



Article

A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques

Balqis Albreiki, Nazar Zaki and Hany Alashwal



Article

A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques

Balqis Albreiki ^{1,2}, Nazar Zaki ^{1,2,*}  and Hany Alashwal ^{1,2}

¹ Department of Computer Science and Software Engineering, College of Information Technology, United Arab Emirates University, Al Ain 15551, United Arab Emirates; 200907523@uaeu.ac.ae (B.A.); halashwal@uaeu.ac.ae (H.A.)

² Big Data Analytics Center, United Arab Emirates University, Al Ain 15551, United Arab Emirates

* Correspondence: Nzaki@uaeu.ac.ae

Abstract: Educational Data Mining plays a critical role in advancing the learning environment by contributing state-of-the-art methods, techniques, and applications. The recent development provides valuable tools for understanding the student learning environment by exploring and utilizing educational data using machine learning and data mining techniques. Modern academic institutions operate in a highly competitive and complex environment. Analyzing performance, providing high-quality education, strategies for evaluating the students' performance, and future actions are among the prevailing challenges universities face. Student intervention plans must be implemented in these universities to overcome problems experienced by the students during their studies. In this systematic review, the relevant EDM literature related to identifying student dropouts and students at risk from 2009 to 2021 is reviewed. The review results indicated that various Machine Learning (ML) techniques are used to understand and overcome the underlying challenges; predicting students at risk and students drop out prediction. Moreover, most studies use two types of datasets: data from student colleges/university databases and online learning platforms. ML methods were confirmed to play essential roles in predicting students at risk and dropout rates, thus improving the students' performance.

Keywords: education data mining; machine learning; MOOC; student performance; prediction; classification



Citation: Albreiki, B.; Zaki, N.; Alashwal, H. A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques. *Educ. Sci.* **2021**, *11*, 552. <https://doi.org/10.3390/educsci11090552>

Academic Editor: Riccardo Pecori

Received: 12 July 2021

Accepted: 12 September 2021

Published: 16 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The recent developments in the education sector have been significantly inspired by Educational Data Mining (EDM). The wide variety of research has discovered and enforced new possibilities and opportunities for technologically enhanced learning systems based on students' needs. The EDM's state-of-the-art methods and application techniques play a central role in advancing the learning environment. For example, the EDM is critical in understanding the student learning environment by evaluating both the educational setting and machine learning techniques. According to information provided in [1], the EDM discipline deals with exploring, researching, and implementing Data Mining (DM) methods. The DM discipline incorporates multi-disciplinary techniques for its success. It has a comprehensive method of extracting valuable and intellectual insights from raw data; the data mining cycle is represented in Figure 1. Machine learning and statistical methods for educational data are analyzed to determine meaningful patterns that improve students' knowledge and academic institutions in general.

Modern learning institutions operate in a highly competitive and complex environment. Thus, analyzing performance, providing high-quality education, formulating strategies for evaluating the students' performance, and identifying future needs are some challenges faced by most universities today. Student intervention plans are implemented in

universities to overcome students' problems during their studies. Student performance prediction at entry-level and during the subsequent periods helps the universities effectively develop and evolve the intervention plans, where both the management and educators are the beneficiaries of the students' performance prediction plans.

E-learning is a rapidly growing and advanced form of education, where students are enrolled in online courses. E-learning platforms such as the Intelligent Tutoring Systems (ITS), Learning Management Systems (LMS), and Massive Open Online Courses (MOOC) take maximum advantage of EDM in developing and building automatic grading systems recommender systems, as well as adaptative systems. These platforms utilize intelligent tools that collect valuable user information such as; frequency of a student's access to the e-learning system, the accuracy of the student's answers to questions, and the number of hours spent reading texts and watching video tutorials [2].

The acquired information, over time, is processed and analyzed using different machine learning methods to improve both the usability and build interactive tools on the learning platform. According to Dr. Yoshua Bengio [3] at the University of Monreal, "research using Machine Learning (ML) is part of Artificial Intelligence (AI), seeking to provide knowledge to computers through data, observations, and close interaction with the world. The acquired knowledge allows the computer to generalize to new settings correctly". Machine learning is a sub-field of AI, where ML systems learn from the data, analyze patterns and predict the outcomes. The growing volumes of data, cheaper storage, and robust computational systems are the reasons behind the rebirth of the machine from just a pattern recognition algorithm to Deep Learning (DL) methods. ML models can automatically and quickly analyze bigger and more complex data with accurate results and avoid unexpected risks.

Although e-learning is widely regarded as a less expensive and more flexible form of education compared to traditional on-campus, it is still regarded as a challenging learning environment since there is no direct interaction between the students and course instructors. Specifically, three main challenges are associated with e-learning systems; (i) the lack of standardized assessment measures for students making it impossible to benchmark across other learning platforms; thus, it is difficult to determine the effectiveness of each learning platform; (ii) e-learning systems have high dropout rates than on-campus studies, due to loss of motivation, especially in the self-paced courses and (iii) predicting students' specialized needs is difficult due to lack of direct communication, especially in case of a student's disability [4,5]. The long-term log data from e-learning platforms such as MOOC, LMS, and Digital Environment to Enable Data-driven (DEED) can be used for student and course assessment.

However, understanding the log data is challenging as not all teachers and course directors understand such valuable data. MOOC and LMS provide free higher education all over the world. These platforms provide student-teacher interaction through their online learning portals [6]. In these portals, the student can select, register and undertake selected courses from anywhere [7]. Machine learning algorithms are useful tools for early predicting students at risk and their dropout chances by utilizing the derived log data. This technique is more advanced than the traditional on-campus where students' records, such as quizzes, attendance, exams, and marks, are used to evaluate and predict their academic performance. The EDM research community utilizes session logs and student databases for processing and analyzing student performance prediction using a machine learning algorithm. This review investigates the application of different techniques of data mining and machine learning to;

1. Predict the performance of students at risk in academic institutions
2. Determine and predict students' dropout from on-going courses
3. Evaluate students' performance based on dynamic and static data
4. Determine the remedial plans for the observed cases in the first three objectives



Figure 1. The typical cycle of Data Mining methodology, image is derived from [8].

There are some previous attempts to survey the literature on academic performance [9,10]; however, most of them are general literature reviews and targeted towards the generic students performance prediction. We aimed to collect and review the best practices of data mining and machine learning. Moreover, we aimed to provide a systematic literature review, as the transparency of the methodology and search strategy reduce the replicability of the review. In this, Grey literature (such as government reports and policy documents) is not included, which may bias perspectives. Although there is one recent publication on the Systematic Literature Review (SLR) of EDM [11], the inclusion and exclusion criteria are different, and they targeted historical data only as compared to our work which is more inclined on recent advances of last 13 years

2. Research Method

A systematic literature review is performed with a research method that must be unbiased and ensure completeness to evaluate all available research related to the respective field. We adopted Okoli's guide [12] for conducting a standalone Systematic Literature Review. Although Kitchenham B. [13], Piper, Rory J. [14], Mohit, et al. [15], and many other researchers provided a comprehensive procedure systematic literature review, most of them concentrate on only substantial parts of the process, and only a few followed the entire process. The chosen method introduces the rigorous, standardized methodology for the systematic literature review. Although the research is mainly tailored to information system research, it is sufficiently broad to be applicable and valuable to scholars from any social science field. Figure 2 provides the detailed flowchart of Okoli's guide for systematic literature review.

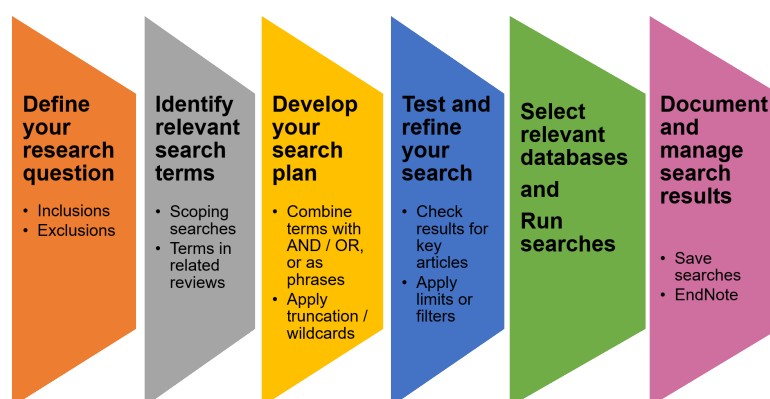


Figure 2. Okoli's guide [12] for conducting a standalone Systematic Literature Review.

Since research questions are the top priority for a reviewer to identify and handle in SLR, we tried to tackle the following research questions throughout the review.

2.1. Research Questions

- What type of problems exist in the literature for Student Performance Prediction?
- What solutions are proposed to address these problems?
- What is the overall research productivity in this field?

2.2. Data Sources

In order to carry out an extensive systematic literature review based on the objectives of this review, We exploited six research databases to find the primary data and to search for the relevant papers. The databases consulted in the entire research process are provided in Table 1. These repositories were investigated in detail using different queries related to ML techniques to predict students at risk and their dropout rates between 2009 and 2021. The pre-determined queries returned many research papers that were manually filtered to retain only the most relevant publications for this review.

Table 1. Data Sources.

Identifiers	Databases	Access Date	URL	Results
Sr.1	ResearchGate	4 February 2021	https://www.researchgate.net/	83
Sr.2	IEEE Xplore Digital Library	4 February 2021	https://ieeexplore.ieee.org/	78
Sr.3	Springer Link	6 February 2021	https://link.springer.com/	20
Sr.4	Association for Computing Machinery	4 February 2021	https://dl.acm.org/	39
Sr.5	Scopus	4 February 2021	https://www.scopus.com/	33
Sr.6	Directory of Open Access Journals	4 February 2021	https://doaj.org/	54

2.3. Used Search Terms

Following search terms (one by one) were used to find out data from the databases according to our research questions:

- EDM OR Performance OR eLearning OR Machine Learning OR Data Mining
- Educational Data Mining OR Student Performance Prediction OR Evaluations of Students OR Performance Analysis of Students OR Learning Curve Prediction
- Students' Intervention OR Dropout Prediction OR Student's risks OR Students monitoring OR Requirements of students OR Performance management of students OR student classification.
- Predict* AND student AND machine learning

2.4. The Paper Selection Procedure for Review

The selection procedure of the paper is comprised of identification, screening, eligibility checking, and inclusion criteria meeting of the research papers. Authors independently collected the research papers and were agreed on the included papers. Figure 3 provides the detailed structure of the review selection procedure after applying Okli's guide [12] for systematic review.

2.5. Inclusion and Exclusion Criteria

2.5.1. Inclusion

- Studies related to Student's Performance Prediction;
- Research Papers that were accepted and published in a blind peer-reviewed Journals or conferences;
- Papers that were from 2009 to 2021 era;
- Paper that were in the English language.

2.5.2. Exclusion

- Studies other than Student's Performance Prediction using ML.
- Papers which had not conducted experiments or had validation of proposed methods.
- Short papers, Editorials, Business Posters, Patents, already conducted Reviews, Technical Reports, Wikipedia Articles, Survey Studies, and extended papers of already reviewed papers.

