

Review

# Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review

Juan L. Rastrollo-Guerrero, Juan A. Gómez-Pulido \*  and Arturo Durán-Domínguez

Escuela Politécnica, Universidad de Extremadura, 10003 Cáceres, Spain; juanluisrg@unex.es (J.L.R.-G.); arduaran@unex.es (A.D.-D.)

\* Correspondence: jangomez@unex.es

Received: 25 November 2019; Accepted: 23 January 2020; Published: 4 February 2020



**Abstract:** Predicting students' performance is one of the most important topics for learning contexts such as schools and universities, since it helps to design effective mechanisms that improve academic results and avoid dropout, among other things. These are benefited by the automation of many processes involved in usual students' activities which handle massive volumes of data collected from software tools for technology-enhanced learning. Thus, analyzing and processing these data carefully can give us useful information about the students' knowledge and the relationship between them and the academic tasks. This information is the source that feeds promising algorithms and methods able to predict students' performance. In this study, almost 70 papers were analyzed to show different modern techniques widely applied for predicting students' performance, together with the objectives they must reach in this field. These techniques and methods, which pertain to the area of Artificial Intelligence, are mainly Machine Learning, Collaborative Filtering, Recommender Systems, and Artificial Neural Networks, among others.

**Keywords:** prediction; students' performance; dropout; machine learning; supervised learning; unsupervised learning; collaborative filtering; recommender systems; artificial neural networks; deep learning

## 1. Introduction

There is often a great need to be able to predict future students' behavior in order to improve curriculum design and plan interventions for academic support and guidance on the curriculum offered to the students. This is where Data Mining (DM) [1] comes into play. DM techniques analyze datasets and extract information to transform it into understandable structures for later use. Machine Learning (ML), Collaborative Filtering (CF), Recommender Systems (RS) and Artificial Neural Networks (ANN) are the main computational techniques that process this information to predict students' performance, their grades or the risk of dropping out of school.

Nowadays, there is a considerable amount of research and studies that follow along the lines of predicting students' behaviour, among other related topics of interest in the educational area. Indeed, many articles have been published in journals and presented in conferences on this topic. Therefore, the main goal of this study is to present an in depth overview of the different techniques and algorithms proposed that have been applied to this subject.

## 2. Methodology

This article is the result of a qualitative research study of 64 recent articles (almost 90% were published in the last 6 years) related to the different techniques applied for predicting students' behaviour. The literature considered for this study stems from different book chapters, journals and conferences. IEEE, Science Direct, Springer, IEEE Computer Society, iJET, ACM Digital Library,

Taylor & Francis Online, JEO, Sage Journals, J-STAGE, Inderscience Publishers, WIT Press, Science Publications, EJER, and Wiley Online Library were some of the online databases consulted to extract the corresponding literature.

We have excluded papers without enough quality or contribution. The journal papers without an impact factor listed in the ISI Journal Citation Report or not peer-reviewed were excluded. The conference papers corresponding with conferences not organized/supported/published by IEEE, ACM, Springer or renowned organizations and editorials were excluded too. As a result, 35% of the papers analyzed correspond to journal articles; of these, 64% have JCR impact factor and the rest correspond to peer-reviewed journals indexed in other scientific lists.

For the search processes used for these databases we mainly considered the following descriptors: "Predicting students' performance", "Predicting algorithm students", "Machine learning prediction students", "Collaborative filtering prediction students", "Recommender systems prediction students", "Artificial neural network prediction students", "Algorithms analytics students" and "Students analytics prediction performance", among other similar terms.

The literature review provided throughout this article is mainly classified from two points of view: techniques and objectives. We describe the techniques first in this article, since they are applied to reach the objectives considered in each reference. These techniques, in turn, are implemented by means of several algorithmic methods.

Table 1 summarizes the main features of the literature review, showing four groups of columns: students' level, objectives, techniques, and algorithms and methods.

- Students' level: Each reference analyzes datasets built from students of a particular level. We consider a classification of wide levels, corresponding to School (S), High School (HS) and University (U).
- Objectives: The objectives are connected to the interests and risks in the students' learning processes.
- Techniques: The techniques consider the different algorithms, methods and tools that process the data to analyze and predict the above objectives.
- Algorithms and methods: The main algorithms and computational methods applied in each case are detailed in the Table 1. Other algorithms with related names or versions not shown in this table could be also applied. The shadowed cells corresponds with the best algorithms found when several methods were compared for the same purpose.

Figure 1 presents graphically the basic statistics about the techniques, objectives, type of students, and algorithms considered in the literature review. These graphs are built from Table 1 in order to understand better the impact of the literature review that is explained in the next sections.

A first consideration about predicting students' performance by means of ML is the academic level of the students. This information can be useful to know because the datasets built from the students' behaviour imply latent factors that can be different according to the academic level. As we can see in Figure 1, most of the cases correspond to the university level, followed by the high-school level.





