

STUDENTS' ACADEMIC PERFORMANCE AND DROPOUT PREDICTIONS: A REVIEW

Ahmed O. Ameen¹, Moshood A. Alarape², Kayode S. Adewole¹

¹Department of Computer Science, University of Ilorin, Ilorin, Nigeria.

²Department of Computer Science, Federal Polytechnic, Bida, Nigeria.
aminamed@unilorin.edu.ng, acica1427ah@yahoo.com, adewole.ks@unilorin.edu.ng

ABSTRACT

Students' Academic Performance (SAP) is an important metric in determining the status of students in any academic institution. It allows the instructors and other education managers to get an accurate evaluation of the students in different courses in a particular semester and also serve as an indicator to the students to review their strategies for better performance in the subsequent semesters. Predicting SAP is therefore important to help learners in obtaining the best from their studies. A number of researches in Educational Psychology (EP), Learning Analytics (LA) and Educational Data Mining (EDM) has been carried out to study and predict SAP, most especially in determining failures or dropouts with the goal of preventing the occurrence of the negative final outcome. This paper presents a comprehensive review of related studies that deal with SAP and dropout predictions. To group the studies, this review proposes taxonomy of the methods and features used in the literature for SAP and dropout prediction. The paper identifies some key issues and challenges for SAP and dropout predictions that require substantial research efforts. Limitations of the existing approaches for SAP and dropout prediction are identified. Finally, the paper exposes the current research directions in the area.

Keywords: Dropout Prediction; Educational Data Mining; Learning Analytics; Machine Learning

Received for review: 21-08-2019; Published: 22-11-2019

1. Introduction

The sacred goal of every academic institution is to entrench learning and learning cannot be said to have taken place except that the learners are correctly evaluated. This evaluation which used to be in the form of class tests, assignments, practical/laboratory works, and semester examinations is considered to be holistically done to evolve the academic status of each of the students in a class in different courses. The performance of the student is, therefore, one of the crucial aspect of every educational institution (Solomon, Patil, & Agrawal, 2018).

Students go through several semesters before the completion of their programmes and they are evaluated on semester or term bases; the final academic status on graduation or even in a semester other than the first semester is always a factor of the previous semesters, which gives an opportunity for the prediction of a subsequent semester performance based on the previous ones. This prediction is nowadays being aided by several Data Mining (DM) tools and techniques for accurate analysis of the results of predictions. The domain of the prediction of SAP is the Educational Data Mining (EDM). Often, prediction models are generated by EDM tools to help in facilitating SAP prediction used in monitoring students' academic progress so as to determine important strategies to employ by the students and other education stakeholders.

Dropout prediction, on the other hand, deals with the prediction of students who are likely not to be retained in the educational institution because of various factors. These factors can be socio-economic, academic and psychological factors. Students who are affected by these factors are not likely to remain in the system and that is the more reason why the education stakeholders need to put more focus on this category of students to help them remain in the system. Early prediction of these type of students most especially using an Early Warning System (EWS) will go a long way to improve the image of the institution and as well improve the revenue generation status of such institution (Marquez-Vera, Cano, Romero, Noaman, Fardoun & Ventura, 2016). When dropout instances are a fallout of SAP, it is said to be a dropout due to academic factors. Both SAP and Dropout evolved from EDM domain. The mention of both SAP and dropout together in this review does not mean the two are taken to be the same but the two are being considered together because of their close relationship in the life of a student.

There have been several studies in EDM that focused on the predictions of SAP and dropouts which featured the use of DM techniques like Naïve Bayes (NB), Support Vector Machine (SVM), Decision Trees (DT) and its variants, Artificial Neural Network (ANN), k-Nearest Neighbour (kNN), Logistic Regression (LR), Extreme Learning Machine (ELM) and ensemble methods like the Random Forest (RF); Adaptive Boosting (AB) and Bagging for predicting the students' academic performance and dropouts in particular. Other DM techniques like the use of Nature-Inspired algorithms, Deep Learning and the hybrid methods had also been applied to predict SAP and dropouts even though the studies in this respect are scanty. These studies revealed the eligibility of data mining techniques to efficiently predict the performance of students at different levels of their studies using various academic performance dependent factors. However, there still exists the need for more researches in conducting and producing an improved framework for SAP and dropout predictions. This paper is set out to review some related and relevant studies in this area and reveal their strengths and limitations with a view to identifying research gaps.

The rest of the paper is organized as follows. Section 2 discusses some of the factors that affect SAP; this is followed by Section 3, which focuses on datasets that have been used for SAP and dropout prediction. Section 4 presents the taxonomy of the methods used in SAP and dropout predictions and Section 5 focuses on taxonomy of the features. Discussion on the various identified limitations in the existing studies is given in section 6. Finally, Section 7 concludes the paper and highlights the future directions for further studies on SAP and dropout's prediction.

2. Factors Affecting Students Academic Performance

Researchers in educational psychology had discovered so many factors that are responsible for Students' Academic Performance (SAP); these researches in educational psychology had also gone a long way to help the EDM researchers in generating variables for study when mining education data. Some of these factors are given in this section. Hijazi and Naqvi (2006) agreed with Tinto (2006) that student performance is the product of socio-economic, psychological and environmental factors.

These factors were explained by others as classroom management, students motivation, balance between moral and social development of students, students' performance monitoring strategies with emphasis on their different nature, parents involvement and comprehensive assessment (Seifert and Sutton, 2009), family stress, communication, learning facilities and proper guidance by both parents and teachers (Mushtaq and Khan, 2012). Age, guardian, father's socio-economic status and daily study hours (Ali et al., 2013). Also mentioned is the student's attendance in classes (Aden, Yahye, and Dahir, 2013), prior academic success, incentives and expectations (Nayebzadeh, Dehnavi, Nejad, Mir, and Sadrabadi, 2013). Truancy and physical environment were recognized by Fareo (2013). Availability of teaching resources, teacher quality, and performance, socio-economic status, entry grade, motivation and attitudes were gathered in Enu, Agyman, and Nkum (2015) and a host of other factors. Most of the factors mentioned above are either found to belong to the perceived academic control or as indicating the student's academic emotion (Respondek et al., 2017).

3. Students' Academic Performance Datasets

SAP and dropout datasets can be categorized as purely institution-based or traditional survey-based depending on the mode of dataset acquisition. The datasets employed in the SAP and dropout studies reviewed in this paper are from either of three data sources, namely: the students' record system, the learning management system and the traditional survey. Both students' record system and learning management system fall under the institution-based datasets collection avenues.

The students' record system as a data source is private to the individual institution and it basically contains the direct data values relating to students' academic performance within the institution databases. The Learning Management System (LMS) as a data source allows the researchers to extract software generated data values relating to students' performance from the system. Data values of attributes that are connected to students' academic performance are mostly obtained from the students using fact-finding tools like the questionnaire and the interview. Although some of these facts to be obtained from students using traditional survey can also be acquired from the students profile entered during sessional registration of courses by the students, the specialized survey can be carried out to obtain such data in case they are not available in students' profiles.

The availability of the SAP dataset is mostly private and is rarely publicly available. There are studies among the reviewed ones that made use of the three data sources for the SAP dataset like the studies of Adejo and Connolly (2018). Some others made use of the LMS data source for their dataset like the studies of Romero *et al.*, (2008). The student's record system that usually contained the exact data values indicating the academic performance of the students in every institution is a widely used data source for all SAP and dropout studies. Datasets used for SAP and dropout studies can be of large size if the researcher considered obtaining several years data and they can contain several features depending on the objective of the research.

4. Taxonomy of Methods Used for SAP and Dropout Predictions

A classification of the various methods used for SAP and dropout predictions in the literature is given in this section with emphasis on the data mining classifiers examples belonging to each classification discussed. SAP and dropout predictions have been approached in three fields of human endeavour, namely: the Educational Psychology, the Learning Analytics, and Data Mining as can be seen in Figure 1. In the Educational Psychology field, researchers tend to postulate theories relating to the effects of certain psychological factors on the final outcome of a student performance, a good example of the works in this area is the study of Respondek *et al.*, (2017) which examines the effect of perceived academic control and academic emotion on dropout intention and students' academic achievements.

Learning Analytics, defined by LAK (2011) as the measurement, collection, analysis and reporting of data about learners and their contexts, for the purpose of understanding and optimizing learning and the environments in which learning occurs, take a deeper steps in analyzing students' performance in learning using a host of techniques when compare to the adopted method by the educational psychologist. These techniques include Statistics, Business Intelligence, Web Analytics, Operational Research, Artificial Intelligence, Social Network Analysis, and Information Visualization. There are several studies in this area, for instance, the work of Daud *et al.*, (2017) which used advanced learning analytics to predict student performance. Data Mining (DM) has offered a lot of its applications in SAP and dropout predictions in a new emerging field of Educational Data Mining (EDM). The DM approaches employed so far can be classified into three, namely: supervised learning approach, semi-supervised approach and unsupervised learning approach.

A supervised learning approach is one in which the class labels are known from the datasets employed. The techniques or methods used in predicting SAP and dropout that are supervised learning are categorized as either generative or discriminative models. A generative model is a statistical model of the joint probability distribution $p(x, y)$ on input x and output y while a discriminative model is a statistical model that defines the conditional probability $P(y/x)$ of output y given input x (Andrew and Michael, 2001). Discriminative models, unlike the generative model, solve classification problems directly rather than solving a more general problem as an intermediate step (Nivre, 2003).

Naïve Bayes (NB) and Decision Trees (DT) are both examples of generative models commonly used while the Logistic Regression (LR), Support Vector Machine (SVM), Neural Network (NN), k-

Nearest Neighbor (kNN) and Random Forest (RF) are examples of commonly used discriminative models. These specific data mining methods that are being employed in SAP and dropout predictions in each category of the two models are briefly discussed as follows.