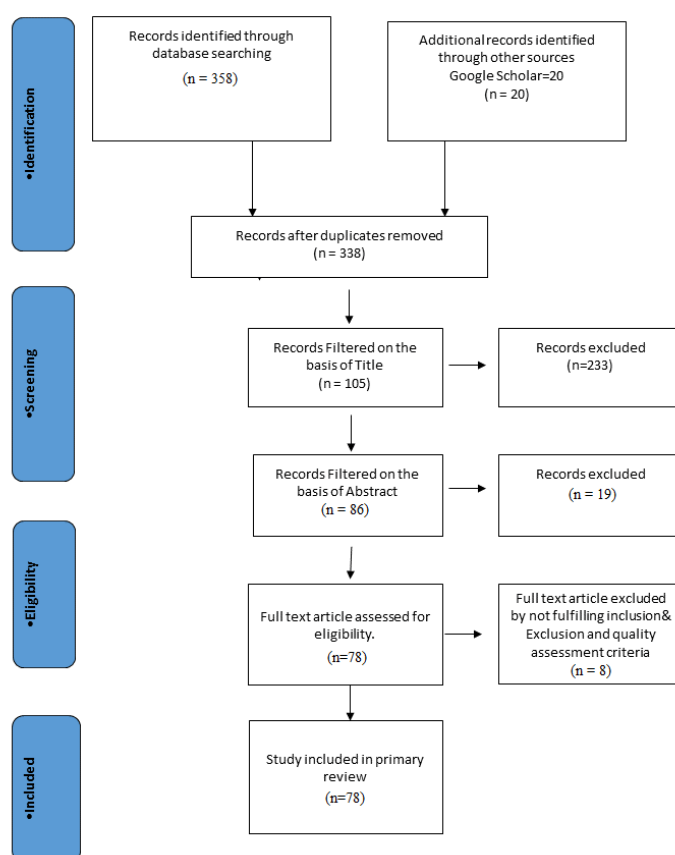


Figure 3. Detailed structure of review selection procedure after applying Okli's [12] for systematic review.

2.6. Selection Execution

The search is executed to obtain the list of studies that can be used for further evaluation. The bibliography management of the studies is performed by a bibliography tool

named Mendeley. These bibliographies contain those studies that are entirely fit the inclusion criteria. After successfully implementing inclusion and exclusion criteria, the resulting 78 papers are described in detail in the following section. Table 2 presents the number of papers selected from each year. All the papers mentioned below have been included in the review.

Table 2. Number of research articles from 2009 to 2021.

Year	References	Count
2009	[4,16,17]	3
2010	[1,18–21]	5
2011	[22]	1
2012	[23–27]	5
2013	[28–31]	4
2014	[7,32–37]	7
2015	[5,6,38–45]	10
2016	[31,46–52]	8
2017	[51,53–60]	9
2018	[33,45,51,61–67]	10
2019	[2,57,68–70]	5
2020	[71–77]	7
2021	[78–80]	4

2.7. Quality Assessment Criteria

The following quality criteria are defined for the systematic literature review:

- QC1: are review objectives clearly defined?
- QC2: are proposed methods well defined?
- QC3: is proposed accuracy measured and validated?
- QC4: are limitations of the review explicitly stated?

3. Results and Discussion

3.1. Predicting the Performance of Students at Risk Using ML

The students' performance prediction provides excellent benefits for increasing student retention rates, effective enrollment management, alumni management, improved targeted marketing, and overall educational institute effectiveness. The intervention programs in schools help those students who are at risk of failing to graduate. The success of such programs is based on accurate and timely identification and prioritization of the students requiring assistance. This section presents a chronological review of published literature from 2009 to 2021, documenting at-risk student performance using ML techniques. Research studies related to dataset type, feature selection methods, criteria applied for classification, experimentation tools, and outcome of the proposed approaches are also summarized.

Kuzilek et al. [5] were focused on General Unary Hypotheses Automation (GUHA) and Markov Chain-based analysis to analyze the student activities in VLE systems. A set of scenarios was developed containing 13 scenarios. The dataset used in this review contained two types of information, i.e., (a) student assignment marks and (b) the VLE activity log that represented student's interaction with the VLE system. Implementation was undertaken using the LISp-Miner tool. Their investigation concluded that both methods could discover valuable insights into the dataset. Markov Chain-based graphical model can help in visualizing the fact, which can be easier to understand. The patterns extracted using the

methods mentioned above provide sub-station support to the intervention plan. Analyzing student behavioural data helps predict student performance during their academic journey.

He et al. [6], examine students at risk identification in MOOCs. They proposed two transfer learning algorithms, namely “Sequentially Smoothed Logistic Regression (LR-SEQ) and Simultaneously Smoothed Logistic Regression (LR-SIM)”. The proposed algorithms are evaluated using DisOpt 1 and DisOpt2 datasets. Comparing the results with the baseline Logistic Regression (LR) algorithm, LR-SIM outperformed the LR- SEQ in terms of AUC, where the LR-SIM had a high ACU value in the first week. This result indicated a promising prediction at the early stage of admission.

Kovacic, Z. [18] analyzed the early prediction of student success using machine learning techniques. The review investigated the socio-demographic features, i.e., education, work, gender, status, disability, etc., and course features such as course program, course block, etc., for effective prediction. These features containing the dataset were collected from the Open University of New Zealand. The machine learning algorithms for feature selection are used to identify the essential features affecting the students’ success. The key finding from the investigation was that ethnicity, course program, and course block are the top three main features affecting students’ success.

Kotsiaritis et al. [19] proposed a technique named the combinational incremental ensemble of classifiers for student performance prediction. In the proposed technique, three classifiers are combined where each of the classifiers calculates the prediction output. A voting methodology is used to select the overall final prediction. Such a technique is helpful for continuously generated data, and when a new sample arrives, each classifier predicts the outcome. The final prediction is selected using the voting system. In this review, the training data is provided by Hellenic Open University. The dataset comprises writing assignments marks containing 1347 instances, with each having four attributes with four features for written assignment scores. The three algorithms used to build the system using a combinational incremental ensemble are naive Bayes (NB), Neural Network (NN), and WINDOW. The system works so that the models are initially trained using the training set, followed by the test of the models using the test set. When a new instance of observation arrives, all three classifiers predict the value, and the ones with high accuracy are automatically selected. Craige et al. [22] used statistical approaches, NN, and Bayesian data reduction approaches to help to determine the effectiveness of the Student Evaluation of Teaching Effectiveness (SETE) test. The results show no support for SETE as a general indicator of teaching effectiveness or student learning on the online platform. In another review by Kotsiantis, Sotiris B. [23] proposed a decision support system for a tutor to predict students’ performance. This review considers student demographic data, e-learning system logs, academic data, and admission information. The dataset is comprised of 354 student’s data having 17 attributes each. Five classifiers are used, namely; Model Tree (MT), NN, Linear Regression (LR), Locally Weighted Linear Regression, and Support Vector Machine (SVM). MT predictor attains high Mean Absolute Error (MAE).

Osmanbegovic et al. [24] analyze Naive Bayes (NB), Decision Tree (DT), and Multilayer perception (MLP) algorithms to predict students’ success. The data is comprised of two parts. The first part of data is collected from the survey conducted at the University of Tuzla in 2010–2011. The participants were the students of the first year from the department of economics. The second part of the data is acquired from the enrollment database. Collectively, the dataset has 257 instances with 12 attributes. They used Weka software as an implementation tool. The classifiers are evaluated using accuracy, learning time, and error rate. The NB attains a high accuracy score of 76.65% with a training time of less than 1 s and high error rates. Baradwaj and Pal [25] also review data mining approaches for student performance prediction. They investigate the accuracy of DT, where the DT is used to extract valuable rules from the dataset. The dataset utilized in their review was obtained from Purvarichal University, India, comprising 50 students’ records, each having eight attributes.

Watson et al. [28] considered the student activity log enrolled in the introductory programming of a course to predict their performance. This review advised a predictor based on automatically measured criteria rather than a direct basis to determine the evolving performance of students over time. They proposed a scoring algorithm called WATWIN that assigns specific scores to each activity of student programming. The scoring algorithm considers the student's ability to deal with the programming errors and the time to solve such errors. This review used the programming activity log data of 45 students from 14 sessions as a dataset. The activity of each student was assigned a WATWIN score, which is then used in linear regression. Linear regression using the WATWIN score achieves 76% accuracy. For effective prediction, the dataset must be balanced. Balanced data mean that each of the prediction classes has an equal number of attributes.

Marquez-Vera et al. [29] shed light on the unbalanced nature of the dataset available for student performance prediction. Genetic algorithms are very rarely used for prediction. This review has been compared between 10 standard classification algorithms implemented in Weka and three variations of genetic algorithms. The 10 Weka implemented classification algorithms are Jrip, NNge, OneR, Prison, Ridor, ADTree, J48, Random Tree, REPTree, and Simple CART, whereas three variations of the genetic algorithm are (Interpretable Classification Rule Mining) ICRM v1, ICRM v2, and ICRM v3 that employs Grammar Based Genetic Algorithm (G3P). For class balancing, the author used SMOTE, also implemented in Weka. The results show that the genetic algorithm ICRM v2 score high accuracy when the data is balanced, whereas the performance is slightly low when the data is not balanced. The data used in this review have three types of attributes, including a specific survey (45 attributes), a General survey (25 attributes), and scores (seven attributes).

Hu et al. [32] explore the time-dependent attributes for predicting student online learning performance. They proposed an early warning system to predict students' performance at risk in an online learning environment. They advised the need for time-dependent variables as an essential factor for determining student performance in Learning Management Systems (LMS). The paper focused on three main objectives as follows; (i) investigation of data mining technique for early warning, (ii) determination of the impacts of time-dependent variables, (iii) selection of data mining technique with superior predictive power. Using data from 330 students of online courses from the LMS, they evaluated the performance of three machine learning classification models, namely "C4.5 Classification and Regression Tree (CART), Logistic Regression (LGR), and Adaptive Boosting (AdaBoost)". Each of the instances in the dataset consisted of 10 features, and the performance of the classifiers is evaluated using accuracy, type I, and type II errors. CART algorithm outperforms the other algorithms achieving accuracy greater than 95%.

Lakkaraju et al. [38] proposed a machine learning framework for identifying the student at risk of failing to graduate or students at risk of not graduating on time. Using this framework, the student's data were collected from two schools in two districts. Five machine learning algorithms used for experimentation purposes include; Support Vector Machine (SVM), Random Forest, Logistic Regression, Adaboost, and Decision Tree. These algorithms are evaluated using precision, recall, accuracy, and AUC for binary classification. Each student is ranked based on the risk score estimated from the classification, as mentioned above model. The results revealed that Random Forest attains the best performance. The algorithms are evaluated using precision and recall at top positions. In order to understand the most likely mistake, the proposed framework can make the authors suggest five critical steps for the educators. These include; (a) identification of frequent patterns in data using FP-Growth algorithm, (b) use risk of the score for ranking the students, (c) addition of a new field in the data and assign a score of one (1) if the framework failed to predict correctly otherwise a score of zero (0), (d) computation of probability mistake for each of the frequent patterns, and (e) sorting of the patterns based on the mistake probability.

Ahmed et al. [45] collected student data between 2005 and 2010 from the educational institute student database. The dataset contains 1547 instances having ten attributes. The

selected attributes gathered information such as; departments, high school degrees, mid-term marks, lab test grades, seminar performance, assignment scores, student participation, attendance, homework, and final grade marks. Two machine learning classification methods, DT and ID3 Decision Tree, are used for data classification. Weka data mining tool is then used for experimentation. The information gained is used to select the root node; the midterm attribute has been chosen to be the root node.

The performance prediction of the new intakes is studied by Ahmed et al. [45]. They contemplated a machine learning framework for predicting the performance of first-year students at FIC, UniSZA Malaysia. This review collected students' data from university databases, where nine attributes were extracted, including gender, race, hometown, GPA, family income, university mode entry, and SPM grades in English, Malay languages, and Math. After pre-processing and cleaning the dataset, demographic data of 399 students from 2006–2007 to 2013–2014 is extracted. Three classifiers, including Decision Tree, Rule-based and Naive Bayes performance, were examined. The results conceded the rule-based classifier as the best performing with 71.3% accuracy. Weka tool is used for experimentation purposes. Students' performance prediction in the online learning environment is significant as the rate of dropouts is very high as compared to the traditional learning system [6].