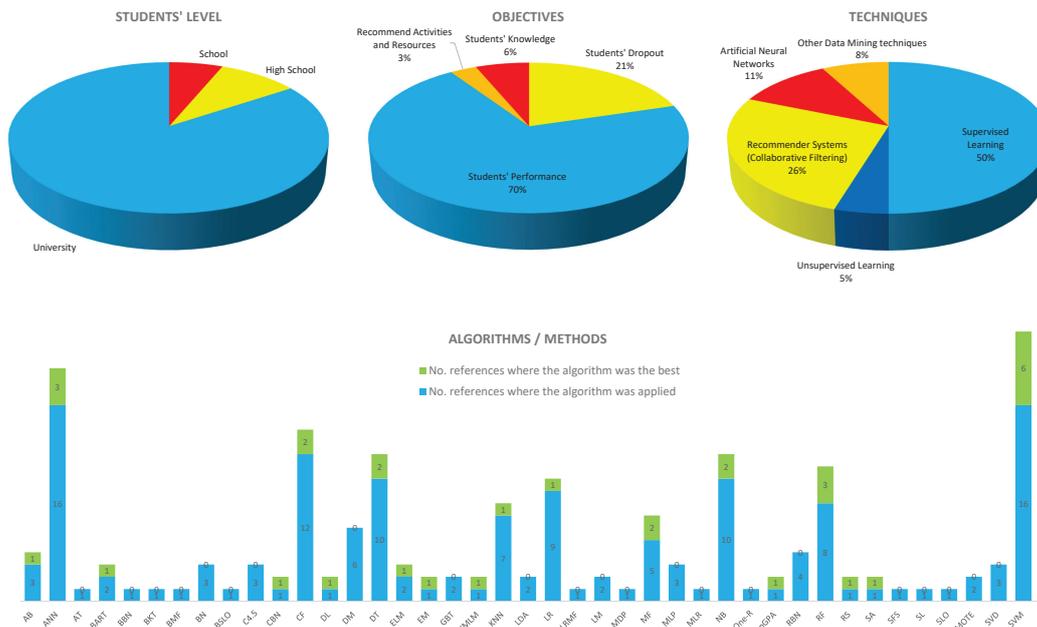


Figure 1. Basic statistics about the techniques, objectives and algorithms tackled in the literature review.

### 3. Techniques

The application of techniques such as ML, CF, RS, and ANN to predict students' behavior take into account different types of data, for example, demographic characteristics and the grades from some tasks. A good starting point was the study conducted by the Hellenic Open University, where several machine-supervised learning algorithms were applied to a particular dataset. This research found that the Naïves Bayes (NB) algorithm was the most appropriate for predicting both performance and probability of student dropout [2]. Nevertheless, each case study has its own characteristics and nature, hence different techniques can be selected as the best option to predict students' behaviour.

We have gathered the different techniques into main four groups: supervised ML, unsupervised ML, CF and ANN. An additional group dealing with other DM techniques is added in order to include some works where similar objectives were tackled. Figure 1 shows the weight amount of each of these groups of techniques in the literature, which can indicate the number of problems and cases where each technique is more suitable. In this sense, supervised ML makes up almost half of the cases, followed by CF with a quarter. On the contrary, unsupervised ML has been applied in very few cases.

#### 3.1. Machine Learning

Machine Learning is a set of techniques that gives computers the ability to learn without the intervention of human programming [3]. ML has supported a wide range of applications such as medical diagnostics, stock market analysis, DNA sequence classification, games, robotics, predictive analysis, etc. We are particularly interested in the area of predictive analysis, where ML allows us to implement complex models that are used for prediction purposes. These models can be of great help to users by providing relevant data to facilitate decision-making.

ML algorithms are classified into two main streams: supervised and unsupervised.

##### 3.1.1. Supervised Learning

Supervised Learning (SL) seeks algorithms able to reason from instances externally supplied in order to produce general hypotheses, which then make predictions about future instances [66]. In other

words, the goal of SL is to build a clear model of the distribution of class labels in terms of predictor characteristics.

Rule Induction is an efficient SL method to make predictions, which was able to reach an accuracy level of 94% when predicting dropout of new students in nursing courses, from 3978 records on 528 students [4].

When using classification techniques, it is necessary to be careful if there are unbalanced datasets, since they can produce misleading predictive accuracy. For this purpose, several improvements were proposed in [5] when predicting dropout, such as exploring a wide range of learning methods, selecting attributes, evaluating the effectiveness of theory, and studying factors between dropout and non-dropout students. The classifier algorithms explored in this study were One-R, C4.5, ADTrees, NB, BN, and Radial Basis Networks (RBN). In this sense, applying several algorithms and comparing their results will be always very useful, as in [6], where four classification algorithms (Logistic Regression (LR) [67], DT, ANN, and SVM) were compared with three data balancing techniques: Over-Sampling, Under-Sampling, and Synthetic Minority Over-Sampling (SMOTE). In this case, SVM with SMOTE gave the best accuracy (90.24%) for retention prediction.

A promising technique was proposed in [7] for predicting the risk of dropout at early stages in online courses, where high dropout rate is a serious problem for this kind of courses at university level. This technique is based on a parallel combination of three ML techniques (K-Nearest Neighbor (KNN), RBN, and SVM), which make use of 28 attributes per student. Considering students' attributes, in [8] a set of ML algorithms (ANN, DT, and BN) took into account the personal characteristics of the students and their academic performance together with input attributes for building prediction models. The effectiveness of the prediction was evaluated using indicators such as the accuracy rate, recovery rate, overall accuracy rate and a particular measure. Moreover, if we take into account the cognitive characteristics of the students, the prediction accuracy improves using DT [9].

An SA framework for early identification of at-risk students was compared to other ML approaches [10], since more than 60% of dropouts occur in the first 2 years, especially in the areas of Science, Technology, Engineering, and Mathematics. Other ML algorithms (DT, NB, KNN, Gradient Boosted Tree (GBT), linear models, and Deep Learning (DL)) were proposed in [11] with similar purposes. Among them, DL and GBT showed the best accuracy. Other studies highlight the quality of SL techniques in predicting dropout: NB and SVM were proposed to predict of individual dropouts [12]; and Sequential Forward Selection (SFS), C4.5, RF, KNN and NB, among other classifiers, were proposed to identify students with difficulties in the third week with 97% accuracy [13]. Along these lines, the use of Random Forests (RF) showed excellent performance in predicting school dropout in terms of various performance metrics for binary classification [14]. Finally, ANN, SVM, LR, NB, and DT were analyzed in [15] for similar purposes by using the data recorded by e-learning tools. In this case, ANN and SVM achieved the highest accuracies.