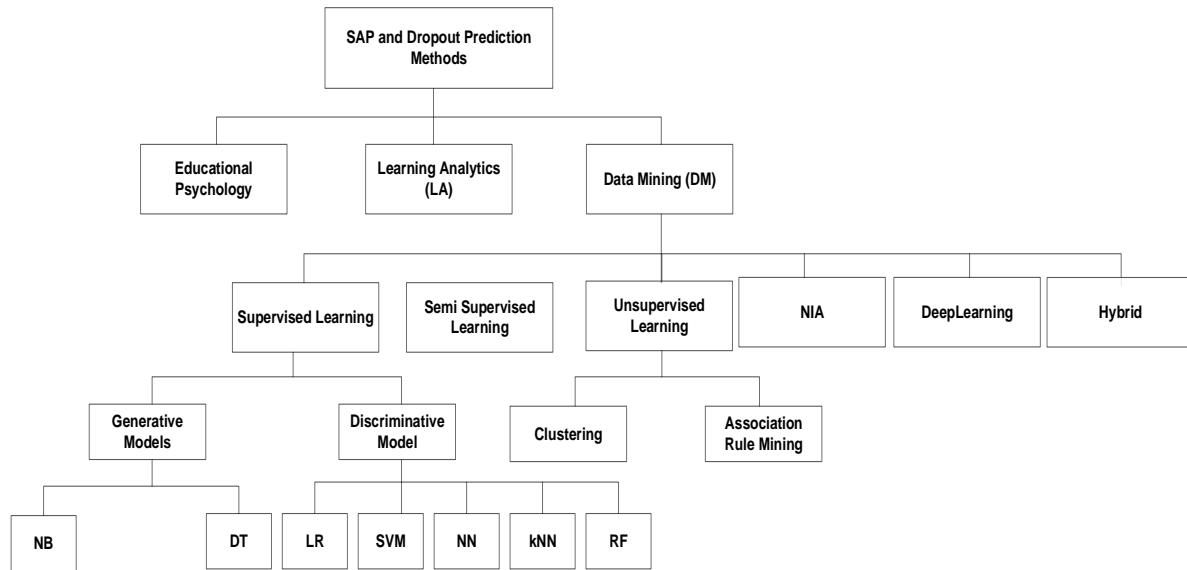


Figure 1. Taxonomy of methods for SAP and dropout predictions

4.1 Naïve Bayes (NB)

Naïve Bayes classifier uses the Bayes’ probability theory which assumes that all attributes of a given class in a dataset are independent. That is to say, it assumed that they all contribute equally to the outcome of classification task performed on the dataset – this is the strong (naïve) independence assumptions. It represents a descriptive and predictive approach to predict the class membership for a target tuple (Pandey and Pal, 2011). The Naïve Bayes Classifier works as follows:

- i. Given a training set of tuples D and their associated class labels C_1, C_2, \dots, C_n . If X represents each of the attributes in D which is to be classified.
- ii. We find the probability of each of the classes, $P(C_i)$ – class prior probabilities.
- iii. Compute the conditional probabilities for each of the components of X with respect to the classes C_i , that is $P(X|C_i)$.
- iv. Using the probabilities in (iii) above, we obtain the posterior probabilities $P(C_i|X)$ for each of the classes. By Bayes theorem; this is given as:

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \tag{1}$$

- v. The Naïve Bayes prediction will favour the class whose posterior probability $P(X|C_i)P(C_i)$ is maximum since $P(X)$ is constant for all classes.

That is to say that the classifier predicts that the label of tuple X is the class C_i if and only if

$$P(X|C_i)P(C_i) > P(X|C_j)P(C_j) \quad \text{for } 1 \leq j \leq n, j \neq i. \tag{2}$$

Hence, it is called the maximum posteriori probability.

Han, Kamber and Pei (2012)

Several studies have applied the Naïve Bayes method in predicting SAP and dropout, these works include Pal (2012) which used the Naïve Bayes method to manage the students’ dropout rate in an engineering college. Aziz, Ismail and Ahmad (2013) also presented a framework for predicting the performance of the first-year students in Computer Science course using merged data from two

academic databases. The presented framework contains four stages of data collection, data transformation, pattern extraction and prototype development. The study normalized into 6 parameters that were mined using Naïve Bayes classifier. The success of the pattern extraction stage was used to determine the last stage of prototype development. The study determined whether the parameters used contribute to the students' academic performance or not and also found out the relationship between independent parameters and the dependent parameter, that is GPA.

Aziz et al. (2015) also applied Naïve Bayes classifier (NBC) on an integrated dataset extracted from the academic databases of students in Unisza and the Student Entry Management Database (SEMD) to analyze the academic performance of students. The study uses Grade Point Average (GPA) as the dependent variable with labels: Poor, Average, to predict the students' performance using other five independent attributes. Their results showed that the NBC had the highest accuracy of 57.4% when 3-fold cross validation was used, also the most influential predictor of students' performance among the five independent parameters used was family income with 56.8% probability. In the overall, the study discovered that at 3-fold cross validation; the predictive model gives better classification for 'Average' students' category and failed to predict the poor student category.

Kaur, Singh and Josan (2015) used Naïve Bayes algorithm alongside with other algorithms like MLP, SMO, J48 to predict slow learners among students and displaying it by a predictive data mining model using classification-based algorithm. Dataset from a high school was used and tested with various classification algorithms (that is, Multi-Layer Perceptron, Naïve Bayes, SMO, J48 and REPTree). Measures such as TP Rate, FP Rate, precision, recall, F-measure, ROC area and accuracy were used to test and validate the resulting models. It was observed that the Multi-Layer Perceptron outperform all other classifiers in the study with accuracy of 75%.

Ahmad, Ismail and Aziz (2015) implemented NB coupled with DT and Rule-based algorithms to classify SAP with a few to compare the methods. Amrieh, Hamtini and Aljarah (2016) introduced a new feature of learning behavioural features to predict students' academic performance Naïve Bayesian, ANN and Decision Tree. The study applied ensemble methods of bagging, boosting and Random Forest to improve the performance of the three classifiers. The dataset was filtered using a filter feature selection method to rank the features extracted from the database of Kalboard of LMS. The experimental result of their study shows that the performance measures observed, that is accuracy, precision, recall and F-measure were high for ANN with and without behavioural features and when tested and validated the measures were all high in values for the prediction models.

Naïve Bayes was used alongside J48, a variant of DT by Kaur and Singh (2016) on a set of dummy data containing nine attributes. The results of their classification showed that Naïve Bayes which had 63.59% prediction accuracy outperformed the J48 algorithm with 61.53% accuracy. The comparisons between the techniques are mainly based on the prediction accuracy of the classifiers. Mueen, Zafar, & Manzoor (2016) also applied three different classification algorithms, Naïve Bayes, Neural Network and Decision Tree, on undergraduate students' dataset. Their results showed that the Naïve Bayes algorithm outperforms the other two algorithms with the highest predictive accuracy of 86%.

Makhtar, Nawang, Nor & Shamsuddin (2017) which uses Naïve Bayes algorithm with 10-fold cross validation to predict the relationship between subjects that affects students' academic performance in Sijil Pelajaran Malaysia institution. Their results showed that Naïve Bayes has the highest predictive accuracy of 80.94% among the four iterations carried out when compared with other selected DM techniques like random tree, nearest neighborhood, multiclass classifier and conjunctive rule. Amra & Maghari (2017) in their study applied both Naïve Bayes and k-Nearest Neighbor (k-NN) algorithms to dataset of secondary general certificate students of Gaza strip. Their results showed that Naïve Bayesian classification had the highest predicative accuracy of 93.17% when compared with that of the k-NN.

Four classifiers, namely: decision tree, random forest, Naïve Bayes and rule induction were applied on students dataset containing their historical records collected online from two schools in Portugal by Agrawal, Vishwakarma, & Sharma (2017). Their results showed that decision tree outperformed the other classifiers at different fold of cross validation of 10, 20, 30, 40 and 50, however the performances of all the classifiers used including the Naïve Bayes were in the same range and were very close.

Asif *et al.*, (2017) used a host of DM algorithms with the inclusion of NB coupled with a pragmatic policy to analyze the performance of undergraduate students' achievement at the end of a four-year study programme and the typical progressions of these students. The application of the two methods was combined to arrive at two groups of students (cohort 1 and 2) of low and high achieving students. The study combined three approaches of educational data mining namely: prediction, clustering and distillation of data for human judgment. In the prediction approach; it uses ten classifiers, six variants of decision tree algorithms, Neural Networks, Naïve Bayes, a rule-based classifier called the Rule Induction with Information Gain (RI-IG) and I-Nearest Neighbour classifier. X-means algorithm was used for the clustering approach. The distillation of data for human judgment in the study was in the form of what the authors called the pragmatic policy on the two cohorts to examine the revealed progressions as in the clusters. It was observed by the authors that the pragmatic policy was able to detect the two groups of students in the two cohorts under study.

Another study that also utilize the Naïve Bayes algorithm is the study of Razaque *et al.*, (2018) which applied the Naïve Bayes algorithm on the bachelor of Computing students dataset. The performances of NB were observed on the data clusters labeled C1 to C4 and the results showed the best predictive accuracy of 98.8% in cluster C3.

4.2 Decision Trees (DT)

Decision trees are widely used data mining techniques. They are built in such a way that they begin with an attribute at the root node (first node) which is further split into leave(s) and these leaves can also be split into other leaf nodes (internal node with one incoming edge and one or more outgoing edges) based on certain criteria. The tree terminates when we have the last node in the tree without any outgoing edges. In other words, a decision tree is a supervised classification technique that builds a top-down tree-like model from a given dataset using the attributes in it. The last node of a decision tree (i.e. the leaf node) represents the final (predicted) class label of a particular instance (Shalev-Shwartz and Ben-David, 2014).

All decision tree classifiers have two phases, namely: the growth phase or building phase and the pruning phase. The tree is built in the first phase by recursively splitting the training set based on local optimal criteria until all or most of the records belonging to each of the partitions bear the same class label. However, there may be overfitting, the effect of which is tackled by the second phase called pruning. Pruning generalizes the tree by removing noise and outliers to increase classification accuracy (Yadav, Bharadwaj and Pal, 2011). There are several variants of decision trees like the ID3 (Interactive Dichotomizer 3), C4.5 (or J48), C5.0 and CART (Classification and Regression Tree).

Studies that have employed various decision tree algorithms in SAP predictions include Ogor (2007), which applied C5.0 algorithm - a variant of DT, on students' dataset using the Clementine software to predict students' academic performance. The rules generated by the C5.0 decision tree algorithm were used in the development of prototype software written in Microsoft Visual Basic to monitor and evaluate students' academic performance. Some of the attributes considered in the work are AgeClass, CreditAttemptClass, Campus, CreditPassClass, the CGPAClass and so on. The system developed was able to predict accurately the students' performance as evidenced in the visualization results of predicted CGPA plotted against the actual graduating GPA.

Dekker, Pechenizkiy and Vleeshouwers (2009) used two variants of DT – CART & C4.5 along with other DM algorithms to generate predictive models to predict the Electrical Engineering students' dropouts. Yadav *et al.*, (2011) applied three most commonly used decision tree algorithms: CART, ID3 and C4.5 on a dataset of 48 students of Masters in Computer Application (MCA). WEKA explorer was used on the 7 parameters extracted from the database. The result obtained shows that CART was the best classifier among the three for the dataset with the highest correctly classified instances (56.25%) despite the fact that it has the highest execution time. Yadav and Pal (2012) study was similar to that of

Yadav *et al.*, (2011) in that same algorithms were implemented but to improve the engineering students' performance.