Al-Barrak and Al-Razgan [46] considered the student grades in previous courses to predict the final GPA. For this purpose, they used students' transcript data and applied a decision tree algorithm for extracting classification rules. The application of these rules helps identify required courses that have significant impacts on the student's final GPA. The work of Marbouti et al. [47] differed from the previous studies in that their investigation analyzed predictive models to identify students at risk in a course that uses standard-based grading. Furthermore, to reduce the size of the feature space, they adopted feature selection methods using the data for the first-year engineering course at Midwestern US University from the years 2013 and 2014. The student performance dataset had class attendance grades, quizzes grades, homework, team participation, project milestones, mathematical modeling activity test, and examination scores. Six machine learning classifiers analyzed included LR, SVM, DT, MLP, NB, and KNN. These classifiers were evaluated using different accuracy measures such as; overall accuracy, accuracy for pass students, and accuracy for failed students. The feature selection method used Pearson's correlation coefficient value, where features with a correlation coefficient value > 0.3 were used in the prediction process. Naive Bayes classifiers had higher accuracy (88%) utilizing 16 features.

In a similar review, Iqbal et al. [53] also predicted student GPA using three machine learning approaches; CF, MF, and RBM. The dataset they used in this review is collected from Information Technology University (ITU), Lahore, Pakistan. They proposed a feedback model to calculate the student's understanding of a specific course. They also suggested a fitting procedure for the Hidden Markov model to predict student performance in a specific course. For the experiment, the data split was 70% for the training set and 30% for the testing data. The ML-based classifiers were evaluated using RMSE, MSE, and Mean Absolute Error (MAE). During the data analysis, RBM achieved low scores of 0.3, 0.09, and 0.23 for RMSE, MSE, and MAE, respectively. Zhang et al. [54] optimized the parameter of the Gradient Boosting Decision Tree (GBDT) classifier to predict the student's grade in the graduation thesis in Chinese universities. With customized parameters, GBDT outperforms KNN, SVM, Random Forest (RF), DT, LDA, and Adaboost in terms of overall prediction accuracy and AUC. The dataset used in this review comprised 771 samples with 84 features from Zhejiang University, China. The data split was 80% training set and 20% testing set.

Hilal Almarabeh [55] investigated the performance of different classifiers for the analysis of student performance. A comparison between five ML-based classifiers has been made in this review. These classifiers include Naive Bayes, Bayesian Network, ID3, J48, and Neural Networks. Weka implementation of all these algorithms is used in experiments. The data for analysis is obtained from the college database with 225 instances, where each instance comprised ten attributes. The results are shown in this review reveal the Bayesian

Network as the most practical prediction algorithm. Jie Xu et al. [56] proposed a new machine learning method having two prominent features. The first feature was a layered structure for prediction considering the ever-evolving performance behavior of students. The layered structure is comprised of multiple bases and ensemble predictors. The second important feature of the proposed method considered the data-driven approach used to discover course relevance. The dataset consisted of 1169 students record enrolled in the Aerospace and Mechanical Engineering departments of UCLA. The proposed method showed promising results in terms of the Mean Square Error (MSE).

Al-shehri et al. [57] carried out a similar review that compared the performance of supervised learning classifiers, SVM and KNN, using data from the University of Minho that had 33 attributes. The dataset was first converted from nominal to numeric forms before analyzing statistically. The dataset was initially collected using questionnaires and reports from two schools in Portugal. The original data also contained 33 features, among which nominal, binary and numeric attributes distribution was 4, 13, and 16, respectively, with 395 the total number of instances. The Weka tool was used in the experiment where the algorithms were tested using different data partition sets. The result was that the SVM achieves high accuracy when using 10-Fold cross-validation and partition ratio.

The application of advanced learning analytics for student performance prediction was examined by Alowibdi, J. [58]. They considered those students on scholarships in Pakistan. This research analyzed discriminative models CART, SVM, C4.5, and generative models Bayes Network and NB. Precision recall and F-score were used to evaluate the predictor performance. Three thousand students' data from 2004 to 2011 was initially collected, which was reduced to 776 students after pre-processing and redundancy elimination. Among these 776 students, 690 completed their degree successfully, whereas 86 were failed to complete the degree programs. A total of 33 features were categorized into four groups: family expenditure, family income, student personal information, and family assets. The review found that natural gas expenditure, electricity expenditure, self-employment, and location" were the most prominent predicting student academic performance. SVM classifier outperforms all other approaches by scoring a 0.867 F1 score.

A hybrid classification approach is proposed by Al-Obeidat et al. [61] combining PROAFTAN- a multi-criteria classifier- and DT classifiers. The proposed algorithm works in three stages; in the first stage, the C4.5 algorithm is applied to the dataset with discretization followed by the data filtering and pre-processing stage, and finally, enhances C4.5 with PROAFTAN with attribute selection and discretization. They used the same UCI dataset as used in [81]. The dataset was comprised of students enrolled in languages and Math courses. The proposed hybrid classification algorithm is evaluated using precision, recall, and F-measure. The authors recorded significant improvement in accuracy for both Languages (82.82%) and Math (82.27%) in the students' dataset. In comparisons with RF, NB, Meta Bagging (MB), Attribute Selected Classifier, Simple Logistic (SL), and Decision Table (DT) algorithms, the proposed hybrid approach attained accuracy, precision, recall, and F-measure scores.

Kaviyarasi and Balasubramanian [62] examined the factors affecting the students' performance. The authors classified the students into three classes; Fast learner, Average Learner, and Slow Learner. The data used was belonged to affiliated colleges of the Periyar University, where 45 features were extracted from the dataset. For classification, Extra Tree (ET) classifier was used to calculate the importance of these features. Twelve top features were identified as important features for predicting student academic performance. Zaffar et al. [63] compared the performance of feature selection methods using two datasets. Dataset 1 consisted of 500 student records with 16 features. Whereas dataset 2 contained 300 students records with 24 features. Weka tool was used for experimentation, and the results revealed that the performance of the feature selection methods depends on the classifiers used and the nature of the dataset.

Chui et al. [64] considered the extended training time of the classifier and proposed a Reduced Training Vector-Based—Support Vector Machine (RTV-SVM) classifier to predict

marginal or at-risk students based on their academic performance. RTV-SVM is a four stages algorithm. The first stage is the input definition, followed by the multivariable approach, as the second stage is used for tier-1 elimination of training vectors. In the third stage of RTV-SVM, vector transformation for tier-2 elimination of the training vectors is used, and the SVM model is in the final stage using the SMO algorithm. OULA dataset was used in [64], comprising of 32,593 student records containing both student demographics data and session log of student interaction with the VLE system. TRV-SVM scored high accuracy of 93.8% and 93.5% for predicting the student at risk and marginal student respectively while significantly reducing the training time by 59%.

Msaci et al. [65] proposed machine learning and statistical methods to examine the determinants of the PISA 2005 test score. The author analyzed PISA 2005 data from several countries, including Germany, the USA, UK, Spain, Italy, France, Australia, Japan, and Canada. This investigation aimed to explore the students' and academic institutions' characteristics that may influence the student's achievements, where the proposed approaches work in two steps. In the first step, a multilevel regression tree is applied considering students nested with schools, and student-level characteristics related to student achievement are identified. In the second step, school value-additions are estimated, allowing for identifying school-related characteristics using regression tree and boosting techniques. The PISA 2015 dataset from the nine countries was used where the total number of attributes at the school level and student level was 19 and 18, respectively. The number of students in sample size (Table 3). The results obtained suggested that both the student-level and school-level characteristics have an impact on students' achievements.

Livieris et al. [68] suggested a semi-supervised machine learning approach to predict the performance of secondary school students. The approach considered in this review included self-training and Yet Another Two-Stage Idea (YATSI). The dataset had performance data of 3716 students collected by Microsoft Showcase School. Each instance in the dataset has 12 attributes. The semi-supervised approaches perform well on the data as compared to supervised and unsupervised learning approaches. For better decision-making, Nieto et al. [69] compared the performance of SVM and ANN, where 6130 students' data was collected, and after pre-processing and cleaning, 5520 instances with multiple features were extracted. KNIME software tool was used for the implementation of SVM and ANN. It was realized that the SVM attained a high accuracy of 84.54% and high AUC values.

Table 3. Student sample size in the selected nine countries based on 2015 PISA dataset.

No.	Country	Sample Size
1	Canada	20,058
2	Australia	14,530
3	UK	14,157
4	Italy	11,586
5	Spain	6736
6	Japan	6647
7	Germany	6504
8	France	6108
9	US	5712

Aggarwal et al. [78] compared academic features and discussed the significance of nonacademic features, such as demographic information, by applying eight different ML algorithms. They utilized a dataset from a technical college in India, which has information about 6807 students with academic and non-academic features. They applied Synthetic minority oversampling methods to reduce the skewness in the dataset. They claimed J48

93.2% F1 score with Decision Tree, 90.3% with Logistic, 91.5% with Multi-Layer Perceptron, 92.4% with Support Vector Machine, 92.4% with AdaBoost, 91.8% with Bagging, 93.8% with Random Forest, and 92.3% with Voting. They also suggested that academic performance is not only dependent on the academic features, but it has a high influence on demographic information as well; they suggested using the non-academic features with a combination of academic features for predicting the student's performance.

Zeineddine et al. [79] utilized the concept of AutoML to enhance the accuracy of student's performance prediction by exploiting the features prior to the starting of the new academic program (prestart data). They achieved 75.9% accuracy with AutoML with a lower false prediction rate. With a Kapa of 0.5. Accordingly, they encourage researchers in this field to adopt AutoML in their search for an optimal student performance prediction model, especially when using pre-start data. They suggested employing the pre-admission data and start intervention and consulting sessions before starting the academic progress, so the students who need immediate help may survive in society. They observed the available data in unbalanced, and they employed SMOTE pre-processing method and then employed the auto-generated Ensemble methods to predict the failing students with an overall accuracy of 83%. The authors acknowledged the overgeneralization limitation of SMOTE and discussed some methods to reduce the unbalancing data problem without the overgeneralization problem. OuahiMariame and Samira [80] evaluated the usage of neural networks in the field of EDM with the perspective of feature selection to classification. They utilized various Neural networks on different student databases to check their performance. They claimed that the NN had surpassed various algorithms such as Naïve Bayes, support vector machine (SVM), RandomForest, and Artificial Neural Network (ANN), to successfully evaluate the student's performance.

Thai-Nghe et al. [82] proposed a recommender system for students' performance prediction. In this method, Matrix Factorization, Logistic Regression, and User-Item collaborative filtering approach performance are compared using the KDD challenge 2010 dataset. The dataset contains a log file of students obtained when the students interact with the computer-aided tutoring system. The results of this review suggested that recommender systems based on Matrix Factorization and User-Item collaborative filtering approaches have a low Average Root Mean Squared Error (RMSE) of 0.30016.

Buenaño-Fernández [83] proposed the usage of ML methods for the final grades prediction of students by using the historical data. They applied the historical data of computer engineering from the universities of Ecuador. One of the strategic aims of this research was to cultivate extensive yet comprehensive data. Their implementation had yielded a panoptic amount of data which can be converted into several useful education-related applications if processed appropriately. This paper proposed a novel technique for pre-processing and grouping of students having the same patterns. Afterward, they applied many supervised learning methods to identify the students who had similar patterns and their predicted final grades. Finally, the results from ML methods were analyzed and compared with the previous state of art methods. They claimed 91.5% accuracy with ensemble techniques, which shows the effectiveness of ML methods to estimate the performance of students.

