci.*, vol. 7, no. 1, pp. 1–24, 2017, doi: 10.1186/s13673-016-0083-0.
20. R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," *Comput. Educ.*, vol. 113, 2017, doi: 10.1016/j.compedu.2017.05.007.
21. 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.*, no. December 2017, pp. 0–1, 2018, doi: 10.1016/j.chb.2018.06.032.

22. A. Sandoval, C. Gonzalez, R. Alarcon, K. Pichara, and M. Montenegro, "Centralized student performance prediction in large courses based on low-cost variables in an institutional context," *Internet High. Educ.*, vol. 37, no. January, pp. 76–89, 2018, doi: 10.1016/j.iheduc.2018.02.002.
23. A. Sarra, L. Fontanella, and S. Di Zio, "Identifying Students at Risk of Academic Failure Within the Educational Data Mining Framework," *Soc. Indic. Res.*, pp. 1–20, 2018, doi: 10.1007/s11205-018-1901-8.
24. J. -P. Vandamme, N. Meskens, and J. -F. Superby, "Predicting Academic Performance by Data Mining Methods," *Educ. Econ.*, vol. 15, no. 4, pp. 405–419, 2007, doi: 10.1080/09645290701409939.
25. N. T. Nghe, P. Janecek, and P. Haddawy, "A comparative analysis of techniques for predicting academic performance," *Proc. - Front. Educ. Conf. FIE*, no. November, 2007, doi: 10.1109/FIE.2007.4417993.
26. E. Osmanbegović and M. Suljic, "Data Mining Approach for Predicting Student Performance," *Econ. Rev. J. Econ. Bus.*, vol. X, no. 1, pp. 3–12, 2012.
27. 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, 2014, doi: 10.1109/ICCIC.2014.7238346.
28. T. Mishra, D. Kumar, and S. Gupta, "Mining students' data for prediction performance," *Int. Conf. Adv. Comput. Commun. Technol. ACCT*, pp. 255–262, 2014, doi: 10.1109/ACCT.2014.105.
29. A. Pradeep and J. Thomas, "Predicting College Students Dropout using EDM Techniques," *Int. J. Comput. Appl.*, vol. 123, no. 5, pp. 26–34, 2015, doi: 10.5120/ijca2015905328.
30. M. A. Salahli, T. Gasimzade, F. Alasgarova, and A. Guliyev, "The Use of Predictive Models in Intelligent Recommendation Systems," *Procedia Comput. Sci.*, vol. 102, no. August, pp. 515–519, 2016, doi: 10.1016/j.procs.2016.09.436.
31. C. J. Villagrá-Arnedo, F. J. Gallego-Durán, F. Llorens-Largo, P. Compañ-Rosique, R. Satorre-Cuerda, and R. Molina-Carmona, "Improving the expressiveness of black-box models for predicting student performance," *Comput. Human Behav.*, vol. 72, pp. 621–631, 2017, doi: 10.1016/j.chb.2016.09.001.
32. P. Golding and O. Donaldson, "Predicting academic performance," *Proc. - Front. Educ. Conf. FIE*, pp. 21–26, 2006, doi: 10.1109/FIE.2006.322661.
33. D. Delen, "A comparative analysis of machine learning techniques for student retention management," *Decis. Support Syst.*, vol. 49, no. 4, pp. 498–506, 2010, doi: 10.1016/j.dss.2010.06.003.
34. P. M. Arsad, N. Buniyamin, and J. L. A. Manan, "A neural network students' performance prediction model (NNSPPM)," *2013 IEEE Int. Conf. Smart Instrumentation, Meas. Appl. ICSIMA 2013*, no. November, pp. 26–27, 2013, doi: 10.1109/ICSIMA.2013.6717966.
35. S. Natek and M. Zwilling, "Student data mining solution-knowledge management system related to higher education institutions," *Expert Syst. Appl.*, vol. 41, no. 14, pp. 6400–6407, 2014, doi: 10.1016/j.eswa.2014.04.024.
36. N. D. Paterson, "Predictors of first year retention rates at the university of the West Indies, Jamaica," *Int. J. Educ. Dev.*, vol. 55, no. June, pp. 63–68, 2017, doi: 10.1016/j.ijedudev.2017.06.001.
37. D. Delen, K. Topuz, and E. Eryarsoy, "Development of a Bayesian Belief Network-based DSS for predicting and understanding freshmen student attrition," *Eur. J. Oper. Res.*, no. xxxx, 2019, doi: 10.1016/j.ejor.2019.03.037.