Ogunde and Ajibade (2014) used ID3 decision tree algorithm to relate the final graduation grades of students with the entry results. The study uses the data collected from the Academic department of Redeemer's University, Nigeria consisting of sex, student entry grades in secondary school (O'level), entrance examination scores and grades obtained at graduation (B.Sc) for all graduates between 2008/2009 and 2011/2012 sessions. ID3 algorithm was applied on the dataset using WEKA and the rules generated from the application were used to form the knowledge base part of the prediction system developed in Java. The prediction system developed was able to help in the prediction of the final grades of students on graduation even at the point of their entry into the university.

A weighted modified ID3 algorithm developed by Joseph and Devadas (2015) was used to predict students' performance using 56 instances from dataset obtained from the first batch students of the department of CSE, College of Engineering Munnar. It was observed that the modified weighted ID3 (with 76% prediction accuracy) outperforms C4.5 (45.83%), ID3 (52.08%) and CART (56.25%). The study used six related attributes of previous semester mark, class test mark, and assignment performance, attendance in class, lab work and End Semester mark. The focus of the study was only to test the efficiency of the modified algorithm.

The three DT algorithms used by Yadav *et al.*, (2011) were implemented differently with CHAID algorithm by Saa (2016) to generate predictive classifiers using multiple features of SAP. The study uses both RapidMiner and WEKA tools to apply C4.5, ID3, CART and CHAID (Chi-Square Automatic Interactive Detection) algorithms on the dataset with 10-fold cross-validation to verify and validate the outcomes of the four algorithms. Using accuracy and precision measures, the result of the study shows that CART had the best accuracy of 40%, followed by C4.5 (35.18%), then CHAID (34.07%) and the least is the ID3 algorithm with 33.33%. The study also applied Naïve Bayes classifier which gives an accuracy of 36.40%. The study compares the algorithms based on prediction accuracy only. SAP prediction models were also developed from DT combined with fuzzy genetic algorithm in Hamsa, Indiradevi and Kizhakkethottam (2016). The study modeled students' academic performance prediction for Bachelor and Master Degree students in Computer Science and Electronic and Communications stream using decision tree and fuzzy genetic algorithm. Parameters selected for the study includes internal marks, sessional marks and admission score. The application of the two techniques on the dataset revealed that there are more students at risk class when a decision tree is used while the fuzzy genetic algorithm gives more passed students. Asif, Hina and Haque (2017) used some variants of DT together with a variant of ensemble method RF to predict students' graduation performance in the final year.

The students' final GPA was predicted using J48 algorithm – a variant of DT by Al-barrak and Al-razgan (2016) where J48 decision tree algorithm was applied to King Saud University Computer Sciences College students of the year 2012 dataset using WEKA tool. The result of the mining shows that among all the courses; Software Engineering was the most important mandatory course in the determination of students' final grade, Java2 course among all programming languages was closely related to the final GPA more than Java1. It was discovered that the addition of students' grade in Data Structure course confirmed the influential nature of Java2 as observed earlier and when J48 algorithm was run repeatedly for six times, each run representing each semester; the resultant tree showed that Java1 is the most influential course for the determination of final GPA.

Afeni, Oloyede & Okurinboye (2019) applied ID3 and C4.5 algorithms on the data of students of Joseph Ayo Babalola University in six academic departments to predict their performances. The result of the prediction model was validated using four performance metrics of accuracy, sensitivity, false alarm rate and precision. The results obtained showed that the ID3 algorithm outperform the C4.5 algorithm with accuracy value of 61%, True Positive rate value of 1.000, False positive rate of 0 and precision of 1.000.

4.3 Logistic Regression (LR)

This is a linear model that model categorical response variables. It estimates the probability that the dependent variable will have a given value. Logistic regression is only used when the output variable of the model is defined as a categorical binary, though the inputs can be quantitative. After having calculated the logistic regression probability for all inputs, the conclusion can be made based on the final value of the probability of the output so as to know the more probable categorical value whether the one with value 0 or the one with value 1 (Kantardizc, 2011).

Logistic regression was used for SAP predictions in some studies like that of Agarwal, Pandey and Tiwari (2012). The authors used LR and other DM algorithms to generate several classification models that predict SAP. In the study, several classification algorithms were applied on a dataset from a community college database. The dataset has 4 attributes and 2000 records of students' performance details. The open source machine learning tool, WEKA, was used to generate classifiers models using the dataset. Eight classifiers, namely: logistic regression, multi-layer perceptron, Support Vector Machine (SVM), RBFNetwork, Voted Perceptron, SIMD, Winnow and simple logistic. The result shows that the Support Vector Machine is the best classifier among others in the study with maximum accuracy and minimum root mean square error (RMSE). The comparison is based only on the prediction accuracy. LR was also used alongside DT and NN by Simeunović and Preradović, (2014) to predict students' success in their studies.

Aulck et al., (2016) modeled students drop-out using data gathered from the registrar databases of the University of Washington in the USA. The dataset contains 32,538 students' data and applied regularized logistic regression, k-nearest neighbour and Random Forest to predict the binary drop-out on the features like race, gender, resident status, GPA and so on totalling 784 additional features. Their results indicate that predicting eventual student attrition from a balanced dataset of over 32,500 students with regularized logistic regression provide the strongest prediction. GPA in Math, English, elementary and psychology courses was among the strongest individual predictors of attrition.

4.4 Support Vector Machines (SVM)

Support Vector Machine (SVM) is an algorithm equipped with special features. It is based on the idea of 'kernel trick', that is the way to solve a non-linear separation problem by mapping the original non-linearly separable points into a higher-dimensional space, where a linear classifier is subsequently used (Gorunescu, 2011). SVM provides a solution to the problems of computational complexity and overfitting as inherent in other learning algorithms. It is based on an algorithm that finds a special kind of linear model, the maximum margin hyperplane. The maximum margin hyperplane is the one that gives the greatest separation between the classes, in a two-class dataset whose classes are linearly separable (Witten and Frank, 2005). SVM supports both classification and regression tasks and can handle multiple continuous and categorical variables.

In other words, support vector machine is primarily a classifier that performs classification tasks by constructing hyperplanes (linear models) in a multidimensional space that separate cases of different class labels. New examples are mapped into that same space and then predicted to belong to a category based on which side of the gap they fall (Rajeshinigo and Jebamalar, 2017). SAP and dropout predictions studies that employed the Support Vector Machines include Tekin (2014) where SVM, NN, and ELM were separately used in predicting students' GPA at graduation.

Bhagvatula et al., (2015) also used SVM, NB and J48 algorithms in their separate implementations to predict students' performance with a view to investigating the contribution of students' academic efforts. A machine learning framework was developed using SVM and four other DM algorithms by Lakkaraju et al., (2015) to identify students at risk of not graduating from the high schools in the two districts employed to test the framework.

Asogbon, Samuel, Omisore, & Ojokoh (2016) which developed a multi-class support vector machine to predict the students' performance using the educational dataset of students from the University of Lagos. The multi-class SVM was able to predict the performance of students adequately using 7-fold cross validation.

Eashwar, Venkatesan & Ganesh (2017) applied SVM on Post Graduate (PG) students dataset collected through questionnaire to predict students who were at the edge of their performances while

using k-means for clustering. Their results showed that a predictive accuracy of 96.7% was achieved with SVM. SVM was also used with four other algorithms by Rovira, Puertas and Igual (2017) to predict academic grades and dropout with a view to developing a data-driven prediction system. In the study, a data-driven system was developed to predict the academic grades of students in different courses and the dropout. Five selected data mining algorithms, namely: Logistic Regression, Gaussian Naïve Bayes, Support Vector Machines, Random Forest and Adaptive Boosting. The performances of these algorithms were evaluated and the best of them was chosen for prediction. Olorunfoba & Akinode (2017) applied SVM on students' dataset to investigate the relationships between students' preadmission academic profile and final academic performance. The study tuned the SVM parameters and found that the RBF kernel with Cost = 100 gave the best predictive accuracy of 97%. When the result was compared with that of other machine learning techniques such as k-NN, decision trees and linear regression, it was found that SVM outperforms all of them.

In the study of Burman & Som (2019), a multi-classifier SVM with linear and radial basis functions (RBF) kernels were used to classify students' performances. The two variants of SVM were applied on students' records containing psychological parameters collected through questionnaires. Their results showed that RBF kernel SVM outperform its linear counterpart with a higher predictive accuracy of 90.97%.

4.5 Neural Network (NN)

The neural network is a set of connected input/output units where each connection has a weight attached to it. It was modeled after the biological neurons of human brains. The network is formed from large numbers of simulated neurons, connected to each other in a manner similar to brain neurons. Like in the human brain, the strength of neuron connections may change (or be changed by learning algorithm) in response to a presented stimulus or an obtained output, which enable the network to learn. During the learning phase, network learns by adjusting weights so as to be able to predict the correct class labels of the input tuples (Gorunescu, 2011; Akinola, Akinkunmi and Alo, 2012).

An artificial neuron can be seen as a single device that has a certain number ' p ' of real inputs x_i , weighted by respective weights w_i , summed and then passed on to an activation function ϕ to produce a certain output, depending on a predetermined 'threshold' T (Gorunescu, 2011). Multilayer Perceptron is a kind of neural networks that have one or more hidden layers whose (computing) elements are called hidden neurons, which are meant to serve as intermediaries between the input layer and the output layer. Information from the environment enters the network through the input layer and then processed by the second layer (the first hidden layer) which then becomes an input to the next hidden layer. This continues till the whole hidden layers are served and the outcome is then passed to the output layer.

Research studies that used NN include Akinola *et al.*, (2012) which used ANN to predict the computer programming proficiency of undergraduate students. It applied Multi-Layer Perceptron feed-forward back-propagation neural network to the enrollment data of 200 level Computer Science students with their results in both 100 level Mathematics and Physics and a programming course at 200 level, using an open source tool called Neuroshell. The results from the study showed that a priori knowledge of Physics and Mathematics are essential in order for a student to excel in computer programming. Simeunović and Preradović (2014) also used NN with DT and LR separate implementations to predict student success in their studies. Osofisan, Adeyemo and Oluwasusi (2014) used MLP, a variant of NN alongside J48 algorithm to study the behaviour of student's performance data. The study applied the two algorithms on M.Sc students' data from Computer Science department of the University of Ibadan, Nigeria to investigate their performances in mining education data. The comparative performance analysis of the two models ran in WEKA environment revealed that MLP (98.3%) outperformed the J48 algorithm (85.4%) in prediction accuracy despite the fact that the time taken to build the J48 model (0.25secs) was lesser than that of the MLP which is 2.7secs on the training dataset. The same trend repeated itself for the test dataset with 60.2% accuracy and 5.93secs for MLP and 52.8% accuracy and 0.04secs for the J48 model. The study, therefore, concluded that Artificial Neural Network gives the best classification results as well as prediction capability in mining education data.