Reddy and Rohith [84] discussed that many researchers had utilized the advanced ML algorithms to predict the student's performance effectively; however, they did not provide any competent leads to under-performing students. They aimed to beat the limitation and worked to identify the explainable human characteristics that may determine the student will have poor tutorial performance. They used the data from the University of Minnesota and applied SVM, RF, Gradient Boosting, and Decision Trees. They claimed more than 75% accuracy to identify the factors which are generic enough to spot out which students will be failing this term.

Anal Archrya and Devadatta Sinha [85] also proposed an early prediction system using ML-based classification methods utilizing the embedded feature selection methods for reducing the feature set size. The total number of features in this review is 15, which are

collected through questionnaires. The survey participants are educators and students of computer science from different colleges in Kolkata, India. The authors reported the C4.5 classifier as the best performing algorithm comparing to Multi-Layer Perception (MLP), Naive Bayes (NB), K-NN, and SMO. In another review conducted at The Open University, the United Kingdom by Kuzilek et al. [5], developed a system comprised of three predictive algorithms to identify the students at risk. The three ML-based algorithms (including Naive Bayes, K-NN, and CART) attain a predictive score using two datasets. The first dataset is a demographic dataset collected from the university database and the second dataset consisted of log data with structural interaction on the Virtual Learning Environment (VLE) system. The final score of each student was calculated as the sum of the predictive score by each algorithm. If the final score was >2 , the student was determined to be at risk, and appropriate measures are implemented. However, if the final score was <3 , the student is not at risk and does not require intervention. Precision and recall score helps evaluate the performance of the proposed system.

E-learning platforms received considerable attention from the EDM research community in recent years. Hussain et al. [81] examined ML methods to predict the difficulties that students encounter in an e-learning system called Digital Electronics Education and Design Suites (DEEDS). EDM techniques consider the student's interaction with the system to identify meaningful patterns that help the educator improve their policies. In this work, data of 100 first-year BSc students from the University of Genoa are used. The data is comprised of session logs created when the students interact with the DEEDS tool and are publicly available at the UCI machine learning repository. Five features selected for student performance prediction included average time, the total number of activities, average idle time, the average number of critical storcks, and total related activities. THIS REVIEW'S five ML algorithms included; ANN, LR, SVM, NBC, and DT. The performance of the classifiers was evaluated using the RMSE, Receiver Operator Characteristics (ROC) curve, and Cohen's Kappa Coefficient. Accuracy, precision, recall, and F-score were also used as performance parameters. ANN and SVM had identical results in terms of RMSE and performance parameters. The authors argued the importance of SVM and ANN algorithms and proposed a modified DEEDS system where ANN and SVM are part of such systems for student performance prediction.

Comparisons of Performance Prediction Approaches

Accurate prediction of students' performance and identification of students at risk on e-learning platform utilized four approaches; (i) prediction of academic performance, (ii) identification of students at risk, (iii) determination of difficulties in an e-learning platform, and (iv) evaluation of the learning platform. Of these approaches, most research studies show that the prediction of student's academic performance is a crucial area of interest, with a total of 16 research studies undertaken between 2009 and 2021. Identifying students at risk was the second after the performance prediction, with 12 research studies undertaken in the same period. Each research is unique in the methodology used and the type attributes selected to determine the relevant algorithms applied during classification. Students' interaction with the e-learning platform was the most sought attribute, where the 1st year students were the most considered during the research process. Very few studies (5) sought to understand the e-learning platform and its impact on students' performance. Overall, the commonly applied algorithms were; DT, LR, NB, MT, and SVM. Table 4 provides details for performance prediction and identification of students at risk.

Table 4. Prediction of student performance and identification of students at risk in e-learning.

Approach	Methodology	Attributes	Algorithms	Count	References
Performance prediction	Early prediction- ML	Socio-demographic	Rule- base	2	[18,58]
	Incremental ensemble	Teaching effectiveness	NB, 1-NN, and WINDOW	2	[19,22]
	Recommender system	Student's platform interaction	MT, NN, LR, LWLR, SVM, NB, DT, MLP	5	[23,24,82]
	Automatic measurement	Students' activity log	WATWIN	2	[25,28]
	Dynamic approach	1st-year students	LR-SEQ, LR-SIM, DT, Rule-based & NB	3	[28,29]
	Semi-supervised ML	Secondary schools	YATSI, SVM, ANN	6	[6,34,39] [68,69,78–80]
Identification of students at-risk	At-risk of failing to graduate			3	
	ML framework	Early prediction of at-risk students	SVM, RF, LR, Adaboost, CART, and DT	1	[32,38,45]
	Reducing feature set size	Final GPA results	CART, C4.5, MLP, NB, KNN & SMO CF		[85]
	Student previous grades	Identification of students at risk	ME, RBM, GBDT, KNN, SVM, RF, DT, LDA, Adaboost	3	[46,53,54]
	Predictive models- grading	Fast Learner,	LR, SVM, DT, MLP, NB, and KNN	2	[47,57]
	Factors affecting- at- risk	Average & Slow Learner	Extra Tree (ET), RTV-SVM	3	[62–64]
Predict the difficulties of the learning platform	Examination of ML methods	Difficulties encountered on the e-learning system	ANN, LR, SVM, NBC, and DT	2	[61,81]
Performance of classifiers	Cross comparison	Comparison between five ML-based classifiers	NB, BN, ID3, J48, and NN	2	[55,56]
Evaluation of MOOC in developed countries	Discriminants of the PISA 2005 test score	Characteristics of students and academic institutions	ML and statistical methods	1	[65]

3.2. Students Dropout Prediction Using ML

Accurate prediction of students' dropout during the early stages helps eliminate the underlying problem by developing and implementing rapid and consistent intervention mechanisms. This section describes in detail dropout prediction using machine learning techniques through a review of related research based on datasets, features used in ML methods, and the outcome of the studies.

The early review by Quadri and Kalyankar [17,20] used decision tree and logistic regression to identify the features for dropout prediction. In these studies, the authors used the students' session logs dataset, where a Decision Tree was used to extract dropout factors while the logistic regression was used to quantify the dropout rates. The combination of ML techniques for dropout prediction was also investigated by Loumos, V. [16] using three machine learning algorithms: the Feed-Forward Network, SVM, and ARTMAP. Three decision schemes for dropout reduction using the above method ML techniques were suggested, i.e., in decision scheme 1, the student was considered dropout if at least one algorithm classifies him/her as a dropout. In decision scheme 2, a student is considered a potential dropout if two algorithms classify a student in this manner, and in decision scheme 3, a student is considered a dropout if all the algorithms declare the student as a dropout. The dataset used in their review comprised records of students' between 2007 and 2008 registered in two e-learning courses. The total number of students is 193, including; gender, residency, working experience, education level, MCQ test grades, project grade, project submission date, and section activity. For experimentation purposes, the year 2007 data is used for training, while 2008 data were used for testing purposes. Accuracy, sensitivity, and precision measures were used to evaluate the performance of the classifiers, and the results indicated that decision scheme 1 is the appropriate scheme to predict students' dropout.

Oyededeji et al. [71] applied machine learning-based techniques in order to analyze the student's academic performance to help out educationists and institutions that are curious to extract the methods that can improve the individual's performance in academia. Their review performed the analysis of past results with the combination of individual attributes

such as the student's age, their demographic distribution, individual attitude towards review, and family background by employing various machine learning-based algorithms. They concluded three significant models for the comparative performance analysis, i.e., Linear regression supervised learning and deep learning. They suggested MAE of 3.26, 6.43, and 4.6, respectively.

Ghorbani and Ghousi [77] compared numerous resampling techniques to predict the student's dropout using two datasets. These techniques include Random Over-Sampling, Borderline SMOTE, SMOTE, SVM-SMOTE, SMOTE-Tomek, and SMOTE-ENN. Their primary goal is to handle the data unbalancing problems while proposing an adequate solution for the performance prediction. They applied many algorithms on balanced data such as RF, KNN, ANN, XG-boost, SVM, DT, LG, and NB. They claimed that the combination of the Random Forest classifier with the balancing technique of SVM-SMOTE provided the best results as 77.97% accuracy by employing shuffle 5-fold cross-validation tests on multiple datasets.

Alhusban et al. [72] employed machine learning analysis to Measure and Enhance the Undergraduate Student's dropout. They collected the data from the students of Al-Al Bayt University and measured various factors for practical analysis such as gender, enrolment type, admission marks, birth city, marital status, Nationality, and subjects studied at the k-12 stage. Many features are included, which makes the big sample data; they exploited Hadoop, a machine learning-based open-source platform. They found that there is a significant effect of admission test marks on the specialization. Moreover, they suggested that specific genders, such as there dominating certain fields, are a massive number of girls specializing in medical compared to the boys. They also claimed the effects of students' social status on performance. Finally, they claimed that Single students performed better compared to the married ones or the ones in some relationships.

Hussain et al. [76] applied a machine-learning-based methodology for the expected performance prediction of students. They collected the curricular data and non-curricular data from university daily activities. They suggested the application of a fuzzy neural network that was trained by exploiting the metaheuristic method. As the original FNN was based on gradient-based error correction and was limited in overall efficiency, they suggested applying Henry Gas Solubility Optimization and fine-tuning the FNN parameters. They compared the proposed methodology with several state-of-the-art BA, ABC, PSO, CS, NB, k-NN, RF, ANN, DNN, ADDE, and Hybrid Stacking. They conducted rigorous experiments on the proposed methodology and claimed 96.04% accuracy in the early prediction student's performance. Wakelam et al. [75] described an experiment that can be conducted on the students of the university that are in their final year by using a module cohort of 23. They use readily available features like lecture attendance, learning environment, Quiz marks, and intermediate assessments. They found these factors as a potential feature for the prediction of an individual's performance. They employed DT, KNN, and RF on the self-generated data and claimed 75% average accuracy in the performance prediction with only little data and small, easily accessible attributes.

Walia et al. [74] applied classification algorithms in order to predict the student's academic performance, such as NB, DT, RF, ZeroR, and JRip. They found that the school, student's attitude, gender, and review time affect the performance in terms of the final grade. They have done a massive amount of experiments with the help of the Weka tool and claimed more than 80.00% accuracy in their self-generated dataset. Similarly, Gafarov, F. M., et al. [73] applied the data analysis on the records of students from Kazan Federal University. The data was collected with the collaboration of the institution ranging from 2012 to 2019. They applied standard analysis tools like Weka and IBM-SPSS and devised different results. They concluded that if sufficient data is collected, it can be a lot easier to apply the advanced algorithm and achieve more than 98% accuracy using modern programming tools and languages.

The dropout rate is considered high in distance education courses as compared to traditional on-campus courses. Kotsiantis, S. [17] argue that student dropout prediction

is essential for the universities providing distance education. The dataset collected for predicting dropout is an imbalance in nature as most of the instances belong to one class. In [17], Hellenic Open University (HOU) distance learning dataset was used for experimental purposes. The feature set comprised two types of features; curriculum-based features and student performance-based features. In this review, the authors suggest a cost-sensitive prediction algorithm that is based on K-NN. The proposed algorithm achieved promising performance on the imbalance dataset as compared to the baseline method. Marquez-vera [21] also considered class imbalance issues to determine the students' failure using the information of 670 students from Zacatecas. For class balancing, the SMOTE algorithm was implemented using the Weka software tool. SMOTE is an oversampling method for data resampling. The results show that applying ML algorithms on balanced data and feature selection based on their frequencies can enhance the classifier's performance to predict the possibility of students' dropout.