Several ML algorithms were compared in [2] to predict the performance of new students, where NB showed the best behaviour in a web tool. SVM was the best of the four techniques analyzed in [16] for predicting academic performance. Also Bayesian Belief Network (BNN) was used to predict the students' performance (grade point average) early [17]. Also LR and SVM were applied for this purpose [18]. Nevertheless, the accuracy of the prediction systems can be improved through careful study and implementing different algorithmic features. Thus, preprocessing techniques have been applied together with classification algorithms (SVM, DT and NB) to improve prediction results [19].

A different focus on students' performance can be found in [20], where the main characteristics for observing performance are deduced from students' daily interaction events with certain modules of Moodle. For this purpose, RF and SVM developed the prediction models, and the best results were obtained by RF. With a similar focus, other SL algorithms analyzed datasets directly from websites to evaluate students' performance [21]. Also software platforms in e-learning made it possible to analyze and take advantage of the results of DM and ML algorithms in order to make decisions and justify educational approaches [22].

A data analysis approach to determine next trimester's courses was proposed in [23]. Here, different ML techniques predicted students' performance, which was used to build transition probabilities of a Markov Decision Process (MDP). The Jacobian Matrix-Based Learning Machine (JMLM) was used to analyze the students' learning performance [24], and AdaBoost assembly algorithm was proposed to predict student classification and showed best performance against techniques as DT, ANN, and SVM [25]. Adaboost was also the best meta-decision classifier for predicting student results [26].

SL algorithms are useful for a wide variety of predicting purposes. Predicting whether a student can successfully obtain a certificate was tackled by LR, SVM, NB, KNN, and BN [27]. Predicting graduation grade point averages was tackled by ANN, SVM, and Extreme Learning Machine (ELM) [28], where SVM gave the highest accurate prediction (97.98%). Student performance in the previous semester along with test grades from the current semester were used as input attributes for a series of algorithms (SVM, NB, RF and Gradient Boosting) that predict student grades [29].

Finally, other SL approaches were satisfactorily applied for predicting students' performance. Bayesian Additive Regressive Trees (BART) was used to predict the final grade of students in the sixth week [30]. A model based on SVM weekly predicted the probability of each student belonging to one of these three types: high, medium or low performance [31]. Latent Dirichlet Allocation (LDA) predicted student grades according to how the students described their learning situations after each lesson [32].

### 3.1.2. Unsupervised Learning

Unsupervised Learning (UL) is also known as class discovery. One of the main differences between UL and SL is that there is no training dataset in UL. As a consequence, there is no obvious role for cross validation [68]. Another important difference is that, although most clustering algorithms are expressed in terms of an optimal criterion, there is generally no guarantee that the optimal solution has been obtained.

A method based on a UL Sparse Auto-Encoder developed a classification model to predict students' performance by automatically learning multiple levels of representation [33]. Classification and clustering algorithms such as K-means and Hierarchical Clustering can be applied to evaluate students' performance [34]. Along these lines, Recursive Clustering was applied in [35] to group students from the programming course into performance-based groups.

### 3.2. Recommender Systems

Recommender systems collect information on the users' preferences for a set of elements (e.g., books, applications, websites, travel destinations, e-learning material, etc.). In the context of students' performance, the information can be acquired explicitly (by collecting users' scores) or implicitly (by monitoring users' behaviour, such as visits to teaching materials, documents downloaded, etc) [69]. RS consider different sources of information to provide predictions and recommendations. They try to balance factors such as precision, novelty, dispersion and stability in recommendations.

#### Collaborative Filtering

Collaborative Filtering methods play an important role in recommendation, although they are often used together with other filtering techniques such as content-based, knowledge-based or social [69]. Just as humans base their decisions according to past experiences and knowledge, CF acts in the same way to perform predictions.

Some studies predicted different issues with regard to students' performance through CF approaches. Thus, similarities among students were found in [36,37], where students' knowledge was represented as a set of grades from their previous courses. In this case, CF demonstrated a effectiveness similar to ML. Personalized predictions of student grades in required courses were generated from CF using improved similarities [38]. A typical CF method was compared to an article recommendation

method based on student's grade in order to recommend personalized articles in an online forum [39]. Students groups, defined by academic characteristics and course influenced matriculation patterns, can be used to design predictive grade models for CF based on neighborhood and MF, and approaches to classification based on popularity [40]. Most of these research studies for predicting students' performance tackle large data matrices. This is the reason why prediction accuracy was not so good when CF was applied for this purpose at small universities [41].

We can find some studies where CF inspires novel methods and tools that try to improve the results in particular environments. A novel student performance prediction model called PSFK combines user-based CF and the user modeling method called Bayesian Knowledge Tracing (BKT) [42]. A method called Hints-Model predicts students' performance [43]. It is combined with a factorization method called Regularized Single-Element-Based Non-Negative Matrix Factorization, achieving a significant improvement in predicting performance. A tool called Grade Prediction Advisor (pGPA) is based on CF and predicts grades in upcoming courses [44]. Two variants of the Low Range Matrix Factorization (LRMF) problem as a predictive task, weighted standard LRMF and non-negative weighted LRMF, were solved by applying the Expectation-Maximization procedure to solve it [45]. A CF technique (matrix decomposition) allows performance prediction of grades for combinations of student courses not observed so far, allowing personalized study planning and orientation for students [46]. A CF tool predicts the unknown performances by analyzing the database that contains students' performances for particular tasks [47]. The optimal parameters of this tool (learning rate and regularization factor) were selected with different metaheuristics in order to improve prediction accuracy. A prototype of RS for online courses improves the performance of new students. It uses CF and knowledge-based techniques to make use of the experience and results of old students in order to be able to suggest resources and activities to help new students [48].