Ruby and David (2015) used MLP to predict SAP and analyzing influencing factors in SAP predictions. The study used a Multi-Layer Perceptron (MLP) algorithm to model students' performance

using data collected from PG Computer Application course. WEKA open source software was used to generate the MLP models for the two datasets of high influential factors (7 attributes) and the whole dataset of 12 attributes. Model accuracy was done and their results show that the models of the high influential factors perform better than that of the whole 12 attributes. Asif *et al.*, (2017) also used NN together with other different algorithms combined with a pragmatic policy to analyze the performance of undergraduate students.

4.6 k-Nearest Neighbour (kNN)

This is a classification method in which a new object is labelled based on its closest (k) neighbouring objects. According to (Gorunescu, 2011), in principle given a training dataset and new object to be classified, the distance (some kind of similarity) between the new object and the training objects is first computed and the nearest k objects are then chosen. The algorithm, therefore, has three requirements, namely: a set of stored records (training dataset), a distance (metric) to compute the similarity between objects and then the value of 'k'. This technique was employed in studies like Asif, Merceron, *et al.*, (2017) with series of methods to analyze the students' performance. Aulck *et al.*, (2016) used kNN together with other methods to predict students' dropout.

Verma, Singh & Verma (2016) used kNN among other algorithms to predict SAP for their performance evaluation. The study applied six algorithms, namely: J48 (C4.5), two Bayesian classifiers – Naïve Bayes and BayesNet, an instance based learner (k-NN), and two rule learners – OneR and JRip to the dataset of the students of SPSU University containing 10,330 records described by twenty parameters. The parameters used include gender, birth year, place, current semester total university score and so on. The results of the study showed that students' university admission score and a number of failures in the first year exams were among the highly influencing factors in the classification. Both 10-fold cross-validation and percentage split test were used in the evaluation of the performance of the algorithms. It was found out that in the overall, the J48 algorithm outperformed others using the overall weighted average values in the two evaluations. Kabakchieva (2012) also used IBk, a variant of kNN among other algorithms to predict students' performance using the holdout method of percentage split.

4.7 Random Forest (RF)

This is an ensemble method that has several decision tree classifiers in it so that their collection becomes a forest. The decision tree components are generated by random selection of attributes at each node to determine the split. Each tree depends on the value of a random vector sampled independently and with the same distribution for all trees in the forest. When RF is employed for classification, each trees votes and the most popular class is returned (Han *et al.*, 2012).

Random Forest was used by Gilbert (2017) to predict student outcomes. RF and genetic algorithm (GA) were applied on California State University freshmen and transfer students data from Fall 2000 through Fall 2010 of over 31,000 students. The study was able to capture the interactions between factors that might be otherwise missed in a linear system through the use of RF – a non-linear ensemble method, and also able to optimize feature selection process thereby striking a balance between recall and precision by using GA. The results of the study show that there is strong predictive capability for 1 and 2 year retention periods and graduation outcomes.

RF was also employed by Mishra, Kumar and Gupta (2014) together with J48 algorithm to predict students' performance using their social and academic integration features. The study explored the link between emotional skills of the students along with socio-economic and previous academic performance parameters to predict academic performance using data mining techniques. The emotional skills like assertion, leadership, stress management were obtained using standard Emotional Skill Assessment Process (ESAP). The results of the study showed among other things that out of all emotional attributes leadership and drive of the students have been found to affect the performance. The second approach used in data mining is the unsupervised learning approach. An unsupervised learning approach is the opposite of the supervised learning and it is, therefore, an approach where the class labels are unknown or there is no label at all. Clustering method and the Association rule mining are examples of unsupervised learning approach employed in the study of SAP predictions.

4.8 Clustering

Clustering is a process of grouping a set of data objects into multiple groups or clusters so that objects within a cluster have high similarity but different from the objects in other clusters (Han *et al.*, 2012; Witten and Frank, 2005). Unlike classification, the label is unknown. Examples of clustering algorithms are the k-means algorithm and the k-medoid algorithm. This technique was combined with other common DM methods in the work of Asif *et al.*, (2017) used X-means algorithm, a modification of k-means clustering algorithm to cluster the students in the datasets into cohorts.

4.9 Association Rule Mining

Association rule mining involves the act of finding strong rules as a way of correlation between different item-set in a large dataset. This is done in the following two phases:

- Discovering the large item set, that is, the sets of items that have transaction support ‘s’ above a predetermined minimum threshold; and
- Use the large item set to generate the association rules for the database that have confidence ‘c’ above a predetermined minimum threshold (Kantardizc, 2011).

A very prominent algorithm used in generating association rules is the Apriori algorithm. The study of Oladipupo *et al.*, (2017) used the Apriori algorithm to generate strong rules for the assessment of the influence of attendance on SAP prediction. The study examined the influence of class attendance on the academic performance of students using association rule mining and it was discovered that the influence of class attendance on the academic performance of students was low. This according to the study indicates that class attendance is not the only factors that determine the academic performance of students.

A version of Apriori algorithm called the AprioriC algorithm was compared with other DM techniques by Romero *et al.*, (2008) to classify students. This study classified students to FAIL, PASS, GOOD and EXCELLENT by applying several algorithms in the KEEL framework. It used the KEEL framework to develop a Moodle Data Mining tool experiment on the dataset. Results showed that the best algorithms (with more than 65% global percentage correctly classified – PCC) with the original data used are CART, GAP, Grammar-based Genetic Programming (GGP), and Neural Network Evolutionary Programming (NNEP), the best algorithms (with over 65% global PCC) using the categorical data are CART and C4.5 and the best algorithms (with over 60% global PCC) using the balanced data are Corcoran, XCS, AprioriC and MaxLogicBoost. It was also observed that no algorithm exceeds 70% global PCC results which the authors attributed to the incomplete data used. The moodle DM tool developed does not give an overall generalization to include real traditional performance dataset.

4.10 Nature-Inspired Algorithms

Nature Inspired algorithms are computational intelligence algorithms that evolved through the wonderful behavior observed of some biological system in a particular environment. This observed cooperation and foraging among these natural systems are then adapted to solve various optimization problems – finding optimal solution whether in a local situation or globally (Du & Swamy, 2016). There are many of these algorithms, however some popularly used ones include Genetic Algorithm, Genetic Programming (GP), Ant Colony Optimization (ACO), Particle Swarm Optimisation (PSO), Cuckoo Optimization Algorithm (COA) and so on.

Studies that uses these category of techniques for predicting students’ academic performance and dropouts include Yildiz, Bal, Gulsecen & Kentli (2012) which modelled the education data of students to predict their academic performances using fuzzy logic and then applied Genetic Algorithm for optimization of the model for better predictive accuracy of 84.52% when compared to the accuracy of the fuzzy logic model.

Chen, Hsieh & Do (2014) used the standard CS and COA train the feed forward neural network for prediction of SAP. CS and COA were used to optimize the weights between layers and biases of the

neural network. Simulation results showed that neural network was well improved by these two algorithms for predicting SAP; however, the ANN-COA outperforms the ANN-CS in terms of RMSE, MAPE and R obtained.

Another study is the work of Marquez-Vera *et al.*, (2016) the authors predict the students' dropouts at several stages of the course being learned using an Interpretable Classification Rule Mining (ICRM) which has a variant of GP, specifically Grammar Based Genetic Programming (GBGP) as its core. The results obtained in their various iterations of their experiments showed that the proposed method is better in prediction when compared to other traditional DM techniques like the SMO, Naïve Bayes and so on.

The study of Hasheminijad & Sarvmili (2019) proposed a rule-based method named S3PSO method to predict the students' academic performances using the Particle Swarm Optimization algorithm. The method uses Association Rule Mining to extract rules that are used to predict students' performances. The result obtained showed that S3PSO improved by 31% of the fitness function when compared with other rule-based classification algorithms like CART, ID3 and C4.5. It also improved by 9% of accuracy when compared with traditional DM techniques like SVM, KNN, Naïve Bayes and so on.

4.11 Deep Learning

Deep Learning technique of data mining is a technique whereby complex functions are approximated to the same accuracy using a 'deep' architecture. By 'deep'; we mean using multiple layers with fewer number of neurons in total (Zhou, Greenspan & Shen, 2017). The basic building blocks of most deep learning models are the auto-encoders like we have it in Deep Neural Network (DNN). However, the Deep Generative models like Deep Belief Network (DBN) and Deep Boltzmann Machine (DBM) uses Restricted Boltzmann Machines (RBM) as their building blocks (Zhou, et al, 2017).

Another common form of deep learning model is the Convolutional Neural Network (CNN) as mentioned in Aggarwal (2018). Studies that made use of this techniques for predicting students' performance are still very scanty in the literature, however some of the existing ones include the studies of Bendangnuksung & Prabu (2018) which uses Deep Neural Network to predict students' performances and Kim, Vizitei & Ganapathi (2018).

4.12 Hybrid Methods

Hybrid methods of predicting SAP and dropouts are the methods which combines two or more traditional DM techniques to achieve a better result. Studies that have evolved in this regard include the work of Tran, Dang, Dinh, Truong, Vuong & Phan (2017) proposed a hybrid prediction model which combines collaborative filtering strategy using matrix factorization method with linear regression strategy using SVM algorithm with training of dataset with the inclusion of skills-related features from 1268 undergraduate students of the National University in Vietnam. The results of their experiment with the hybrid and the single implementation of the component algorithms showed that the hybrid outperformed them using the Root Mean Square Error (RMSE) as the performance metric.

Another study is that of Altaher & Barukab (2018) which proposed a hybrid model combining an Adaptive Neuro-Fuzzy Inference System (ANFIS) with Genetic Algorithm (GA) for the prediction of SAP. SAP dataset of 100 Computer Science students' records were passed into the ANFIS and the result of training and testing were later optimized using GA. The developed hybrid model, HGANFIS was compared with Neural Network (NN) and ANFIS approaches using the RMSE as the performance metric. The results show that the HGANFIS model performed better than the other two algorithms by having the least RMSE values of 0.101 and 0.104 for both training and testing data respectively.

Another important study is that of Chen, Feng, Sun, Wu, Yang & Chen (2019) which combined Decision Tree (DT) algorithm with Extreme Learning Machine (ELM) for predicting MOOC (Massive Open Online Course) dropout. It uses the DT algorithm to extract important features from MOOC students' learning behavior records. This DT feature extraction process were later map with the ELM algorithm based on entry 2015 datasets and its results compared with eight other algorithms, namely: GA-ELM, DT, SVM, LR, BP (Backward Propagation Neural Network), EN (Entropy-Net), ELM and LSTM. It was observed that the DT-ELM model performed better than all others in terms of Accuracy

(0.9642), AUC (0.9412) and F1-score (0.9667) obtained for week 4 among the five weeks observations in their experiments.