Mark Plagge [30] studied Artificial Neural Networks (ANN) algorithms to predict the retention rate of first-year students between 2005 and 2010 registered at Columbus State University. In this review, two ANN algorithms, (i) feed-forward neural network and (ii) cascade feed-forward neural network, were investigated. The results suggested that 2-layered feed-forward ANN achieved a high accuracy of 89%. Saurabh Pal [26] proposed a predictive model to identify the possible student dropouts by utilizing the decision tree variants including CART, ADT, IDB, and C4.5 in the prediction process. The dataset contained 1650 instances with 14 attributes each. The Weka tool was also used to implement the variance of decision tree variants in which the result showed that ID3 attained a high accuracy of 90.90% followed by C4.5, CART, and ADT with an accuracy score of 89.09%, 86.66%, and 82.27% respectively.

Owning the temporal nature of dropout factors [35] Mi and Yeung [40] proposed applicable temporal models. Their review proposed two versions of the Hidden Markov Model (HMM) and named as Input-Output Hidden Markov Model (IOHMM) IOHMM1 and IOHMM2. Moreover, a modified version of the Recurrent Neural Network with LSTM cell as a hiding unit was also proposed in this review. The performance of the proposed methods was then compared with the baseline classification model using a dataset collected from MOOC's platform. The results showed the dominance of RNN combined with LSTM as the best classification model, whereas IOHMM1 and IOHMM2 perform in line with the baseline. Kloft et al. [7] considered the clickstream data for dropout classification in MOOCs environment using EMNLP 2014 dataset. The output of the review proposed feature selection and extraction pipeline for feature engineering.

Yukselturk et al. [36] review to investigate the data mining techniques for dropout prediction in which the data was collected through online questionnaires. The online questionnaire had ten attributes: age, gender, education level, previous online experience, coverage, prior knowledge, self-efficacy, occupation, and locus of control. The total number of participants was 189 students. The review employed four machine learning classifiers and used the Genetic Algorithm-based feature selection method. The results show 3NN as the best classifier while achieving 87% accuracy. Considering the larger dataset, the performance of ANN, DT, and BN is studied by [37]. The dataset used in this review contains data from 62,375 online learning students. The dataset's attributes were grouped into two categories, i.e., students' characteristics and academic performance. The final result indicated that the Decision Tree classifier reached a high accuracy score and overall effectiveness.

Dropout prediction at the early stage of the course can provide the management and instructors with early intervention. Sara et al. [41] used a large dataset from the Macom Lectio review administration system used by Danis high schools. The dataset included 72,598 instances, where each instance comprised 17 attributes values. Weka software was used as an implementation tool for RF, CART, SVM, and NB algorithms. The performance of classifiers was evaluated using accuracy and AUC, wherein RF reaches high values for both measures. Kostopoulos [42] work served as a pioneering review to

use semi-supervised machine learning techniques to predict student dropout. The KEEL software tool was then used to implement the semi-supervised learning methods, and the performances were compared. The dataset contained 244 instances with 12 attributes each. The results obtained from the investigation suggested Tri-Training Multi-Classified semi-supervised learning algorithm as the most effective method.

In recent years MOOCs platform has taken the center stage in education mining research [48,49,59,60]. All these studies focused on the early detection of students' dropouts. In [48], the authors shed light on the significance of temporal features in student dropout prediction. The temporal features captured the evolving characteristics of student performance using data obtained from quiz scores and information gathered from discussion forums through Canvas API. The extracted features from the data include; dropout week, number of discussion posts, number of forum views, number of quiz views, number of module views, social network degree, and active days. General Bayesian Network (GBN) and Decision Tree (DT) were the two classification approaches employed in classification.

In [59], the authors investigated deep learning models capable of automatic feature extraction from raw MOCs data. A deep learning method named "ConRec Network" was proposed by combining CNN and RNN, in which the feature extraction automatically takes place at the pooling layer. The proposed ConRec Network model attained high precision, recall, F-score, and ACU values. Liang and Zhen [49] analyzed data from student learning activities to measure the probabilities of student dropout in the next couple of days. The proposed framework comprised data collection from the XuetangX platform, data pre-processing, feature extraction, selection, and machine learning methods. The XuetangX online learning dataset covered 39 courses based on Open Edx. The data contained the students' behavior logs over 40 days. The log data need pre-processing so that it can be used to train ML algorithms. One hundred twelve features were extracted into three categories; user features, course features, and enrollment features. The dataset was then divided into training and testing sets containing 120,054 and 80,360 instances, respectively. Gradient Boosting Tree (GBT), SVM, and RF classifiers are used where GBT score high average AUC score.

The potential of ML for student dropout prediction at an early stage was also highlighted by [50,51]. University of Washington's student dataset containing 69,116 students record enrolled between the years 1998 and 2003 was used in experiments by [50]. From this dataset, 75.5% of the instances belonged to the graduate class, whereas the rest 24.5% belonged to the not-graduated class. The majority class was resampled using the "undersampling" technique by deleting random instances, and the number of instances is reduced to 16,269 for each class. The feature set contains demographic features, pre-college entry information, and transcript information. Logistic Regression (LG), RF, and KNN algorithms are used for classification, where LG performs better for accuracy and ROC values. They further added that GPAs in Math, Psychology, Chemistry, and English are potent predictors for student retention. In [51], the authors suggested an Early Warning System (EWS) based on machine learning methods. The authors proposed Grammar-based Genetic Programming (GBCP), a modified version of the Interpretable Classification Rule Mining (ICRM) proposed in 2013. ICRM2 algorithm can work on both balance and imbalance datasets. The author measures the performance of ICRM2 with SVM, NB, and DT and found ICRM as the best predictor even if the dataset is imbalanced.

Burgas et al. [51] analyze course grade data to predict the dropout. According to their experiments, they proposed that combining both the prediction and tutoring plan reduces the dropout rate by 14%. More recently, Gordner and Brook [66] proposed Model Selection Task (MST) for predictive model selection and extraction of features. They suggested two stages based on Friedman and Nemenyi statistical tests for model selection. This review collected data comprised of 298,909 students from the MOOCs platform for six online courses. This dataset contained 28 features grouped into three categories, i.e., clickstream features, academic features, and discussion forum features. CART and Adaboosting Tree classifiers were utilized for prediction purposes. This review concluded that the click-

stream feature was more beneficial while in the selection of the ML method, the Critical Distance between the classifier's performance was a better measure. Desmarais et al. [70] showed the importance of deep learning methods for dropout prediction and compared the performance with KNN, SVM, and DT algorithms. The deep learning algorithm achieved higher AUC and accuracy values than ML algorithms, where the dataset contained students' clickstream and forum discussion data having 13 total features.

Comparisons of Dropout Prediction Approaches

Early prediction of possible students' dropout is critical in determining necessary remedial measures. The most used approaches included identifying dropout features, curriculum and student performance, retention rate, dropout factors, and early prediction. The student's characteristics and academic performance were commonly used attributes by most researchers in determining the dropout features. Early prediction of potential student dropout was undertaken using both dynamic and static datasets. The commonly applied algorithms in dropout prediction were; DT, SVM, CART, KNN, and NB (Table 5).

Table 5. Prediction of student dropout using ML techniques and EDM methods.

Approach	Attributes	Algorithms	Count	References
Features for dropout prediction including temporal features	students' personal characteristics and academic performance	DT, LR, SVM, ARTMAP, RF, CART, and NB	10	[17,20,37] [41–43] [48,59,60] [49]
Curriculum-based and student performance-based features	Students performance class imbalance issues	K-NN, SMOTE	2	[17,21]
Retention rate	Freshman students	DT, Artificial Neural Networks (ANN)	2	[26,30]
Dropout factors	Evaluation of useful temporal models (Hidden Markov Model (HMM))	RNN combined with LSTM	3	[35,36,40]
Early-stage prediction of possible student dropout	pre-college entry information, and transcript information	ICRM2 with SVM, NB, DT, ID3, DL, and KNN, CART, and Adaboosting Tree	4	[26,51,66] [70]

3.3. Evaluation of Students' Performance Based on Static Data and Dynamic Data

The student performance data used to predict the students' performance can be categorized into two groups; (a) static data and (b) dynamic data. According to [27], the dynamic student performance data contain student success and failure logs gathered as they interact with the learning system. Student interaction logs with the e-learning system are an example of dynamic data as the characteristics of the dataset changes with time. On the other hand, static student performance data is acquired once and cannot change with time. An example includes students' enrolment and demographic data. The following sections present discussions on the usage of static and dynamic data in educational data mining.

Thaker et al. [27] proposed a dynamic student knowledge model framework for adaptive textbooks. The proposed framework utilizes student reading and quiz activity data to predict the students' current state of knowledge. The framework contains two advanced versions of the basic Behavioral Model (BM), i.e., (i) Behavior-Performance Model (BPM) and (ii) Individualized Behavior-Performance Model (IBPM). Feature Aware Student Knowledge Tracing (FAST) tool was used to implement the proposed models. The proposed approach achieved low RMSE and high ACU values comparing to the basic Behavior Model.

Carlos et al. [52] present a classification-based model to predict students' performance includes a data collection method to collect student learning and behavioural data from training activities. The SVM algorithm was used as a classification method that classifies the students into three categories based on their performance: high, medium, and low-

performance levels. Data of 336 students were collected with 61 features. Four experiments were conducted as follows; in experiment 1, only behavioral features were considered for classification; in experiment 2, only learning features were used; in experiment 3, learning and behavioral feature were combined for classification, and in experiment 4, only selected features were used for student performance prediction. Generally, the dataset contained eight behavioural features and 53 learning features, and students' performance was predicted over ten weeks. The results showed that the accuracy of the classifier increased in the subsequent week as the data grows. Furthermore, both the behavioral and learning features combined achieved high classification performance with an accuracy of 74.10% during week 10.

Desmarais et al. [70] proposed four-linear models based on matrix factorization using static student data for student's skill assessment where the performance of the proposed linear models was compared with well-known Item Response Theory (IRT) and k-nearest neighbor. In this review, three datasets were utilized, namely; (a) fraction algebra comprised of 20 questions and 149 students, (b) UNIX shell comprising of 34 questions and 48 students, and (c) college math comprised of 60 questions and 250 students. The experimental results showed that traditional IRT approaches attained higher accuracy than the proposed linear model and k-nearest neighbor approaches.

Application of Static and Dynamic Data Approaches

Early prediction of student performance, identification of at-risk students is essential in determining the potential dropout and accurate remedial measures. A total of 15 research studies used dynamic data for the students, such as student reading, quiz results, and activity logs from the e-learning system (Table 6). Only nine studies utilized static data that focused on enrolment details and demographic information, while 14 used both dynamic and static datasets. This indicates that the students' performances and activities on the learning platform provide much feedback needed for performance prediction. The commonly applied algorithms in early prediction using static and dynamic data were; KNN, NB, SVM, DT, RF, ID3, and ICRM2.

Table 6. Utility of static and dynamic data for early prediction of student performance.

Approach	Attributes	Algorithms	Count	References
Dynamic	Student performance data, student reading and quiz activity	K-NN, SMOTE, BM, SVM, NB, BN, DT, CR, ADTree, J48, and RF	15	[23,25,28,53,54,82] [21,47,57,62–64] [33,44,49,86]
Static	Enrolment and demographic data	Item Response Theory (IRT), ICRM2, SVM, NB, DT, ID3, DL, and KNN, CART and Adaboosting Tree	9	[6,26,30,34,39,51,66] [27,70]
Both	Pre-college entry information, and transcript information	ICRM2 with SVM, DL, ID3, KNN, DT, LR, SVM, ARTMAP, RF, CART, and NB	14	[17,20,26,37,41–43] [45,48,49,59,60,66,70]

3.4. Remedial Action Plan

Early identification of the student at risk is crucial and contributes to developing practical remedial actions, which further contributes to students' performance improvement. This section provided details of recent scholarly work on the remedial action plan to enhance the student output during their studies.