Matrix factorization is a well-proven technique in this field. A study conducted at the University of KwaZulu-Natal investigated the efficacy of MF in solving the prediction problem. In this study, an MF technique called Singular Value Decomposition (SVD) was successfully applied [49]. This method was compared with simple baselines (Uniform Random, Global Mean and Mean of Means) when predicting retention [50]. MF and biased MF were compared with other CF methods when predicting whether or not students would answer multiple choice questions: two reference methods (random and global average), two memory-based algorithms (User-kNN and Item-kNN), and two Slope One algorithms (Slope One and Bipolar Slope One) [51]. Probabilistic MF and Bayesian Probabilistic MF using Markov Chain Monte Carlo were used for predicting grades for courses not yet matriculated in by the students, which can help them to make decisions [52].

### 3.3. Artificial Neural Networks

An ANN consists of a set of highly interconnected entities, called Processing Elements. The structure and function of the network is inspired by the biological central nervous system, particularly the brain. Each Processing Element is designed to mimic its biological counterpart, the neuron [53], which accepts a weighted set of inputs and responds with the corresponding output.

ANNs have been applied to different prediction approaches, basically by considering the evaluation results of students, as the following cases show. A feedforward ANN was trained to predict the scores of evaluation tests considering partial scores during the course [54]. An ANN that uses the Cumulative Grade Point Average predicted the academic performance in the eighth semester [55]. Two models of ANN (Multilayer Perceptron and Generalized Regression Neural Network) were compared in order to identify the best model to predict academic performance of students [56]. Lastly, the potential of ANNs to predict learning results was compared to the multivariate LR model in the area of medical education [57].

Not only mere evaluation results, but also additional information from students can improve prediction performed by ANNs. Thus, basic students' information, along with cognitive and non-cognitive measures, were used to design predictive models of students' performance by using

three ANN models [58]. The non-linear relationship between cognitive and psychological variables that influence academic performance was analyzed by an ANN, which efficiently grouped students into different categories according to their level of expected performance [53]. Finally, an ELM (which is a particular type of ANN) predicted students' performance by considering the value of the subjects that focus on the final national exam [59].

### 3.4. Impact of the Techniques

The techniques described before had different efficiencies with regard to the students' behaviour. As shown in the bar graph of Figure 1, the different algorithms were not only applied to a greater or lesser extent (blue bars), but also had different performance (green bars) when compared to others. Thus, we check that ANN and SVM were more the most applied, followed by CF, DT, and NB.

On the other hand, SVM was the best method in performance terms. This conclusion should be taken with caution, since it is necessary to consider which algorithms were involved in the comparison, as well as the particular case where they were applied. However, these results may show some guidance in making decisions about which techniques to use for particular scenarios.

## 4. Objectives

We have gathered the different objectives into four wide groups: student dropout, students' performance, recommend recommended activities and resources, and students' knowledge. Figure 1 shows the weight of each of these objectives in the literature, which can indicate their importance or interest for research. In this sense, students' performance collect the majority of the prediction efforts (70%), followed by student dropout (21%). Students' knowledge and recommend activities and resources were low-demand objectives (6% and 3% respectively).

### 4.1. Student Dropout

Several studies focused on the dropout rate in nursing courses have tried to find the causes rather than predicting the likelihood of students dropping out. A useful method for trying to make this type of prediction is the induction of rules, using IBM SPSS Answer Tree (AT) software [4] for this purpose. The authors [5] found that the following factors are highly informative in predicting school dropout: family history, socioeconomic status of families, high school grade and exam results.

It was noticed that unbalanced class data was a common problem for prediction [6]. In addition, classification techniques with unbalanced datasets can provide deceptively high prediction accuracy. To solve this problem, the authors compared different data balancing techniques (including SMOTE) to improve accuracy. All these techniques improved the accuracy of predictions, although Support Vector Machine (SVM) combined with SMOTE data balancing technique achieved the best performance.

Nowadays, higher education institutions are attempting to use data collected in university systems to identify students at risk of dropping out [64]. This study uses the data to validate the Moodle Engagement Analytics Plugin learning analysis tool. High dropout rates are a very important problem for e-learning. The authors propose a technique that considers a combination of multiple classifiers to analyze a set of attributes of students' activities over time [7]. Other authors [8] selected students' personal characteristics and academic performance as input attributes. They developed prediction models using ANN, Decision Trees (DT) and Bayesian Networks (BN). Along these lines, another study [65] identified the most important factors for predicting school dropout risk: those that showed student commitment and consistency in the use of online resources. For this purpose, Exploratory Data Analysis was applied.

In particular, higher education institutions in the United States faced a problem of university student attrition, especially in the areas of Science, Technology, Engineering and Mathematics. More than 60% of the dropouts occurred in the first two years. One study develops and evaluates a Survival Analysis (SA) framework for early identification of students at risk of dropping out of school and early intervention to improve student retention [10].

#### 4.2. Student Performance

One of the essential and most challenging issues for educational institutions is the prediction of students' performance. Particularly, this issue could be very useful in e-learning environments at university level. We can find several approaches in the literature for this purpose.

The demographic characteristics of the students and their grades in some tasks can build a good training set for a machine-supervised learning algorithm [2]. Adding other characteristics such as the cumulative grade point of the students, the grades obtained in other courses and the ratings of several exams, can build accurate models. Pursuing this goal, four mathematical models were compared to predict students' performance in a basic course, a high-impact course and a high-enrollment course in engineering dynamics [16]. In this sense, it is advisable to consider several more characteristics, since a relationship among different factors may appear after a detailed analysis of the prediction results. Thus, an analysis of different characteristics of the data obtained from the results of primary school exams in Tamil Nadu (India) showed the relationship between ethnicity, geographic environment, and students' performance [3].