38. H. Hamsa, S. Indiradevi, and J. J. Kizhakkethottam, "Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm," *Procedia Technol.*, vol. 25, pp. 326–332, 2016, doi: 10.1016/j.protcy.2016.08.114.
39. A. S. Hoffait and M. Schyns, "Early detection of university students with potential difficulties," *Decis. Support Syst.*, vol. 101, pp. 1–11, 2017, doi: 10.1016/j.dss.2017.05.003.
40. C. C. Gray and D. Perkins, "Utilizing early engagement and machine learning to predict student outcomes," *Comput. Educ.*, vol. 131, no. December 2018, pp. 22–32, 2019, doi: 10.1016/j.compedu.2018.12.006.
41. T. Wang and A. Mitrovic, "Using neural networks to predict student's performance," *Proc. - Int. Conf. Comput. Educ. ICCE 2002*, no. October, pp. 969–973, 2002, doi: 10.1109/CIE.2002.1186127.
42. I. Lykourentzou, I. Giannoukos, V. Nikolopoulos, G. Mpardis, and V. Loumos, "Dropout prediction in e-learning courses through the combination of machine learning techniques," *Comput. Educ.*, vol. 53, no. 3, pp. 950–965, 2009, doi: 10.1016/j.compedu.2009.05.010.
43. Elakia, Gayathri, Aarthi, and Naren J, "Application of Data Mining in Educational Database for Predicting Behavioural Patterns of the Students," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 3, pp. 4649–4652, 2014.

44. C. Bruffaerts, C. Bruffaerts, C. Dehon, and B. Guisset, "Can Schooling and Socio-Economic Level Be a Millstone to a Student ' s Academic Success ? Can schooling and socio-economic level be a millstone to a student ' s academic success ? ∗," 2011.
45. R. J. Oskouei, M. Askari, and P. R. P. Sajja, "Perceived internet usage behaviours as predictor to outlier detection in students' communities in academic environments," *Indian J. Sci. Technol.*, vol. 6, no. 7, pp. 4923–4935, 2013.
46. S. Huang and N. Fang, "Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models," *Comput. Educ.*, vol. 61, no. 1, pp. 133–145, 2013, doi: 10.1016/j.compedu.2012.08.015.
47. C. Márquez-Vera, A. Cano, C. Romero, A. Y. M. Noaman, H. Mousa Fardoun, and S. Ventura, "Early dropout prediction using data mining: A case study with high school students," *Expert Syst.*, vol. 33, no. 1, pp. 107–124, 2016, doi: 10.1111/exsy.12135.
48. A. Hoffait and M. Schyns, "Early detection of university students with potential difficulties," *Decis. Support Syst.*, vol. 101, pp. 1–11, 2017, doi: 10.1016/j.dss.2017.05.003.
49. L. Paura and I. Arhipova, "Cause Analysis of Students' Dropout Rate in Higher Education Study Program," *Procedia - Soc. Behav. Sci.*, vol. 109, pp. 1282–1286, 2014, doi: 10.1016/j.sbspro.2013.12.625.
50. J. Bayer, H. Bydzovská, and J. Géryk, "Predicting Drop-Out from Social Behaviour of Students.," *Proc. 5th Int. Conf. Educ. Data Min.*, no. Dm, pp. 103–109, 2012, [Online]. Available: http://eric.ed.gov/?id=ED537184.
51. L. Liu *et al.*, "Predicting academic performance by considering student heterogeneity," *Knowledge-Based Syst.*, vol. 161, no. December 2017, pp. 134–146, 2018, doi: 10.1016/j.knosys.2018.07.042.
52. S. Ray and M. Saeed, "Applications of educational data mining and learning analytics tools in handling big data in higher education," *Appl. Big Data Anal. Trends, Issues, Challenges*, pp. 135–160, 2018, doi: 10.1007/978-3-319-76472-6_7.
53. K. Venkatachalapathy, V. Vijayalakshmi, and V. Ohmprakash, "Educational data mining tools: A survey from 2001 to 2016," *Proc. - 2017 2nd Int. Conf. Recent Trends Challenges Comput. Model. ICRTCCM 2017*, no. 14, pp. 67–72, 2017, doi: 10.1109/ICRTCCM.2017.53.
54. S. N. Bonde and D. K. Kirange, "Educational Data Mining Survey for Predicting Student's Academic Performance," *Lect. Notes Data Eng. Commun. Technol.*, vol. 31, no. 10, pp. 293–302, 2020, doi: 10.1007/978-3-030-24643-3_35.