Ahadi et al. [52] suggested machine learning algorithms can detect the low and high-performance students at early stages and proposed an intervention plan during the programming of the course work. Early detection of low and high-performing students can benefit instructors to guide the struggling students and help them during their future studies. In this review, the authors evaluated the work presented by Jaduad, and Watson et al. [28] on the given dataset. Furthermore, they also used machine learning algorithms to predict

the low and high-performing students in the first week of an introductory programming course. The students' dataset of two semesters for introductory programming courses at Helsinki University was used. The dataset was collected using the Test My Code tool [44] to assess student performance automatically. A total of 296 students data from spring (86 students) and fall (210 students) semesters were divided into three groups as follows; (a) "an algorithmic programming question is given in the exam", (b) "the overall course", and (c) "combination of two". During classification, nine ML algorithms of three types were used. These include NB, BN (Bayesian), DT, Conjunctive Rule (CR), PART (Rule Learner), ADTree, J48, RF, and Decision Stump (DS) (Decision Tree). A total of 53 features were extracted from the dataset after applying three feature selection methods: best first method, genetic search, and greedy step-wise. The number of features was then reduced by eliminating those with low information gain. Weka data mining tool was used for classification and feature selection, and algorithm implementation. The results suggested that 88% to 93% accuracy is achieved by the classifiers when evaluated using 10-k fold cross-validation and percent split methods. They also concluded that machine learning approaches performed better than Jadwad and Watson et al. [28] methods. Moreover, the authors suggested additional practices for low-performing students, such as rehearsing and encouraging students to do more experiments rather than only correct experiments.

Jenhani et al. [87] proposed a classification-based remedial action plan that is built on a remedial action dataset. In this review, the authors first constructed remedial action datasets from different sources and multiple semesters where a set of supervised machine learning algorithms was applied to predict practical remedial actions. The proposed system helps the instructor to take appropriate remedial action. The system used was trained on historical data based on experts' and instructors' actions to improve the low learning outcome. Various sources collect the data, including "Blackboard LMS", legacy systems, and instructor gradings. Each instance in the dataset contains 13 attributes, whereas nine class labels represent remedial actions. The attributes included course code, course learning outcome (CLO), NQFDomain, gender, section size, course level, semester, Haslab, assessment. The nine classes of remedial actions used were CCES- Support-Center-and-Tutorial, Practice-Software, Improve class lab coordination, Revise Concept, Extra Quizzes, Practice Examples, Extra Assignments, Discussion Presentation, and Demos, Supplement Textbook and Materials. A total of 10 classifications algorithms were selected and used for classification using the Weka data mining tool, and all the classifiers achieved an average accuracy of 80%.

Elhassan et al. [31] proposed a remedial actions recommender system (RARS) to address student performance shortcomings. The proposed recommender system was based on a multi-label classification approach. This review was an extension of [87], where each instance in the dataset had more than one label. The dataset used contained 1008 instances where the average number of labels per instance was six with seven features each. Weka data mining tool was used for implementation purposes, where the dataset was first split into a 70:30 ratio. The 70% instances were used as the training set while the rest 30% were used as testing sets. Four wrapping methods were employed during the experimentation phase, including "Binary Relevance", "Classifier Chain (CC)", "Random-Kla bEL (RK)", and "Rank+ Threshold (RT)". Classification algorithms C4.5, NB and K-NN, are used for wrapper methods. The performance of the classifiers is evaluated using hamming loss, zero-one loss, One-Error loss, and average accuracy. The results concluded that Decision Tree C4.5 had low error loss (0.0) and a high average accuracy of 98.4% for the Binary-Relevance (BR) wrapper method.

Burgos et al. [51] investigated the use of the knowledge discovery technique and proposed a tutoring action plan to reduce the dropout rate by 14%. Logistic Regression models are used as predictive methods for detecting the potential student dropout using the activity grades. The proposed prediction method uses an iterative function that assessed students' performance every week. The performance of the proposed LOGIT-Act method was compared with SVM, FFNN, PESFAM, and SEDM algorithms. The proposed algorithm

attains high accuracy, precision, recall, and specificity scores of 97.13%, 98.95%, 96.73%, and 97.14%, respectively. This research review also suggested a weekly tutoring action plan that can prevent students from dropping out. The proposed action plan included;

- Courtesy call at the start of the academic year
- The public message of welcome to the course via a virtual classroom
- The video conference welcoming session
- Email to potential dropout
- A telephone call to potential dropout
- A telephone call to potential dropout (from one or more courses)

Remedial Action Approaches

This research has revealed that early detection based on students' performance is significant in determining the required remedial measures. On the other hand, remedial actions are undertaken using the course characteristics and technologies of the e-learning system. The review also revealed that standard early detection and the remedial algorithm was NB, as most earlier studies exploited NB for the task and achieved significant results.

Overall, DT, NB, and SVM algorithms were applied for performance and during dropout predictions using static and dynamic data. Table 7 provides summary of remedial action approaches. Figure 4 shows the common methodological approach used by the majority of the evaluated research studies.

Table 7. Remedial action based on prediction results.

Approach	Attributes	Algorithms	Count	References
Early detection	Student performance	NB, BN, DT, CR, PART (Rule Learner), ADTree, J48, and RF	12	[17,20,28,37,41,42] [43,44,48,49,59,60]
Remedial action	course code, course learning outcome (CLO), NQFDomain, gender, section size, course level, semester, Haslab, assessment, U, M, A, E.	RARS, C4.5, NB and K-NN	7	[26,31,51,66,70] [45]

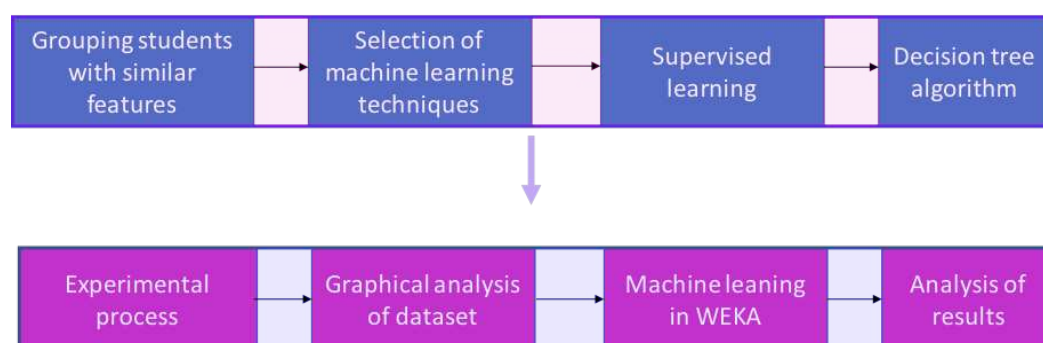


Figure 4. Methodological approach adopted by most of the research studies.

4. Discussion and Critical Review

In order to address our first two research questions, we collected the studies that are trying to address the problems. We identified the problems and their solutions in the literature. To answer the third question, the overall research productivity is shown using country-wise distribution in Figure 5, conference/journal wise distribution in Figure 6 and year-wise distribution in Figure 7 of the studies that are included in the review. It can be observed that the research community from Germany and UK focused more on the field than the other countries. Similarly, 2018 was a hot year for the student's performance prediction topic overall. Meanwhile, more journals focused on the topic as compared to the conferences.

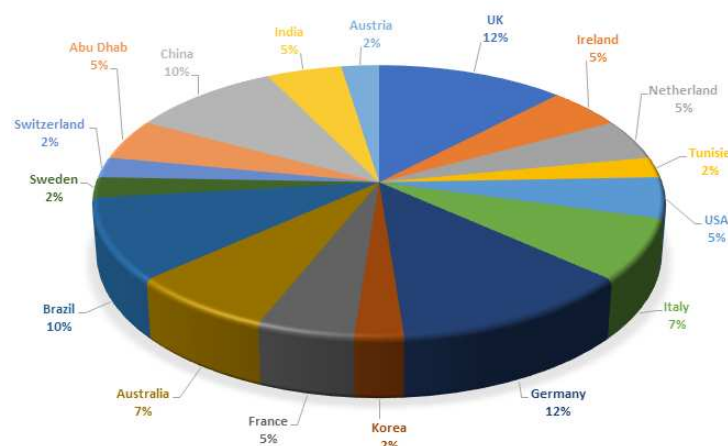


Figure 5. Contry wise distribution of included publications.

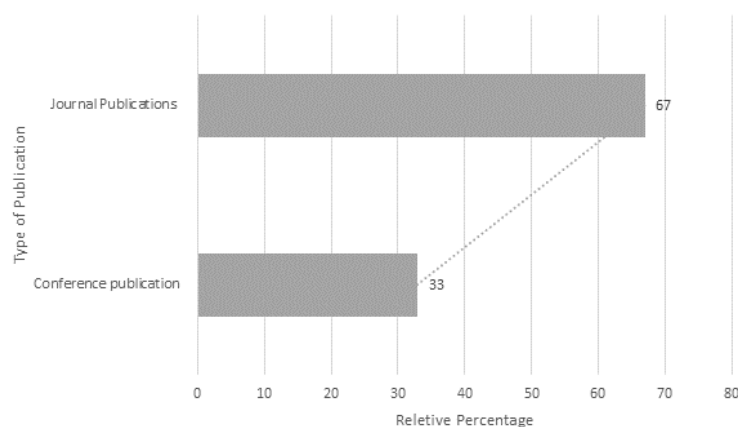


Figure 6. Conference vs. journal wise distribution of included publications.

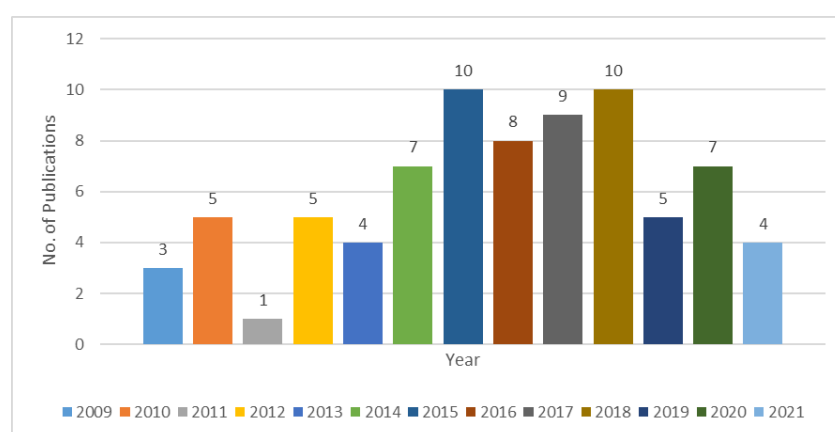


Figure 7. Year wise distribution of included publications.

This paper presents an overview of the machine learning technique used in educational data mining, focusing on two critical aspects; (a) accurate prediction of students at risk and (b) accurate prediction of student dropout. Following an extensive literature review of critical publications between 2009 and 2021, the following conclusions are made;

- Most studies used minimal data to train the machine learning methods. However, it is a fact that ML algorithms need massive data in order to perform accurately.