If we focus on the students' history, in [36,37] the performance is predicted considering particular first semester courses. Our goal was to represent the knowledge as a set of grades from their passed courses and to be able to find similarity among students to predict their performance. In small universities or in courses with few students [41], the research was carried out with large sparse matrices, which represented students, assignments, and grades. The result obtained in this research showed that prediction accuracy was not as good as expected; therefore more information from students or homework was needed. Accuracy is important since it can be very useful in planning educational interventions aimed at improving the results of the teaching-learning process, saving government resources and educators' time and effort [51]. Moreover, the additional use of pre-processing techniques along with classification algorithms has improved performance prediction accuracy [19].

It is possible to predict final students' performance beforehand thanks to behavioural data supplemented with other more relevant data (related to learning results). The system proposed in [31] obtained a weekly ranking of each student's probability of belonging to one of these three classification levels: high, medium or low performance. This performance could have something to do with non-cognitive characteristics which can have a significant impact on the students [9]. This research concluded that the prediction mechanism improves by exploiting the cognitive and non-cognitive characteristics of students, thereby increasing accuracy. In any case, the data obtained from previous records seem to be important, even better than applying course-dependent formulas to predict performance [26].

ML Clustering techniques have been satisfactorily applied in this field. For example, recursive clustering groups the students into specific courses according to their performance. Each of these groups receives a set of programs and notes automatically, depending on which group they belong to. The goal of this technique is to move the majority of the students from lower to higher groups [35]. Nevertheless, each student has particular features to be taken into account. A personalized prediction of the student's performance will aid in finding the right specialization for each student. For example, a method of personalized prediction is presented in [38], where specific characteristics such as basic courses, prerequisites and course levels were analyzed for computer specialization courses.

#### 4.3. Recommender Activities and Resources

Recommender systems have been used to improve the experience of students and teachers. Most of the studies based on RS consider demographics, interests or preferences of the students to improve their systems. For example, an RS was developed considering the experiences previously stored and classified by former students, which were compared with the current students' competencies [48]. Another example is an RS based on student's performance, which recommends personalized articles to students in an online forum, using a "Like" button similar to the one on Facebook for this purpose [39].

#### 4.4. Students' Knowledge

The trend in the use of learning systems aims to analyse the information generated by students [60]. This approach seeks to improve the effectiveness of the education process through the recognition of patterns in students' performance. Along these lines, an automatic approach that detects students' learning styles is proposed in [61] to offer adaptable courses in Moodle. It is based on students' response to the learning style and the analysis of their behavior within Moodle.

In this context, it is very important to discover which students' characteristics are associated with test results, and which school characteristics are associated with the added value of the school [62]. For example, machine learning applications were proposed to acquire knowledge about students' learning in computer science, develop optimal warning models, and discover behavioural indicators from learning analytical reports [63].

## 5. Discussion

In this article, we have reviewed many papers aimed at predicting student behavior in the academic environment. We can draw some conclusions from the analysis of these papers.

We have noted that there is a strong tendency to predict student performance at the university level, as around 70% of the articles included in this review are intended for this purpose. This may encourage us to consider complementary research efforts to fill gaps in other areas. Thus, we consider that it would be interesting to promote working lines to apply these predictions at school level, which would contribute to identify the low performance of students at early ages. The analysis of student dropout during the early stages of their levels is very interesting, as there are still opportunities to research about helpful predictive tools to enable prevention mechanisms. In this sense, a good approach to research would be to apply the same predictive techniques used for academic performance (and other novel ones) to this case, in addition to considering non-university levels.

Based on the data collected in this review, the most widely used technique for predicting students' behavior was supervised learning, as it provides accurate and reliable results. In particular, the SVM algorithm was the most used by the authors and provided the most accurate predictions. In addition to SVM, DT, NB and RF have also been well-studied algorithmic proposals that generated good results.

Recommender systems, in particular collaborative filtering algorithms, have been the next successful technique in this field. However, it should be clarified that success has been more in recommending resources and activities than in predicting student behavior.

As for the neural networks, they are a less used technique, but they obtain a great precision in predicting the students' performance. We believe that a good line of research with these techniques would be to apply them to other related types of predictions in the educational field, different from the strict students' performance.

We emphasize that unsupervised learning is an unattractive technique for researchers, due to the low accuracy of predicting students' behavior in the cases studied. However, this fact can be an incentive for research, as it provides the opportunity to further improve these techniques in order to obtain more reliable and accurate results.

This review can be useful to obtain a wide insight of the possibilities to apply ML for predicting students' performance and related problems. In this regard, Table 1 and Figure 1 may be useful to researchers in planning how to approach the initial stages of their studies. Nevertheless, many researchers will probably tackle this problem in the coming years considering other and new ML tools, since this problem has attracted a high degree of interest nowadays.

**Author Contributions:** Search, classification, and analysis of bibliographic resources, J.L.R.-G.; work methodology, writing and editing of original manuscript, J.A.G.-P.; supervision and access to resources, A.D.-D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partially funded by the Government of Extremadura (Spain) under the project IB16002, and by the ERDF (European Regional Development Fund, EU) and the AEI (State Research Agency, Spain) under the contract TIN2016-76259-P.

**Acknowledgments:** We express our gratitude to the staff of the Service of Library of the University of Extremadura, Spain, for their support and ease in accessing to the different bibliographic resources and databases.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

AB	AdaBoost
ANN	Artificial Neural Networks
AT	Answer Tree
BART	Bayesian Additive Regressive Trees
BBN	Bayesian Belief Network
BKT	Bayesian Knowledge Tracing
BMF	Biased-Matrix Factorization
BN	Bayesian Networks
BSLO	Bipolar Slope One
CBN	Combination of Multiple Classifiers
CF	Collaborative Filtering
DL	Deep Learning
DM	Data Mining
DT	Decision Tree
ELM	Extreme Learning Machine
EM	Expectation-Maximization
GBT	Gradient Boosted Tree
JMLM	Jacobian Matrix-Based Learning Machine
KNN	K-Nearest Neighbor
LDA	Latent Dirichlet Allocation
LR	Logistic Regression
LRMF	Low Range Matrix Factorization
LM	Linear Models
MDP	Markov Decision Process
MF	Matrix Factorization
ML	Machine Learning
MLP	Multilayer Perception
MLR	Multiple Linear Regression
NB	Naïves Bayes
pGPA	Grade Prediction Advisor
RBN	Radial Basis Networks
RF	Random Forests
RS	Recommender Systems
SA	Survival Analysis
SFS	Sequential Forward Selection
SL	Supervised Learning
SLO	Slope One
SMOTE	Synthetic Minority Over-Sampling
SVD	Singular Value Decomposition
SVM	Support Vector Machine
UL	Unsupervised Learning