Also noteworthy is the study of Francis & Babu (2019) which proposed a hybrid prediction model that uses four algorithms of SVM, Naïve Bayes, DT and NN for selecting the best features from real life datasets of students in various disciplines in higher education institutions in a wrapper feature selection mode. The selected features were then passed to the K-means clustering algorithm to predict students' academic performance using majority vote approach. The feature selection phase of the model showed that the combination of academic, behavioural and extra features gives the best accuracy (0.7547) and the new model yielded the best result when compared with DT and NN in terms of precision, recall, F-score and accuracy values of 0.6415 respectively.

The taxonomy presented in this section addresses the popular methods used in SAP and dropout prediction, there are however some researches that used methods like the works of Mohsin *et al.*, (2010) which employed Rough set to determine the factors influencing the final programming practice of students. A multi-model heterogeneous ensemble approach was also applied to the prediction of students' academic performance in Adejo & Connolly (2018). The study used data from multi-sources: institution database, learning management system and survey and applied ensemble of three basic classifiers of decision tree, artificial neural network and support vector machine to predict students' academic performance in an efficient and accurate manner.

Yang *et al.*, (2018) also used multiple linear regressions together with principal component analysis to predict SAP. A neuro-fuzzy approach was also applied to the classification of SAP in Do & Chen (2013). In this study, the neuro-fuzzy classifier was trained with several algorithms including Kalman-filter; Levenberg Marquardt method and others and using 100 iterations with a 10-fold cross-validation to avoid overfitting. The model was coded and implemented in MATLAB R2011b and simulation results obtained. The efficiency of the classifier was determined by comparing the predicted and actual class labels for the testing dataset. The approach was later compared with well-known classifiers like SVM, Naïve Bayes, neural network and decision tree and the result indicated that the presented neuro-fuzzy approach performed better than others. Strecht *et al.*, (2015) did a statistically inclined evaluation of a host of regression and classification algorithms incorporating both common and uncommon techniques for SAP predictions.

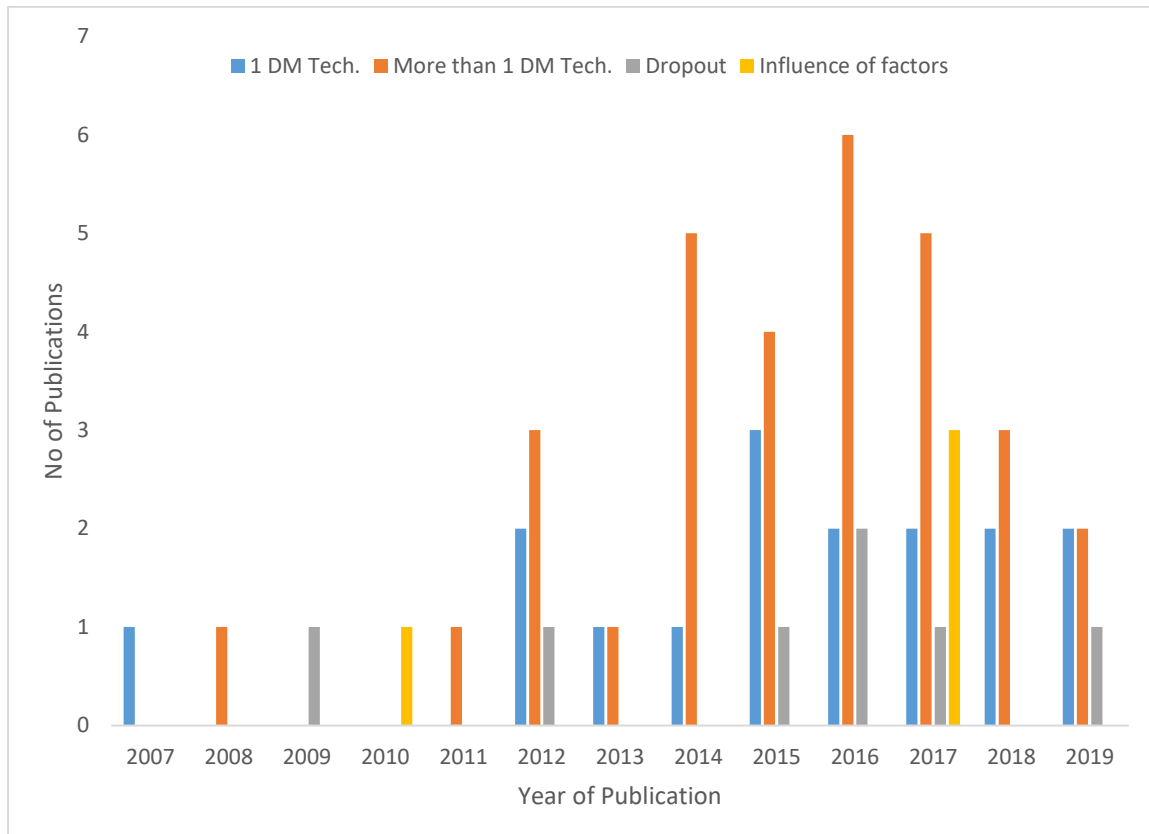


Figure 2. Distribution of the selected papers by year of publications

Table 1. Summary of SAP and Dropout prediction researches from 2007 to 2019

YEAR	AUTHOR(S)	OBJECTIVE	DM TECHNIQUE	LIMITATIONS
2007	Ogor	Development of a methodology through performance prediction for deployment of a student performance assessment & monitoring system	C5.0	Lack the integration of students' attitude and so on to the result of mining student performance.
2008	Romero, Ventura, Espejo & Hervas	Comparison of different DM methods and techniques for classifying students	CART, GAP, GGP, AprioriC and so on.	The Moodle DM tool developed does not give an overall generalization of prediction of SAP for all forms of learning.
2009	Dekker, Pechenizkiy & Vleeshouwer	Generation of predictive models form selected DM techniques to predict Electrical Engineering students' dropouts	CART, C4.5, BayesNet, Logistic model, JRip & RF.	Students' progression was not taken into consideration in the study.
2010	Mohsin, Norwawi, Hibadullahi & AbdWahab	Investigation of factors influencing the final programming performance	Rough Set	Prominent DM techniques like NB, SVM need to be applied on the dataset with other attributes added to confirm results.
2011	Yadav, Bharadwaj & Pal	Development of DT-based models to predict students' performance.	CART, ID3 & C4.5	Only one evaluation criteria were used and that is accuracy which is not enough to conclude.

2012	Akinola, Akinkumi & Alo	To predict the computer programming proficiency of undergraduate students	ANN	The study did not experiment with other DM techniques to ascertain their performance.
2012	Agarwal, Pandey & Tiwari	Generation of classification models using some classification algorithms	LR, MLP, SVM, RBF network, Voted Perceptron, SIMD, Winnow & Simple Logistic.	The study did not consider other performance metrics apart from the prediction accuracy.
2012	Kabakchieva	Predicting Students performance using DM classification algorithms	OneR, J48, MLP, kNN	Classification algorithms like SVM and NB were not experimented to confirm and compare performance with the selected one in the study.
2012	Yadav and Pal	Performance improvement of Engineering students using DM	C4.5, ID3, CART	Only DT techniques were used, other prominent DM techniques were not considered.
2012	Pal	Development of a DM application to generate predictive models for Engineering student's dropout management	NB	The variables used in the study do not include any social behaviour feature.
2012	Yildiz et al.	Using model to evaluate the Distance Education Students' academic performances.	Genetic-Fuzzy hybrid.	Only meta-heuristic algorithms used, no focus on any of the traditional DM algorithms.
2013	Do & Chen	Presentation of a Neuro-fuzzy approach to classification of SAP	Neuro-Fuzzy	The approach being fuzzy based can be rigid with rigid fuzzy rule sets.
2013	Aziz, Ismail & Ahmad	Development of a prototype SAP prediction model	NB	The study did not take the student's social behaviour into consideration.
2014	Simeunovic and Preradovic	Predicting student success in studying using Data Mining	LR, DT, NN	Other commonly used DM algorithms in the SAP prediction community like SVM and NB are not employed and can limit the acceptance of the model among the EDM researchers.
2014	Tekin, A.	Predicting students' GPA at graduation using several DM techniques	NN, SVM and ELM	The attributes considered are purely academic and no socio-demographic data was used. This would have improved the accuracy of the predictions and make it very near to reality.
2014	Mishra, Kumar & Gupta	Performing Prediction using students' social integration, academic integration, and emotional skills using Data Mining	J48 and RF	Results cannot be generalized as only the variants of DT are used in the study.
2014	Osofisan, Adeyemi & Oluwasusi	Using DM techniques in studying the behaviour of student's performance data	MLP and J48	The only two techniques compared in the study were not enough to pick the best from several DM techniques for education data.

2014	Ogunde & Ajibade	Development of a DM system to predict student's graduation grades based on their entry results data using ID3	ID3	More distinctive attributes like social behaviour of students that can have a direct impact on the final prediction needs to be considered for better accuracy.
2014	Chen, Hsieh & Do	To train Neural Network with two meta-heuristic algorithms of CS and COA for SAP prediction.	NN trained with CS & COA	It only serves as an effective standard for NN and not for other DM algorithms.
2015	Aziz, Ismail, Ahmad & Hassan	Development of a framework for predicting the performance of 1 st year Bachelor students in Computer Science	NB	Social behaviour of students was not considered in the study.
2015	Joseph & Devadas	Development of a DT-based classification model to predict SAP	Weighted modified ID3	The developed model was evaluated using only the prediction accuracy as the performance metric.
2015	Lakkaraju et al.	Development of a machine learning framework to identify students at risk	RF, AB, LR, SVM, DT	The study needs to employ other commonly used EDM techniques like NB for generalization of results.
2015	Kaur, Singh & Josan	Prediction of slow learners	MLP, NB, SMO, J48 & REPTree	Some other important factors like social behaviour of students are not considered in the study.
2015	Strecht, Cruz, Soares, Mendes-Moreira & Abreu	Evaluation of the most popular classification and regression algorithms for predicting SAP	SVM, CART, NB, RF, kNN, AB, AB.R2, OLS	The evaluation criteria used in the study is statistically inclined and is not goal oriented performance metrics used in data mining.
2015	Ahmad, Ismail & Aziz	Comparative analysis of selected classification algorithms	DT, NB & Rule-Based (RB)	Only prediction accuracy was used as a performance metric to determine the best of the selected algorithms.
2015	Ruby & David	Analysing influencing factors in SAP predictions	MLP	Social behaviour of students not considered.
2015	Bhagvatula, Shreyas, Gupta & Singh	Investigation of the contribution of students' academic efforts on the prediction of students' performance.	SVM, NB, J48	The study uses only the prediction accuracy as a performance metric.
2016	Marquez-Vera et al.	Predict dropout of high school students at early stage.	ICRM2 with GBGP core.	The prediction in the study does not give room for accumulated performance of the students.
2016	Mueen, Zafar & Manzoor	To predict students' academic performance.	NB, NN & DT	Only predictive accuracy is employed as performance metric.
2016	Asogbon et al.	To investigate the multi-class ability of SVM for SAP prediction	SVM	Model developed met the target but other traditional DM techniques may outperform it.
2016	Amrieh, Hamtini & Aljarah	Introduction of the new feature of learners' behaviours to SAP prediction using several classifiers.	NB, DT, ANN & Ensemble methods	No significant consideration is giving to social behaviour attributes.
2016	Saa	Generating predictive classifiers from the use of multiple features of SAP	C4.5, ID3, CART & CHAID	The volume of the dataset used was not large enough to give more acceptable prediction accuracy.