- The review also revealed that a few studies have focused on class balancing or data balancing. Class balancing is mainly considered important in obtaining high classification performance [50].
- The temporal nature of features used for at-risk and dropout students' predictions has not been studied to its potential. The values of these features change with time due to their dynamic nature. Incorporating temporal features for classification has the ability to enhance the predictor performance [40,48,67]. Khan et al. [67] examine the temporal features for text classification.
- It was also observed that the prediction of students at-risk and dropout studies for on-campus students utilized the dataset with a very minimal number of instances. Machine learning algorithms trained on small datasets might not achieve satisfactory results. Moreover, the data pre-processing technique can contribute significantly to more accurate results.
- Most of the research studies tackled the problem as a classification task. Whereas very few studies focused on clustering algorithms that detected the classes of students' in the dataset. Furthermore, the problems mentioned above are treated as binary classification while several other classes would be introduced to help the management develop more effective intervention plans.
- Less attention has been paid to feature engineering tasks, where the types of features can influence the predictor's performance. Three features were primarily used in the studies, i.e., students' demographics, academic, and e-learning interaction session logs.
- It was also observed that most of the studies used traditional machine learning algorithms such as SVM, DT, NB, KNN, etc., and only a few have investigated the potential of deep learning algorithms.
- Last but not least, the current literature does not consider the dynamic nature of student performance. The students' performance is an evolving process and improves or drops steadily. The performance of predictors on real-time dynamic data is yet to be explored.

As a result, ML has all the potential to speed up the progress in the educational field and it can be noticed that the efficiency of education grows significantly. By applying ML techniques in educational field in a proper and efficient way, this will transform education and fundamentally changing teaching, learning, and research. Educators who are using ML will gain a better understanding of how their students are progressing with learning, therefore will be able to help struggling students earlier and take action to improve success and retention.

5. Conclusions

With recent advancements in data acquisition systems and system performance indicators, educational systems are now studied more effectively yet with much less effort. State-of-the-art data mining and machine learning techniques have been proposed for analyzing and monitoring massive data giving rise to a whole new field of big data analytics. Overall, this review achieved its objectives of enhancing the students' performance by predicting students' at-risk and dropout, highlighting the importance of using both static and dynamic data. This will provide the basis for new advances in Educational Data Mining using machine learning and data mining approaches. However, only a few studies proposed remedial solutions to provide in-time feedback to students, instructors, and educators to address the problems. Future research will focus more on developing a efficient ensemble method to practically deploy the ML-based performance prediction methodology and search for dynamic ways or methods to predict students' performance and provide automatic needed remedial actions to help the students as early as possible. Finally, we emphasize the promising directions for future research using ML techniques in predicting students performance. We are looking to implement some of the excellent existing works and focusing more on dynamic nature of students performance. As a

result, the instructors can gain more hints to build up proper interventions for learners and achieve precision education targets..

Author Contributions: Conceptualization, methodology, software, statistical analysis, writing original draft preparation: B.A., N.Z. and H.A.; data curation: B.A.; writing review and editing: literature review, discussion: B.A., N.Z. and H.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

RF	Random Forest
LG	Logestic Regression
NN	Neural Network
SVM	Support Vector Machine
MLP	Multi layer Perceptron
DT	Decision Tree
NB	Naive Bayes
KNN	K-nearest neighbors
SMOTE	Synthetic Minority Over-sampling Technique

References

- Romero, C.; Ventura, S.; Pechenizkiy, M.; Baker, R.S. *Handbook of Educational Data Mining*; CRC Press: Boca Raton, FL, USA, 2010.
- Hernández-Blanco, A.; Herrera-Flores, B.; Tomás, D.; Navarro-Colorado, B. A systematic review of deep learning approaches to educational data mining. *Complexity* **2019**, *2019*, 1306039. [[CrossRef](#)]
- Bengio, Y.; Lecun, Y.; Hinton, G. Deep Learning for AI. *Commun. ACM* **2021**, *64*, 58–65. [[CrossRef](#)]
- Lykourantzou, I.; Giannoukos, I.; Mpardis, G.; Nikolopoulos, V.; Loumos, V. Early and dynamic student achievement prediction in e-learning courses using neural networks. *J. Am. Soc. Inf. Sci. Technol.* **2009**, *60*, 372–380. [[CrossRef](#)]
- Kuzilek, J.; Hlostá, M.; Herrmannová, D.; Zdrahal, Z.; Wolff, A. OU Analyse: Analysing at-risk students at The Open University. *Learn. Anal. Rev.* **2015**, *2015*, 1–16.
- He, J.; Bailey, J.; Rubinstein, B.I.; Zhang, R. Identifying at-risk students in massive open online courses. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015.
- Kloft, M.; Stiehler, F.; Zheng, Z.; Pinkwart, N. Predicting MOOC dropout over weeks using machine learning methods. In *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*; Department of Computer Science, Humboldt University of Berlin: Berlin, Germany, 2014; pp. 60–65.
- Alapont, J.; Bella-Sanjuán, A.; Ferri, C.; Hernández-Orallo, J.; Llopis-Llopis, J.; Ramírez-Quintana, M. Specialised tools for automating data mining for hospital management. In Proceedings of the First East European Conference on Health Care Modelling and Computation, Craiova, Romania, 31 August–2 September 2005; pp. 7–19.
- Hellas, A.; Ihantola, P.; Petersen, A.; Ajanovski, V.V.; Gutica, M.; Hynninen, T.; Knutas, A.; Leinonen, J.; Messom, C.; Liao, S.N. Predicting academic performance: A systematic literature review. In Proceedings of the Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education, Larnaca, Cyprus, 2–4 July 2018; pp. 175–199.
- Alyahyan, E.; Düşteğör, D. Predicting academic success in higher education: Literature review and best practices. *Int. J. Educ. Technol. High. Educ.* **2020**, *17*, 1–21. [[CrossRef](#)]
- Namoun, A.; Alshanqiti, A. Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Appl. Sci.* **2021**, *11*, 237. [[CrossRef](#)]
- Okoli, C. A guide to conducting a standalone systematic literature review. *Commun. Assoc. Inf. Syst.* **2015**, *37*, 43. [[CrossRef](#)]
- Kitchenham, B. *Procedures for Performing Systematic Reviews*; Keele University: Keele, UK, 2004; Volume 33, pp. 1–26.
- Piper, R.J. How to write a systematic literature review: A guide for medical students. *Natl. AMR Foster. Med. Res.* **2013**, *1*, 1–8.
- Bhandari, M.; Guyatt, G.H.; Montori, V.; Devereaux, P.; Swiontkowski, M.F. User's guide to the orthopaedic literature: How to use a systematic literature review. *JBJS* **2002**, *84*, 1672–1682. [[CrossRef](#)]
- Loumos, V. Dropout prediction in e-learning courses through the combination of machine learning techniques. *Comput. Educ.* **2009**, *53*, 950–965.
- Kotsiantis, S. Educational data mining: A case study for predicting dropout-prone students. *Int. J. Knowl. Eng. Soft Data Paradig.* **2009**, *1*, 101–111. [[CrossRef](#)]
- Kovacic, Z. Early Prediction of Student Success: Mining Students' Enrolment Data. In Proceedings of the Informing Science and Information Technology Education Joint Conference, Cassino, Italy, 19–24 June 2010.

19. Kotsiantis, S.; Patriarcheas, K.; Xenos, M. A combinational incremental ensemble of classifiers as a technique for predicting students. *Perform. Distance Educ. Knowl.-Based Syst.* **2010**, *23*, 529–535. [CrossRef]
20. Quadri, M.; Kalyankar, N. Drop out feature of student data for academic performance using decision tree techniques. *Glob. J. Comput. Sci. Technol.* **2010**, *10*. ISSN 0975-4172. Available online: <https://computerresearch.org/index.php/computer/article/view/891> (accessed on 15 August 2021).
21. Marquez-Vera, C.; Romero, C.; Ventura, S. Predicting school failure using data mining. In Proceedings of the 4th International Conference on Educational Data Mining, Eindhoven, The Netherlands, 6–8 July 2011.
22. Galbraith, C.; Merrill, G.; Kline, D. Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business-related classes? a neural network and Bayesian analyses. *Res. High Educ.* **2011**, *53*, 353–374. [CrossRef]
23. Kotsiantis, S.B. Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. *Artif. Intell. Rev.* **2012**, *37*, 331–344. [CrossRef]
24. Osmanbegovic, E.; Suljic, M. Data mining approach for predicting student performance. *Econ. Rev. J. Econ. Bus.* **2012**, *10*, 3–12.
25. Baradwaj, B.K.; Pal, S. Mining educational data to analyze students' performance. *arXiv* **2012**, arXiv:1201.3417.
26. Pal, S. Mining educational data to reduce dropout rates of engineering students. *Int. J. Inf. Eng. Electron. Bus.* **2012**, *4*, 1–7. [CrossRef]
27. Thaker, K.; Huang, Y.; Brusilovsky, P.; Daqing, H. Dynamic knowledge modeling with heterogeneous activities for adaptive textbooks. In Proceedings of the 11th International Conference on Educational Data Mining, Buffalo, NY, USA, 15–18 July 2018.
28. Watson, C.; Li, F.W.; Godwin, J.L. Predicting performance in an introductory programming course by logging and analyzing student programming behavior. In Proceedings of the IEEE 13th International Conference on Advanced Learning Technologies, Beijing, China, 15–18 July 2013; pp. 319–323.
29. Márquez-Vera, C.; Cano, A.; Romero, C.; Ventura, S. Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. *Appl. Intell.* **2013**, *38*, 315–330. [CrossRef]
30. Plagge, M. Using artificial neural networks to predict the first-year traditional students' second-year retention rates. In Proceedings of the 51st ACM Southeast Conference, Savannah, GA, USA, 4–6 April 2013.
31. Elhassan, A.; Jenhani, I.; Brahim, G. Remedial actions recommendation via multi-label classification: A course learning improvement method. *Int. J. Mach. Learn. Comput.* **2018**, *8*, 583–588.
32. Hu, Y.H.; Lo, C.L.; Shih, S.P. Developing early warning systems to predict students. *Online Learn. Perform. Comput. Hum. Behav.* **2014**, *36*, 469–478. [CrossRef]
33. Villagra-Arnedo, C.J.; Gallego-Duran, F.; Compan, P.; Largo, F.; Molina-Carmona, R. Predicting Academic Performance from Behavioral and Learning Data. 2016. Available online: <http://hdl.handle.net/10045/57216> (accessed on 2 January 2021).
34. Wolff, A. Modelling student online behavior in a virtual learning environment. *arXiv* **2018**, arXiv:1811.06369.
35. Ye, C.; Biswas, G. Early prediction of student dropout and performance in MOOCs using higher granularity temporal information. *J. Learn. Anal.* **2014**, *1*, 169–172. [CrossRef]
36. Yukselturk, E.; Ozekes, S.; Turel, Y. Predicting dropout student: an application of data mining methods in an online education program. *Eur. J. Open Distance e-Learn.* **2014**, *17*, 118–133. [CrossRef]
37. Tan, M.; Shao, P. Prediction of student dropout in e-learning program through the use of machine learning method. *Int. J. Emerg. Technol. Learn. (ijET)* **2015**, *10*, 11–17. [CrossRef]
38. Lakkaraju, H.; Aguiar, E.; Shan, C.; Miller, D.; Bhanpuri, N.; Ghani, R.; Addison, K. A machine learning framework to identify students at risk of adverse academic outcomes. In Proceedings of the 21st ACM SIGKDD, International Conference on Knowledge Discovery and Data, Sydney, NSW, Australia, 10–13 August 2015.
39. Ahmad, F.; Ismail, N.; Aziz, A. The prediction of students academic performance using classification data mining techniques. *Appl. Math. Sci.* **2015**, *9*, 6415–6426. [CrossRef]
40. Fei, M.; Yeung, D.Y. Temporal models for predicting student dropout in massive open online courses. In Proceedings of the IEEE International Conference on Data Mining Workshop (ICDMW), Atlantic City, NJ, USA, 14–17 November 2015.
41. Sara, N.B.; Halland, R.; Igel, C.; Alstrup, S. High-school dropout prediction using machine learning: A danish large-scale study. In Proceedings of the Eu-European Symposium on Artificial Neural Networks, Computational Intelligence, Bruges, Belgium, 22–24 April 2015.
42. Kostopoulos, G.; Kotsiantis, S.; Pintelas, P. Estimating student dropout in distance higher education using semi-supervised techniques. In Proceedings of the 19th Panhellenic Conference on Informatics, Athens, Greece, 1–3 October 2015; pp. 38–43.
43. Xing, W.; Chen, X.; Stein, J.; Marcinkowski, M. Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization. *Comput. Hum. Behav.* **2016**, *58*, 119–129. [CrossRef]
44. Vihavainen, A.; Vikberg, T.; Luukkainen, M.; Pärtel, M. Scaffolding students' learning using test my code. In Proceedings of the 18th ACM Conference on Innovation and Technology in Computer Science Education, Canterbury, England, UK, 1–3 July 2013; pp. 117–122.
45. Ahmed, A.; Elaraby, I. Data mining: A prediction for student's performance using classification method. *World J. Comput. Appl. Technol.* **2014**, *2*, 43–47. [CrossRef]
46. Al-Barrak, M.; Al-Razgan, M. Predicting Students' final GPA using decision trees: A case study. *Int. J. Inf. Educ. Technol.* **2016**, *6*, 528. [CrossRef]