## References

1. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*, 3rd ed.; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 2011.
2. Kotsiantis, S.; Pierrakeas, C.; Pintelas, P. Predicting Students' Performance in Distance Learning using Machine Learning Techniques. *Appl. Artif. Intell.* **2004**, *18*, 411–426. [[CrossRef](#)]
3. Navamani, J.; Kannammal, A. Predicting performance of schools by applying data mining techniques on public examination results. *Res. J. Appl. Sci. Eng. Technol.* **2015**, *9*, 262–271. [[CrossRef](#)]
4. Moseley, L.; Mead, D. Predicting who will drop out of nursing courses: A machine learning exercise. *Nurse Educ. Today* **2008**, *28*, 469–475. [[CrossRef](#)] [[PubMed](#)]

5. Nandeshwar, A.; Menzies, T.; Nelson, A. Learning patterns of university student retention. *Expert Syst. Appl.* **2011**, *38*, 14984–14996. [[CrossRef](#)]
6. Thammasiri, D.; Delen, D.; Meesad, P.; Kasap, N. A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *Expert Syst. Appl.* **2014**, *41*, 321–330. [[CrossRef](#)]
7. Dewan, M.; Lin, F.; Wen, D.; Kinshuk. Predicting dropout-prone students in e-learning education system. In Proceedings of the 2015 IEEE 12th Intl Conference on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conference on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conference on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), Beijing, China, 14 August 2016; pp. 1735–1740. [[CrossRef](#)]
8. Tan, M.; Shao, P. Prediction of student dropout in E-learning program through the use of machine learning method. *Int. J. Emerg. Technol. Learn.* **2015**, *10*, 11–17. [[CrossRef](#)]
9. Sultana, S.; Khan, S.; Abbas, M. Predicting performance of electrical engineering students using cognitive and non-cognitive features for identification of potential dropouts. *Int. J. Electr. Eng. Educ.* **2017**, *54*, 105–118. [[CrossRef](#)]
10. Chen, Y.; Johri, A.; Rangwala, H. Running out of STEM: A comparative study across STEM majors of college students At-Risk of dropping out early. In Proceedings of the 8th International Conference on Learning Analytics and Knowledge, Sydney, Australia, 7–9 March 2018; pp. 270–279.
11. Nagy, M.; Molontay, R. Predicting Dropout in Higher Education Based on Secondary School Performance. In Proceedings of the 2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES), Las Palmas de Gran Canaria, Spain, 21 June 2018; pp. 000389–000394. [[CrossRef](#)]
12. Serra, A.; Perchinunno, P.; Bilancia, M. Predicting student dropouts in higher education using supervised classification algorithms. *Lect. Notes Comput. Sci.* **2018**, *10962 LNCS*, 18–33. [[CrossRef](#)]
13. Gray, C.; Perkins, D. Utilizing early engagement and machine learning to predict student outcomes. *Comput. Educ.* **2019**, *131*, 22–32. [[CrossRef](#)]
14. Chung, J.; Lee, S. Dropout early warning systems for high school students using machine learning. *Child. Youth Serv. Rev.* **2019**, *96*, 346–353. [[CrossRef](#)]
15. Hussain, M.; Zhu, W.; Zhang, W.; Abidi, S.; Ali, S. Using machine learning to predict student difficulties from learning session data. *Artif. Intell. Rev.* **2018**, *52*, 1–27. [[CrossRef](#)]
16. Huang, S.; Fang, N. Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Comput. Educ.* **2013**, *61*, 133–145. [[CrossRef](#)]
17. Slim, A.; Heileman, G.L.; Kozlick, J.; Abdallah, C.T. Predicting student success based on prior performance. In Proceedings of the 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), Singapore, 16 April 2015; pp. 410–415. [[CrossRef](#)]
18. Zhao, C.; Yang, J.; Liang, J.; Li, C. Discover learning behavior patterns to predict certification. In Proceedings of the 2016 11th International Conference on Computer Science & Education (ICCSE), Nagoya, Japan, 23 August 2016; pp. 69–73. [[CrossRef](#)]
19. Chaudhury, P.; Mishra, S.; Tripathy, H.; Kishore, B. Enhancing the capabilities of student result prediction system. In Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies, Udaipur, India, 4–5 March 2016. [[CrossRef](#)]
20. Nespereira, C.; Elhariri, E.; El-Bendary, N.; Vilas, A.; Redondo, R. Machine learning based classification approach for predicting students performance in blended learning. *Adv. Intell. Syst. Comput.* **2016**, *407*, 47–56. [[CrossRef](#)]
21. Sagar, M.; Gupta, A.; Kaushal, R. Performance prediction and behavioral analysis of student programming ability. In Proceedings of the 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, 21 September 2016; pp. 1039–1045. [[CrossRef](#)]
22. Verhun, V.; Batyuk, A.; Voityshyn, V. Learning Analysis as a Tool for Predicting Student Performance. In Proceedings of the 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine, 11 September 2018; Volume 2, pp. 76–79. [[CrossRef](#)]
23. Backenköhler, M.; Wolf, V. Student performance prediction and optimal course selection: An MDP approach. *Lect. Notes Comput. Sci.* **2018**, *10729 LNCS*, 40–47. [[CrossRef](#)]
24. Hsieh, Y.Z.; Su, M.C.; Jeng, Y.L. The jacobian matrix-based learning machine in student. *Lect. Notes Comput. Sci.* **2017**, *10676 LNCS*, 469–474. [[CrossRef](#)]