2016	Kaur & Singh	Finding the best DM technique from two selected for mining	NB, J48	The only two techniques compared in the study are not enough to pick the best from several DM techniques for education data.
2016	Verma, Singh & Verma	Performance evaluation of selected DM algorithms	J48, NB, JRip, OneR, BayesNet, kNN	Only the prediction accuracy was used as a performance metric.
2016	Hamsa, Indiradevi & Kizhakethottam	Development of a SAP prediction model using DT and fuzzy-genetic algorithm.	DT, Fuzzy-genetic	The approach needs to be tested using a large dataset.
2016	Aulck, Velagapudi, Blumenstock & West	Predicting Students dropout	Regularised LR, kNN and RF	The DM methods employed are not the most popular in EDM researches and may limit research result acceptance.
2016	Al-Barrak & Al-Razgan	Prediction of student's GPA based on grades in previous courses	J48	The focus is mainly on the grades in courses not on other influencing factors on GPA
2017	Asif, Merceron, Ali & Haider	To analyze the performance of undergraduate students using EDM	DT, NN, NB, RI-IG, k-NN, X-means with pragmatic policy	The study took student progression into consideration but there is a need for generalizability of the results of DM.
2017	Asif, Hina & Haque	Studying DM techniques to predict the students' graduation performance in the final year.	DT-GI, DT-IG, DT-ACC, NB, NN, RF-GI, RF-IG, RF-ACC	Some other important DM techniques were not studied apart from the DT variants.
2017	Gilbert	Prediction of student outcomes using RF and genetic algorithm for feature selection and pruning	Random Forest (RF)	The DM algorithm employed is not a common one in the area and therefore the obtained result may not be popular.
2017	Rovira, Puertas & Igual	Development of a data-driven system to predict academic grades and dropout	LR, Gaussian NB, SVM, RF, and AB.	There is no focus on standardization of methods rather on the automatic extraction of data for prediction.
2017	Oladipupo, Daramola, Oyelade & Afolabi	Assessment of the influence of attendance on SAP using DM	Association Rule mining	The result indicated the low influence of class attendance; however, some other DM techniques need to be investigated to confirm this.
2017	Makhtar, Nawang, Nor & Shamsuddin	Investigate the relationship between subjects and students' performances.	NB	Only predictive accuracy is employed.
2017	Agrawal, Vishwakarma & Sharma	To use data mining technique to predict undergraduate students' performances.	NB, RF, DT & Rule Induction	Only predictive accuracy is employed.
2017	Amra & Maghari	Predicting SAP using KNN & NB	KNN & NB	Only predictive accuracy is employed.
2017	Eashwar, Venkatesan & Ganesh	Predicting the Post Graduate SAP using SVM	SVM	Only predictive accuracy is employed.
2017	Oloruntoba & Akinode	To investigate the relationships between students' preadmission	SVM	Focus is on influence of features not the classifier's performance.

		academic profile and final academic performance		
2017	Tran <i>et al.</i>	Develop a model for student performance prediction using multiple strategies	CF & SVM	Only RMSE was used as the performance metric. Other popular metrics would have been measured too.
2018	Attaher & Barukab	Developing an intelligent hybrid model of Adaptive Neuro-fuzzy and GA for SAP prediction.	Adaptive Neuro-Fuzzy and GA	Need for investigating the same with other popular traditional DM techniques.
2018	Yang, Lu, Huang, Huang, Ogata & Lin	Improving Multi LR predictive capability with PCA	Multiple Linear Regression, PCA	Prediction model developed is strictly associated with the learning activity and particular data in the study thereby making the model non-flexible.
2018	Adejo & Connolly	Predicting SAP using a multi-model heterogeneous ensemble approach	ANN, SVM, DT, Ensemble	Non-generalizable models.
2018	Bendangnuksung & Prabu	Using Deep Neural Network to predict students' performances.	Deep NN	The study can be improved by adding some other DM techniques.
2018	Razaque <i>et al.</i>	Using NB to perform SAP prediction	NB	Non-generalizable models.
2019	Hasheminejad & Sarvmili	Performing SAP prediction using PSO	PSO	There is need for inclusion of any of the traditional DM techniques.
2019	Burman & Som	Investing RBF and Linear kernels for SVM prediction of SAP	SVM	Focus is only on finding the best among the two kernels of SVM
2019	Afeni, Oloyede & Okurinboye	Predicting Students' Performance using classification Algorithm	ID3 & C4.5	Comparison of classifiers is only among Decision Trees. There is need to extend to other DM techniques.
2019	Chen <i>et al.</i>	Developing a hybrid of DT & ELM for dropout prediction	DT & ELM	Can be used to provoke hybrid of two or more other DM techniques
2019	Francis & Babu	Predicting SAP using Hybrid	SVM, NB, DT & NN with K-means algorithm.	The study only finds the best based on selected group in dataset

5. Taxonomy Based on Features

The features employed in the existing SAP and Dropout studies are very diverse. A taxonomy of these features as presented in Figure 3 of this paper shows that all these features can be classified into either internal feature of students under study or external features linked to the students owing to their relationship with their environments. The internal features can be further divided into two, namely: personal and psychological features. The external features are classified into four categories as academic features, social features, economic features and demographic features.

5.1 Personal features

These are features that are directly reflecting the students in the case studies without any link to the students' outside world. These features include Age, sex/gender, disability status, health status, average study hour, mode of study among others. All existing studies on SAP and dropout employed some of

these variables except few ones like the studies of Ogor (2007); Ogunde and Ajibade (2014); Hamsa *et al.*, (2016).

5.2 Psychological features

These are features that describe the mental development of the students used as a case study in the SAP and dropout researches. They are most emphasized in the dropout studies. These features include stress management ability, first generation learner, perceived academic emotion, perceived academic control, learning style and course satisfaction among others. These features were employed in the studies of Mishra *et al.*, (2014); Ruby and David (2015); Respondek *et al.*, (2017).

5.3 Academic features

These are features in the datasets that are purely related to the academic activities of the students under study. Academic features are classified into two, namely: the pre-Higher Educational Institution (HEI) academic features and the Higher Educational Institution (HEI) academic features.

5.3.1 Pre-Higher Educational Institution (HEI) academic features

These features are also called the pre-University academic features when the case study is a university. They are features acquired by the students before their entrance into the Higher Educational Institutions. They include the admission score, admission type, entrance exams scores, Higher School Grade (HSG) exam, place of previous education, pre-university subjects' grades or scores and so on. These features are mostly used by SAP studies that tend to predict the student's final grade on graduation at the university using the pre-university data or features. Examples of such studies are Dekker *et al.*, (2009); Kabakchieva (2012); Ogunde and Ajibade (2014); and Ahmad *et al.*, (2015).

5.3.2 Higher Educational Institution (HEI) academic features

These are features that are obtained from the academic activities of students while in the higher educational institution. They are major set of features employed in SAP and dropout studies as both the predictors and the predicted variables. They include final grade, Grade Point Average (GPA), Cumulative GPA (CGPA), Matric No, course marks, course grades, class test marks, assignment scores, class attendance, laboratory scores, theory scores, seminar marks and so on. Every SAP and dropout studies must use some of these features or variables most importantly attributes like the GPA, the CGPA and the final grade as the class variable. Some studies like Osofisan *et al.*, (2014); Al-barrak and Al-razgan (2016) and Rovira *et al.*, (2017) had exclusively limited themselves to the use of these variables for the prediction of students' academic performance.

5.4 Social features

Social features are features that describe the relationship of the students under study with others in their environment, most especially as it affects their studies in the university. The social features used in the existing literature include the number of friends, Average hours used with friends, visited resources, sporting and extra-curricular activities, technology impact, adaptation and so on. Studies that had employed these features include Ruby and David (2015); Amrieh *et al.*, (2016); Saa (2016) and (Adejo & Connolly, 2018), and Adejo and Connolly (2018).

5.5 Economic features

The economic features are those features that indicate the economic status and support available to students in a case study of any SAP and dropout research. Economic features can be classified into two: parent-related economic features and the student-related economic features.

5.5.1 Parent-related Economic features

These are economic features that are linked to the students from their parents. These include family income or family support, parent's economic status which depends on the mother's occupation and father's occupation as separate attributes, parent's education which is broken into the mother's qualification and father's qualification. Studies that had employed these features include Pal (2012); Aziz et al., (2013); Saa (2016) and Adejo and Connolly (2018).

5.5.2 Student-related Economic features

These are the features that describe the nature of the economic support available to a student through other means apart from his parent. It includes sponsorship or financial aid from third parties and self-sponsorship. Studies like Gilbert (2017) and Adejo and Connolly (2018) applied these features.

5.6 Demographic features

These features are demographic data related to students in the case study. They include marital status, race, nationality, living location, date of birth, ethnicity, hometown, city and so on. Many SAP studies had employed these features in predicting the students' academic performance, among them are Aulck et al., (2016); Kaur and Singh (2016); Lakkaraju et al., (2015) and Strecht et al., (2015) among others. The list of these features is in-exhaustive and therefore the taxonomy presented here will assist researchers in this area to classify future attributes embedded in the study of the SAP and dropout predictions.

6. Discussion

The summary of SAP and dropout prediction studies in Table 1 shows that the students' academic performance (SAP) prediction studies limitations include the vastness of the attributes selected for the study, number of performance metrics used in the evaluation of the models developed and non-generalizability of models. The distribution of the selected research studies is represented in Figure 2 where all the studies are further classified based on observed methods and targets of each study. It is observed that the year 2015 featured much of the SAP studies in this review with 8 research works out of 39 for a 12 year period in different categories as being grouped into studies that use one DM technique, those that use more than one, the one majorly targeting dropout and those that examine the influence of factors on the final prediction.