47. Marbouti, F.; Diefes-Dux, H.; Madhavan, K. Models for early prediction of at-risk students in a course using standards-based grading. *Comput. Educ.* **2016**, *103*, 1–15. [\[CrossRef\]](#)
48. Wang, W.; Yu, H.; Miao, C. Deep model for dropout prediction in MOOCs. In Proceedings of the 2nd International Conference on Crowd Science and Engineering, Beijing, China, 6–9 July 2017; pp. 26–32.
49. Aulck, L.; Velagapudi, N.; Blumenstock, J.; West, J. Predicting student dropout in higher education. *arXiv* **2016**, arXiv:1606.06364.
50. Marquez-Vera, C.; Cano, A.; Romero, C.; Noaman, A.; Fardoun, H.; Ventura, S. Early dropout prediction using data mining: A case study with high school students. *Expert Syst.* **2016**, *33*, 107–124. [\[CrossRef\]](#)
51. Burgos, C.; Campanario, M.; de la Pena, D.; Lara, J.; Lizcano, D.; Martinez, M. Data mining for modeling students performance: A tutoring action plan to prevent academic dropout. *Comput. Electr. Eng.* **2017**, *66*, 541–556. [\[CrossRef\]](#)
52. Ahadi, A.; Lister, R.; Haapala, H.; Vihavainen, A. Exploring machine learning methods to automatically identify students. need of assistance. In Proceedings of the eleventh annual. International Conference on International Computing Education Research, Omaha, NE, USA, 9–13 July 2015; pp. 121–130.
53. Iqbal, Z.; Qadir, J.; Mian, A.; Kamiran, F. Machine learning-based student grade prediction: A case study. *arXiv* **2017**, arXiv:1708.08744.
54. Zhang, W.; Huang, X.; Wang, S.; Shu, J.; Liu, H.; Chen, H. Student performance prediction via online learning behavior analytics. In Proceedings of the International Symposium on Educational Technology (ISET), Hong Kong, China, 27–29 June 2017.
55. Almarabeh, H. Analysis of students' performance by using different data mining classifiers. *Int. J. Mod. Educ. Comput. Sci.* **2017**, *9*, 9. [\[CrossRef\]](#)
56. Xu, J.; Moon, K.; Schaar, M.D. A machine learning approach for tracking and predicting student performance in degree programs. *IEEE J. Sel. Top. Signal Process.* **2017**, *11*, 742–753. [\[CrossRef\]](#)
57. Al-Shehri, H.; Al-Qarni, A.; Al-Saati, L.; Batoaq, A.; Badukhen, H.; Alrashed, S.; Alhiyafi, J.; Olatunji, S. Student performance prediction using support vector machine and k-nearest neighbor. In Proceedings of the 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Windsor, ON, Canada, 30 April–3 May 2017.
58. Alowibdi, J. Predicting student performance using advanced learning analytics. In Proceedings of the 26th International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee, Perth, Australia, 3–7 April 2017; pp. 415–421.
59. Nagrecha, S.; Dillon, J.; Chawla, N. Mooc dropout prediction: Lessons learned from making pipelines interpretable. In Proceedings of the 26th International Conference, World Wide Web Companion, International World Wide Web Conferences Steering Committee, Perth, Australia, 3–7 April 2017; pp. 351–359.
60. Liang, J.; Li, C.; Zheng, L. Machine learning application in MOOCs: Dropout prediction. In Proceedings of the 11th International Conference on Computer Science & Education (ICCSE), Nagoya, Japan, 23–25 August 2016; pp. 52–57.
61. Al-Obeidat, F.; Tubaishat, A.; Dillon, A.; Shah, B. Analyzing students performance using multi-criteria classification. *Clust. Comput.* **2018**, *21*, 623–632. [\[CrossRef\]](#)
62. Kaviyarasi, R.; Balasubramanian, T. Exploring the high potential factors that affect students. *Acad. Perform. Int. J. Educ. Manag. Eng.* **2018**, *8*, 15.
63. Zaffar, M.; Iskander, S.; Hashmani, M. A study of feature selection algorithms for predicting students academic performance. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 541–549. [\[CrossRef\]](#)
64. Chui, K.; Fung, D.; Lytras, M.; Lam, T. Predicting at-risk university students in a virtual learning environment via a machine learning algorithm. *Comput. Hum. Behav.* **2020**, *107*, 105584. [\[CrossRef\]](#)
65. Masci, C.; Johnes, G.; Agasisti, T. Student and school performance across countries: A machine learning approach. *Eur. J. Oper. Res.* **2018**, *269*, 1072–1085. [\[CrossRef\]](#)
66. Xing, W.; Du, D. Dropout prediction in MOOCs: Using deep learning for personalized intervention. *J. Educ. Comput. Res.* **2019**, *57*, 547–570. [\[CrossRef\]](#)
67. Khan, S.; Islam, M.; Aleem, M.; Iqbal, M. Temporal specificity-based text classification for information retrieval. *Turk. J. Electr. Eng. Comput. Sci.* **2018**, *26*, 2915–2926. [\[CrossRef\]](#)
68. Livieris, I.; Drakopoulou, K.; Tampakas, V.; Mikropoulos, T.; Pintelas, P. Predicting secondary school students. *Perform. Util. Semi-Supervised Learn. Approach J. Educ. Comput. Res.* **2019**, *57*, 448–470.
69. Nieto, Y.; García-Díaz, V.; Montenegro, C.; Crespo, R.G. Supporting academic decision making at higher educational institutions using machine learning-based algorithms. *Soft Comput.* **2019**, *23*, 4145–4153. [\[CrossRef\]](#)
70. Desmarais, M.; Naceur, R.; Beheshti, B. Linear models of student skills for static data. In *UMAP Workshops*; Citeseer: University Park, PA, USA, 2012.
71. Oyedeji, A.O.; Salami, A.M.; Folorunsho, O.; Abolade, O.R. Analysis and Prediction of Student Academic Performance Using Machine Learning. *JITCE (J. Inf. Technol. Comput. Eng.)* **2020**, *4*, 10–15. [\[CrossRef\]](#)
72. Alhusban, S.; Shatnawi, M.; Yasin, M.B.; Hmeidi, I. Measuring and Enhancing the Performance of Undergraduate Student Using Machine Learning Tools. In Proceedings of the 2020 11th International Conference on Information and Communication Systems (ICICS), Copenhagen, Denmark, 24–26 August 2020; pp. 261–265.
73. Gafarov, F.; Rudneva, Y.B.; Sharifov, U.Y.; Trofimova, A.; Bormotov, P. Analysis of Students' Academic Performance by Using Machine Learning Tools. In Proceedings of the International Scientific Conference "Digitalization of Education: History, Trends and Prospects" (DETP 2020), Yekaterinburg, Russia, 23–24 April 2020; Atlantis Press: Paris, France, 2020; pp. 574–579.

74. Walia, N.; Kumar, M.; Nayar, N.; Mehta, G. Student's Academic Performance Prediction in Academic using Data Mining Techniques. In *Proceedings of the International Conference on Innovative Computing & Communications (ICICC)*; Springer: Berlin/Heidelberg, Germany, 2020.
75. Wakelam, E.; Jefferies, A.; Davey, N.; Sun, Y. The potential for student performance prediction in small cohorts with minimal available attributes. *Br. J. Educ. Technol.* **2020**, *51*, 347–370. [[CrossRef](#)]
76. Hussain, K.; Talpur, N.; Aftab, M.U.; NoLastName, Z. A Novel Metaheuristic Approach to Optimization of Neuro-Fuzzy System for Students' Performance Prediction. *J. Soft Comput. Data Min.* **2020**, *1*, 1–9. [[CrossRef](#)]
77. Ghorbani, R.; Ghousi, R. Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques. *IEEE Access* **2020**, *8*, 67899–67911. [[CrossRef](#)]
78. Aggarwal, D.; Mittal, S.; Bali, V. Significance of Non-Academic Parameters for Predicting Student Performance Using Ensemble Learning Techniques. *Int. J. Syst. Dyn. Appl. (IJSDA)* **2021**, *10*, 38–49.
79. Zeineddine, H.; Braendle, U.; Farah, A. Enhancing prediction of student success: Automated machine learning approach. *Comput. Electr. Eng.* **2021**, *89*, 106903. [[CrossRef](#)]
80. OuahiMariame, S.K. Feature Engineering, Mining for Predicting Student Success based on Interaction with the Virtual Learning Environment using Artificial Neural Network. *Ann. Rom. Soc. Cell Biol.* **2021**, *25*, 12734–12746.
81. Hussain, M.; Zhu, W.; Zhang, W.; Abidi, S.; Ali, S. Using machine learning to predict student difficulties from learning session data. *Artif. Intell. Rev.* **2019**, *52*, 381–407. [[CrossRef](#)]
82. Thai-Nghe, N.; Drumond, L.; Krohn-Grimberghe, A.; Schmidt-Thieme, L. Recommender system for predicting student performance. *Procedia Comput. Sci.* **2010**, *1*, 2811–2819. [[CrossRef](#)]
83. Buenaño-Fernández, D.; Gil, D.; Luján-Mora, S. Application of machine learning in predicting performance for computer engineering students: A case study. *Sustainability* **2019**, *11*, 2833. [[CrossRef](#)]
84. Reddy, P.; Reddy, R. Student Performance Analyser Using Supervised Learning Algorithms. 2021. Available online: <https://easychair.org/publications/preprint/QhZK> (accessed on 4 August 2021).
85. Acharya, A.; Sinha, D. Early prediction of students performance using machine learning techniques. *Int. J. Comput. Appl.* **2014**, *107*, 37–43. [[CrossRef](#)]
86. Muzamal, J.H.; Tariq, Z.; Khan, U.G. Crowd Counting with respect to Age and Gender by using Faster R-CNN based Detection. In *Proceedings of the 2019 International Conference on Applied and Engineering Mathematics (ICAEM)*, Taxila, Pakistan, 27–29 August 2019, Volume 10, pp. 157–161.
87. Jenhani, I.; Brahim, G.; Elhassan, A. Course learning outcome performance improvement: A remedial action classification-based approach. In *Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Anaheim, CA, USA, 18–20 December 2016; pp. 408–413.