25. Han, M.; Tong, M.; Chen, M.; Liu, J.; Liu, C. Application of Ensemble Algorithm in Students' Performance Prediction. In Proceedings of the 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), Hamamatsu, Japan, 16 November 2017; pp. 735–740. [[CrossRef](#)]
26. Shanthini, A.; Vinodhini, G.; Chandrasekaran, R. Predicting students' academic performance in the University using meta decision tree classifiers. *J. Comput. Sci.* **2018**, *14*, 654–662. [[CrossRef](#)]
27. Ma, C.; Yao, B.; Ge, F.; Pan, Y.; Guo, Y. Improving prediction of student performance based on multiple feature selection approaches. In Proceedings of the ICEBT 2017, Toronto, ON, Canada, 10–12 September 2017; pp. 36–41. [[CrossRef](#)]
28. Tekin, A. Early prediction of students' grade point averages at graduation: A data mining approach [Öğrencinin mezuniyet notunun erken tahmini: Bir veri madenciliği yaklaşıdotlessmidotless]. *Eğit. Arastirmalari Eurasian J. Educ. Res.* **2014**, 207–226. [[CrossRef](#)]
29. Pushpa, S.; Manjunath, T.; Mrunal, T.; Singh, A.; Suhas, C. Class result prediction using machine learning. In Proceedings of the 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), Bengaluru, India, 19 August 2018; pp. 1208–1212. [[CrossRef](#)]
30. Howard, E.; Meehan, M.; Parnell, A. Contrasting prediction methods for early warning systems at undergraduate level. *Internet High. Educ.* **2018**, *37*, 66–75. [[CrossRef](#)]
31. Villagrà-Arnedo, C.; Gallego-Duran, F.; Compan-Rosique, P.; Llorens-Largo, F.; Molina-Carmona, R. Predicting academic performance from Behavioural and learning data. *Int. J. Des. Nat. Ecodyn.* **2016**, *11*, 239–249. [[CrossRef](#)]
32. Sorour, S.; Goda, K.; Mine, T. Estimation of Student Performance by Considering Consecutive Lessons. In Proceedings of the 4th International Congress on Advanced Applied Informatics, Okayama, Japan, 12 June 2016, pp. 121–126. [[CrossRef](#)]
33. Guo, B.; Zhang, R.; Xu, G.; Shi, C.; Yang, L. Predicting Students Performance in Educational Data Mining. In Proceedings of the 2015 International Symposium on Educational Technology (ISET), Wuhan, China, 24 March 2016; pp. 125–128. [[CrossRef](#)]
34. Rana, S.; Garg, R. Prediction of students performance of an institute using ClassificationViaClustering and ClassificationViaRegression. *Adv. Intell. Syst. Comput.* **2017**, *508*, 333–343. [[CrossRef](#)]
35. Anand, V.K.; Abdul Rahiman, S.K.; Ben George, E.; Huda, A.S. Recursive clustering technique for students' performance evaluation in programming courses. In Proceedings of the 2018 Majan International Conference (MIC), Muscat, Oman, 19 March 2018; pp. 1–5. [[CrossRef](#)]
36. Bydžovská, H. Student performance prediction using collaborative filtering methods. *Lect. Notes Comput. Sci.* **2015**, *9112*, 550–553. [[CrossRef](#)]
37. Bydžovská, H. Are collaborative filtering methods suitable for student performance prediction? *Lect. Notes Comput. Sci.* **2015**, *9273*, 425–430. [[CrossRef](#)]
38. Park, Y. Predicting personalized student performance in computing-related majors via collaborative filtering. In Proceedings of the 19th Annual SIG Conference on Information Technology Education, Fort Lauderdale, FL, USA, 3 October 2018; p. 151. [[CrossRef](#)]
39. Liou, C.H. Personalized article recommendation based on student's rating mechanism in an online discussion forum. In Proceedings of the 2016 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, USA, 5–8 January 2016; pp. 60–65. [[CrossRef](#)]
40. Elbadrawy, A.; Karypis, G. Domain-aware grade prediction and top-n course recommendation. In Proceedings of the 10th ACM Conference on Recommender Systems, Boston, MA, USA, 15 September 2016, pp. 183–190. [[CrossRef](#)]
41. Pero, v.; Horváth, T. Comparison of collaborative-filtering techniques for small-scale student performance prediction task. *Lect. Notes Electr. Eng.* **2015**, *313*, 111–116. [[CrossRef](#)]
42. Song, Y.; Jin, Y.; Zheng, X.; Han, H.; Zhong, Y.; Zhao, X. PSFK: A student performance prediction scheme for first-encounter knowledge in ITS. *Lect. Notes Comput. Sci.* **2015**, *9403*, 639–650. [[CrossRef](#)]
43. Xu, K.; Liu, R.; Sun, Y.; Zou, K.; Huang, Y.; Zhang, X. Improve the prediction of student performance with hint's assistance based on an efficient non-negative factorization. *IEICE Trans. Inf. Syst.* **2017**, *E100D*, 768–775. [[CrossRef](#)]
44. Sheehan, M.; Park, Y. pGPA: A personalized grade prediction tool to aid student success. In Proceedings of the sixth ACM conference on Recommender systems, Dublin, Ireland, 3 September 2012, pp. 309–310. [[CrossRef](#)]