With regards to the attributes considered for which the taxonomy is given in Figure 3, the Grade Point Average (GPA) has been consistently employed as the dependent variable in all the studies while the predicting variables continue to vary based on different studies priority. However, all the attributes considered in all of the studies can be conveniently linked to the two most important attributes or parameters of SAP and dropout predictions, namely: perceived academic control and academic emotion as expatiated in Respondek et al., (2017). It could also be deduced from the various studies in Table 1 that the research direction is towards standardization of methods for the prediction of SAP and dropout which is quite lacking in the area. Standardization of DM methods or techniques for the prediction of SAP and dropout has not been fully explored. Most of the predictions has focused on the prediction of general success and failure of students, no literature has focused on the prediction of excellent students, which if its influencing factors are known, can be used to infer solution the causes of students' failure that can lead to academic dropout.

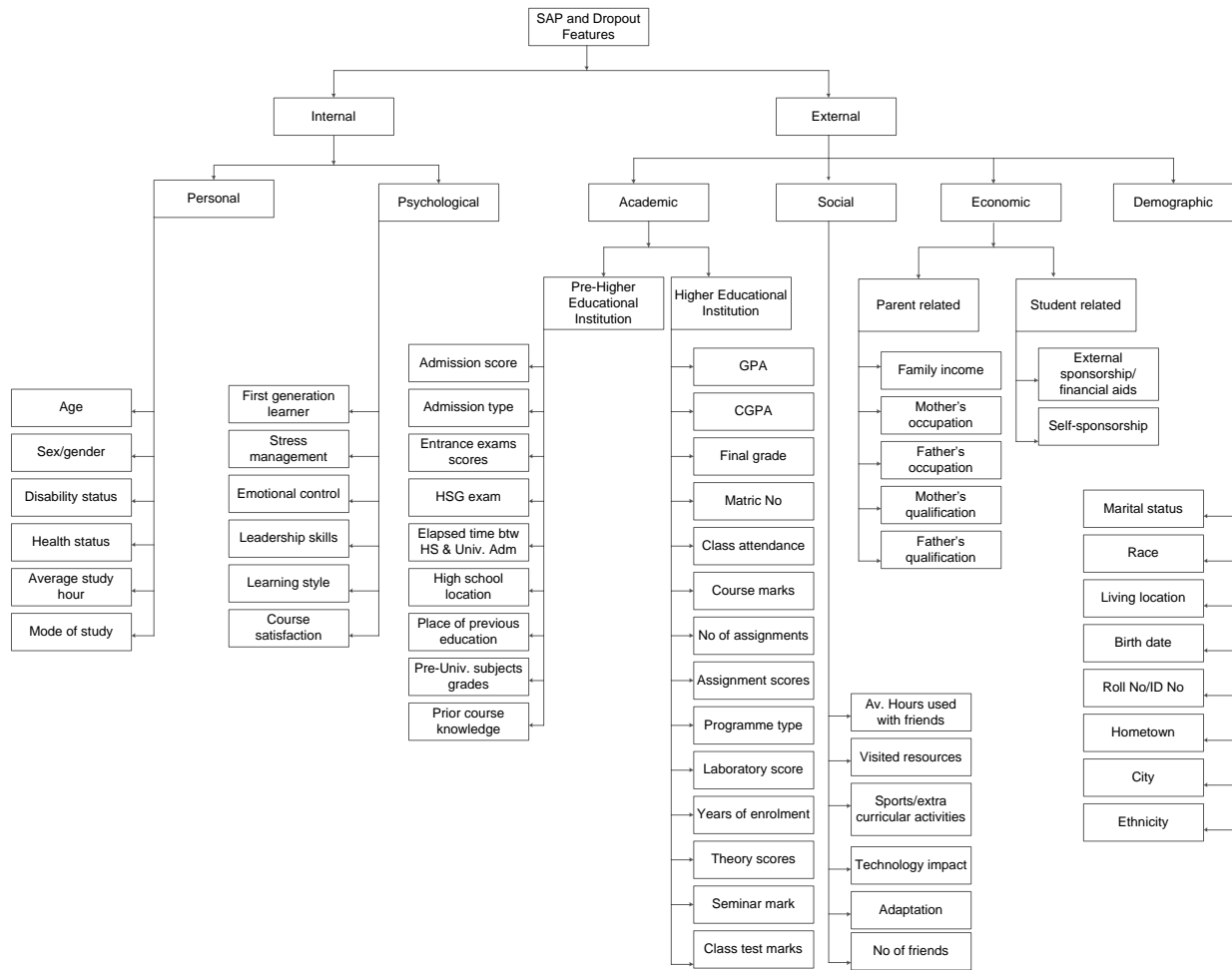


Figure 3. Taxonomy based on features

7. Conclusion and Future Directions

This review had x-rayed some selected studies in the SAP and dropout prediction as a popular area in Educational Data Mining. The taxonomy of data mining methods used in SAP and dropout studies were presented, the SAP datasets described. The study also presented taxonomy of the features employed from the different datasets in the studies reviewed.

It was revealed that the major concerns of the prediction of SAP and dropout studies are the nature of the attributes employed in the mining and prediction and the performance of the DM techniques used. The review shows that serious efforts have not been made in the direction of standardizing the DM techniques for SAP and dropout prediction. Another revelation from this review is that of the fact that the SAP and dropout datasets are scantily made public owing to the fact that each institution of higher learning considered SAP data as too confidential to be made public.

The review of studies on the predictions of students' academic performance and dropouts as presented in this paper can be compared to other review works in the area like the studies of Shahiri, Husain, & Rashid (2015) and that of Kumar, Singh & Handa (2017). However, while the two mentioned reviews only considered the DM techniques used in the study area and the attributes, this paper extends their consideration by x-raying the limitations of the studies selected for review and also presented the classification of the features used in SAP prediction studies.

Studies that used survival analysis approach to the predictions of SAP and dropouts like the study of Ameri *et al.*, (2016) are very scanty. The review also reviewed that using meta-heuristic or nature-inspired algorithms, deep learning and developing hybrid methods for predicting SAP and dropouts are fertile grounds for new and potential researchers.

References

- Adejo, O. W., & Connolly, T. (2018). Predicting student academic performance using multi-model heterogeneous ensemble approach. *Journal of Applied Research in Higher Education*, 10(1), 61–75. <https://doi.org/10.1108/JARHE-09-2017-0113>
- Aden, A. A., Yahye, Z. A., & Dahir, A. M. (2013). The Effect of Student's Attendance on Academic Performance: a Case Study At Simad University Mogadishu. *Academic Research International*, 4(6), 409–417.
- Agarwal, S., Pandey, G. N., & Tiwari, M. D. (2012). Data Mining in Education : Data Classification and Decision Tree Approach. *International Journal of E-Education, e-Business, e-Management and e-Learning*, 2(2), 140–145. <https://doi.org/10.1080/13604819608713449>
- Agrawal, S., Vishwakarma, S. K., & Sharma, A. K. (2017). Using Data Mining Classifier for Predicting Student's Performance in UG Level. *International Journal of Computer Applications*, 172(8), 39–44. <https://doi.org/10.5120/ijca2017915201>
- Ahmad, F., Ismail, N. H., & Aziz, A. A. (2015). The Prediction of Students ' Academic Performance Using Classification Data Mining Techniques. *Applied Mathematical Sciences*, 9(129), 6415–6426.
- Akinola, O. S., Akinkunmi, B. O., & Alo, T. S. (2012). A Data Mining Model for Predicting Computer Programming Proficiency of Computer Science Undergraduate Students. *Afr J Comp & ICT*, 5(1), 43–52.
- Al-barrak, M. A., & Al-razgan, M. (2016). Predicting Students Final GPA Using Decision Trees : A Case Study. *International Journal of Information and Education Technology*, 6(7), 528–533. <https://doi.org/10.7763/IJiet.2016.V6.745>
- Ali, S., Haider, Z., Munir, F., Khan, H., & Ahmed, A. (2013). Factors Contributing to the Students Academic Performance: A Case Study of Islamia University Sub-Campus. *American Journal of Educational Research*, 1(8), 283–289. <https://doi.org/10.12691/education-1-8-3>
- Ameri, S., Fard, M. J., Chinnam, R. B., & Reddy, C. K. (2016). Survival Analysis based Framework for Early Prediction of Student Dropouts. *International Conference on Computational Intelligence and Knowledge Management (CIKM)*, 903–912. <https://doi.org/http://dx.doi.org/10.1145/2983323.2983351>
- Amra, I. A. A., & Maghari, A. Y. A. (2017). Students Performance Prediction Using KNN and Naïve Bayesian. *2017 8th International Conference on Information Technology (ICIT) Students*, 909–913. IEEE.
- Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016). Mining Educational Data to Predict Student's academic Performance using Ensemble Methods. *International Journal of Database Theory and Application*, 9(8), 119–136. <https://doi.org/10.14257/ijdta.2016.9.8.13>
- Andrew, Y. N. and Michael, I. J. (2001). On Discriminative Vs. Generative Classifiers: A Comparison of Logistic Regression and Naïve Bayes. <https://ai.stanford.edu/~ang/Papers/Nips01-Discriminativegenerative.Pdf>. Retrieved from <https://ai.stanford.edu/~ang/papers/nips01-discriminativegenerative.pdf>.
- Asif, R., Hina, S., & Haque, S. I. (2017). Predicting Student Academic Performance using Data Mining Methods. *IJCSNS International Journal of Computer Science and Network Security*, 17(5), 187–191.
- Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers and Education*, 113, 177–194. <https://doi.org/10.1016/j.compedu.2017.05.007>
- Asogbon, M. G., Samuel, O. W., Omisore, M. O., & Ojokoh, B. A. (2016). A Multi-class Support Vector Machine Approach for Students Academic Performance Prediction. *International Journal of Multidisciplinary and Current Research Research*, 4(March), 210–215.
- Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2016). Predicting Student Dropout in Higher Education. *International Conference on Machine Learning (ICML)*, 16–20. <https://doi.org/10.1002/prot.24187>
- Aziz, A. A., Ismail, N. H., & Ahmad, F. (2013). Mining Students' Academic Performance. *Journal of Theoretical and Applied Information Technology*, 53(3), 485–495.
- Aziz, A. A., Ismail, N. H., Ahmad, F., & Hassan, H. (2015). A Framework for Students' Academic