45. Lorenzen, S.; Pham, N.; Alstrup, S. On predicting student performance using low-rank matrix factorization techniques. In Proceedings of the 9th European Conference on E-Learning—ECEL 2010 (ECEL 2010), Porto, Portugal, 5 November 2010; pp. 326–334.
46. Houbraken, M.; Sun, C.; Smirnov, E.; Driessens, K. Discovering hidden course requirements and student competences from grade data. In Proceedings of the UMAP '17: Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization, Bratislava, Slovakia, 9–10 July 2017; pp. 147–152. [[CrossRef](#)]
47. Gómez-Pulido, J.; Cortés-Toro, E.; Durán-Domínguez, A.; Crawford, B.; Soto, R. Novel and Classic Metaheuristics for Tuning a Recommender System for Predicting Student Performance in Online Campus. *Lect. Notes Comput. Sci.* **2018**, *11314 LNCS*, 125–133. [[CrossRef](#)]
48. Chavarriaga, O.; Florian-Gaviria, B.; Solarte, O. A recommender system for students based on social knowledge and assessment data of competences. *Lect. Notes Comput. Sci.* **2014**, *8719 LNCS*, 56–69. [[CrossRef](#)]
49. Jembere, E.; Rawatlal, R.; Pillay, A. Matrix Factorisation for Predicting Student Performance. In Proceedings of the 2017 7th World Engineering Education Forum (WEEF), Kuala Lumpur, Malaysia, 16 November 2018; pp. 513–518. [[CrossRef](#)]
50. Sweeney, M.; Lester, J.; Rangwala, H. Next-term student grade prediction. In Proceedings of the 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, USA, 1 November 2015; pp. 970–975. [[CrossRef](#)]
51. Adán-Coello, J.; Tobar, C. Using collaborative filtering algorithms for predicting student performance. *Lect. Notes Comput. Sci.* **2016**, *9831 LNCS*, 206–218. [[CrossRef](#)]
52. Rechkoski, L.; Ajanovski, V.; Mihova, M. Evaluation of grade prediction using model-based collaborative filtering methods. In Proceedings of the 2018 IEEE Global Engineering Education Conference (EDUCON), Santa Cruz de Tenerife, Spain, 17 April 2018; pp. 1096–1103. [[CrossRef](#)]
53. Adewale Amoo, M.; Olumuyiwa, A.; Lateef, U. Predictive modelling and analysis of academic performance of secondary school students: Artificial Neural Network approach. *Int. J. Sci. Technol. Educ. Res.* **2018**, *9*, 1–8. [[CrossRef](#)]
54. Gedeon, T.; Turner, H. Explaining student grades predicted by a neural network. In Proceedings of the 1993 International Conference on Neural Networks, Nagoya, Japan, 25 October 1993; Volume 1, pp. 609–612.
55. Arsad, P.M.; Buniyamin, N.; Manan, J.A. A neural network students' performance prediction model (NNSPPM). In Proceedings of the 2013 IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), Kuala Lumpur, Malaysia, 27 November 2013; pp. 1–5. [[CrossRef](#)]
56. Iyanda, A.; D. Ninan, O.; Ajayi, A.; G. Anyabolu, O. Predicting Student Academic Performance in Computer Science Courses: A Comparison of Neural Network Models. *Int. J. Mod. Educ. Comput. Sci.* **2018**, *10*, 1–9. [[CrossRef](#)]
57. Dharmasaroja, P.; Kingkaew, N. Application of artificial neural networks for prediction of learning performances. In Proceedings of the 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, China, 13 August 2016; pp. 745–751. [[CrossRef](#)]
58. Musso, M.; Kyndt, E.; Cascallar, E.; Dochy, F. Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. *Frontline Learn. Res.* **2013**, *1*, 42–71. [[CrossRef](#)]
59. Mala Sari Rochman, E.; Rachmad, A.; Damayanti, F. Predicting the Final result of Student National Test with Extreme Learning Machine. *Pancar. Pendidik.* **2018**, *7*. [[CrossRef](#)]
60. Villegas-Ch, W.; Lujan-Mora, S.; Buenano-Fernandez, D.; Roman-Canizares, M. Analysis of web-based learning systems by data mining. In Proceedings of the 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM), Guayas, Ecuador, 16 October 2018; pp. 1–5. [[CrossRef](#)]
61. Karagiannis, I.; Satratzemi, M. An adaptive mechanism for Moodle based on automatic detection of learning styles. *Educ. Inf. Technol.* **2018**, *23*, 1331–1357. [[CrossRef](#)]
62. 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](#)]

63. Johnson, W. Data mining and machine learning in education with focus in undergraduate cs student success. In Proceedings of the 2018 ACM Conference on International Computing Education Research, Espoo, Finland, 13–15 August 2018; pp. 270–271. [[CrossRef](#)]
64. Liu, D.; Richards, D.; Froissard, C.; Atif, A. Validating the effectiveness of the moodle engagement analytics plugin to predict student academic performance. In Proceedings of the 21st Americas Conference on Information Systems (AMCIS 2015), Fajardo, Puerto Rico, 13 August 2015.
65. Saqr, M.; Fors, U.; Tedre, M. How learning analytics can early predict under-achieving students in a blended medical education course. *Med Teach.* **2017**, *39*, 757–767. [[CrossRef](#)] [[PubMed](#)]
66. Kotsiantis, S. Supervised machine learning: A review of classification techniques. *Informatica* **2007**, *31*, 249–268.
67. Hosmer, D.; Lemeshow, S.; Sturdivant, R. *Applied Logistic Regression*; Wiley Series in Probability and Statistics; Wiley: Hoboken, NJ, USA, 2013.
68. Gentleman, R.; Carey, V.J. Unsupervised Machine Learning. In *Bioconductor Case Studies*; Springer: New York, NY, USA, 2008; pp. 137–157. [[CrossRef](#)]
69. Bobadilla, J.; Ortega, F.; Hernando, A.; Gutiérrez, A. Recommender systems survey. *Knowl. Based Syst.* **2013**, *46*, 109–132. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).