- Performance Analysis Using Naïve Bayes Classifier. *Jurnal Teknologi*, 3, 13–19.
- Bhagvatula, S., Shreyas, K. J., Gupta, S., & Singh, M. (2015). Contribution of Students' Academic Efforts on Prediction of Students' Performance Using WEKA. *International Journal of Current Engineering and Scientific Research (IJCESR)*, 2(11), 83–89. Retrieved from troindia.in/journal/ijcesr/vol2iss11/83-89.pdf
- Burman, I., & Som, S. (2019). Predicting Students Academic Performance Using Support Vector Machine. *Proceedings - 2019 Amity International Conference on Artificial Intelligence, AICAI 2019*, 756–759. <https://doi.org/10.1109/AICAI.2019.8701260>
- Daud, A., Aljohani, N. R., Abbasi, R. A., Lytras, M. D., Abbas, F., & Alowibdi, J. S. (2017). Predicting Student Performance using Advanced Learning Analytics. *International World Wide Web Conference Committee (IW3C2)*, 415–421. Retrieved from <https://dl.acm.org/citation.cfm?id=3054164>
- Dekker, G. W. ., Pechenizkiy, M. ., & Vleeshouwers, J. M. . (2009). Predicting Students Drop Out: A Case Study. *International Conference On Educational Data Mining*, 41–50. <https://doi.org/doi:10.1037/0893-3200.21.3.344>
- Do, Q. H., & Chen, J. (2013). A Neuro-Fuzzy Approach in the Classification of Students ' Academic Performance. *Computational Intelligence and Neuroscience*, 2013, 1–8. <https://doi.org/http://dx.doi.org/10.1155/2013/179097> Research
- Eashwar, K. B., Venkatesan, R., & Ganesh, D. (2017). Student Performance Prediction Using Svm. *International Journal of Mechanical Engineering and Technology (IJMET)*, 8(11), 649–662.
- Enu, J., Agyman, O. K., & Nkum, D. (2015). Factors Influencing Students' Mathematics Performance in Some Selected Colleges of Education in Ghana. *International Journal of Education Learning and Development*, 3(3), 68–74. <https://doi.org/10.1007/978-1-4471-5520-1>
- Fareo, D. O. (2013). Truancy and Academic Performance of Secondary School Students in Southwestern Nigeria: Implications for Counselling. *International Journal for Cross-Disciplinary Subjects in Education (IJCDSE), Special Issue*, 3(2), 1424–1428.
- Gilbert, N. (2017). Predicting Success: An Application of Data Mining Techniques to Student Outcomes. *International Journal of Data Mining & Knowledge Management Process*, 7(2), 01–20. <https://doi.org/10.5121/ijdkp.2017.7201>
- Gorunescu, F. (2011). *Data Mining Concepts, Models and Techniques* (J. Kacprzyk & L. C.Jain, eds.). <https://doi.org/10.1007/978-3-642-19721-5> Intelligent
- Hamsa, H., Indiradevi, S., & Kizhakkethottam, J. J. (2016). Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm. *Procedia Technology*, 25, 326–332. <https://doi.org/10.1016/j.protcy.2016.08.114>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques* (Third). USA: Morgan Kaufmann Publishers.
- Hijazi, S. T., & Raza Naqvi, S. M. M. (2006). Factors Affecting Students' Performance. *Bangladesh E-Journal of Sociology*, 3(1), 1–10.
- Joseph, S., & Devadas, L. (2015). Student ' s Performance Prediction Using Weighted Modified ID3 Algorithm. *International Journal of Scientific Research Engineering & Technology (IJSRET)*, 4(5), 571–576.
- Kabakchieva, D. (2012). Student Performance Prediction by Using Data Mining Classification Algorithms. *International Journal of Computer Science and Management Research*, 1(4), 686–690.
- Kantardizc, M. (2011). *Concepts , Models , Methods , and Algorithms* (second; R. Abhari, J. A. G. W. A. F. Canavero, M. E.- Hawary, B.-M. H. M. L. D. Jacobson, & O. P. M. S. N. T. S. G. Zobrist, eds.). New Jersey: John Wiley & Sons Inc.
- Kaur, G., & Singh, W. (2016). Prediction Of Student Performance Using Weka Tool. *International Journal of Engineering Sciences*, 17(January), 2229–6913.
- Kaur, P., Singh, M., & Josan, G. S. (2015). Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector. *Procedia Computer Science*, 57, 500–508. <https://doi.org/10.1016/j.procs.2015.07.372>
- Kim, B., Vizitei, E., & Ganapathi, V. (2018). GritNet : Student Performance Prediction with Deep Learning. *International Conference on Educational Data Mining*. Retrieved from educationaldatamining.org/files/conferences/EDM2018/.../EDM2018_paper_257.pdf

- Kumar, M., Singh, A. J., & Handa, D. (2017). Literature Survey on Educational Dropout Prediction. *International Journal of Education and Management Engineering*, 2, 8–19. <https://doi.org/10.5815/ijeme.2017.02.02>
- LAK. (2011). Calls for Papers. *1st International Conference on Learning Analytics and Knowledge*. Banff, Alberta.
- Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., & Addison, K. L. (2015). A Machine Learning Framework to Identify Students at Risk of Adverse Academic Outcomes. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15*, 1909–1918. <https://doi.org/10.1145/2783258.2788620>
- Makhtar, M., Nawang, H., Nor, S., & Shamsuddin, W. A. N. (2017). ANALYSIS ON STUDENTS PERFORMANCE USING NAÏVE. *Journal of Theoretical and Applied Information Technology*, 95(16), 3993–4000.
- Mishra, T., Kumar, D., & Gupta, S. (2014). Mining students' data for prediction performance. *International Conference on Advanced Computing and Communication Technologies, ACCT*, 255–262. <https://doi.org/10.1109/ACCT.2014.105>
- Mohsin, F., Norwawi, M. N., M., Hidadullah, C. F., & Wahab, M. H. A. (2010). Mining the student programming performance using rough set. *Proceedings of 2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering, ISKE 2010*, 478–483. <https://doi.org/10.1109/ISKE.2010.5680824>
- Mueen, A., Zafar, B., & Manzoor, U. (2016). Modeling and Predicting Students' Academic Performance Using Data Mining Techniques. *International Journal of Modern Education and Computer Science*, 8(11), 36–42. <https://doi.org/10.5815/ijmecs.2016.11.05>
- Mushtaq, I., & Khan, S. N. (2012). Factors Affecting Students' Academic Performance. *Global Journal of Management and Business Research*, 12(9), 16–22.
- Nayebzadeh, S., Dehnavi, H. D., Nejad, M. S., Mir, M., & Sadrabadi, M. (2013). Factors Affecting the Academic Improvement of Accounting Students in Islamic Azad University of Yazd. *International Journal of Academic Research in Economics and Management Sciences*, 2(4), 231–243. <https://doi.org/10.6007/IJAREMS/v2-i4/122>
- Nivre, J. (2003). Discriminative Models. *Stp.Lingfil.Uu.Se/~nivre/Master/Discriminative.Pdf*, 1–5. Retrieved from stp.lingfil.uu.se/~nivre/master/discriminative.pdf
- Ogor, E. N. (2007). Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques. *Fourth Congress of Electronics, Robotics and Automotive Mechanics*, 354–359. <https://doi.org/10.1109/CERMA.2007.78>
- Ogunde, A. O., & Ajibade, D. A. (2014). A Data Mining System for Predicting University Students' Graduation Grades Using ID3 Decision Tree Algorithm. *Computer Science and Information Technology*, 2(1), 21–46.
- Oladipupo, O., Daramola, O., Oyelade, J., & Afolabi, I. (2017). Assessing the Impact of Class Attendance on Student's Academic Performance Using Data Mining. *Asian Journal of Information Technology*, 16(2–5), 218–226.
- Oloruntoba, S. A., & Akinode, J. L. (2017). Student Academic Performance Prediction Using Support Vector Machine. *International Journal Of Engineering Sciences & Research Technology Student*, 6(12), 588–598.
- Osofisan, A., Adeyemo, O., & Oluwasusi, S. (2014). Empirical Study of Decision Tree and Artificial Neural Network Algorithm for Mining Educational Database. *African Journal of Computing & ICT* ©, 7(2), 187–196. Retrieved from http://www.ajocict.net/uploads/V7N2P22-2014_AJOCICT_-_Paper_22.pdf
- Pal, S. (2012). Mining Educational Data Using Classification to Decrease Dropout Rate of Students. *International Journal of Multidisciplinary Sciences and Engineering*, 3(5), 35–39. Retrieved from www.ijmse.org
- Pandey, U. K., & Pal, P. S. (2011). Data Mining : A prediction of performer or underperformer using classification. *(IJCSIT) International Journal of Computer Science and Information Technology*, Vol. 2(2), 2011, 686-690, ISSN:0975-9646.
- Rajeshinigo, D., & Jebamalar, J. P. A. (2017). Educational Mining: A Comparative Study of Classification Algorithms Using WEKA. *International Journal of Innovative Research in*

- Computer and Communication Engineering*, 5(3), 5583–5589. <https://doi.org/10.15680/IJIRCCE.2017>.
- Razaque, F., Soomro, N., Shaikh, S. A., Soomro, S., Samo, J. A., Kumar, N., & Dharejo, H. (2018). Using Naïve Bayes Algorithm to Students' Bachelor Academic Performances Analysis. *4th IEEE International Conference on Engineering Technologies and Applied Sciences, ICETAS 2017, 2018-Janua*(November), 1–5. <https://doi.org/10.1109/ICETAS.2017.8277884>
- Respondek, L., Seufert, T., Stupnisky, R., & Nett, U. E. (2017). Perceived academic control and academic emotions predict undergraduate university student success: Examining effects on dropout intention and achievement. *Frontiers in Psychology*, 8(MAR), 1–18. <https://doi.org/10.3389/fpsyg.2017.00243>
- Romero, C., Ventura, S., Espejo, P. G., & Hervás, C. (2008). Data Mining Algorithms to Classify Students. *1st International Conference on Educational Data Mining*, 8–17. Retrieved from <https://pdfs.semanticscholar.org/c6e2/9d4a3d70cd09c1397e736a38afd25def7bd3.pdf>
- Rovira, S., Puertas, E., & Igual, L. (2017). Data-driven System to Predict Academic Grades and Dropout. *PLoS ONE*, 12(2), 1–21. <https://doi.org/10.1371/journal.pone.0171207>
- Ruby, J., & David, K. (2015). Analysis of Influencing Factors in Predicting Students Performance Using MLP -A Comparative Study. *International Journal of Innovative Research in Computer and Communication Engineering*, 3297(2), 1085–1092. <https://doi.org/10.15680/ijircce.2015.0302070>
- Saa, A. A. (2016). Educational Data Mining and Students' Performance Prediction. *International Journal of Advanced Computer Science and Applications*, 7(5), 212–220. <https://doi.org/10.14569/IJACSA.2016.070531>
- Seifert, K., & Sutton, R. (2009). *Educational Psychology Second Edition* (Second; J. S. Kelvin Seifert, Marisa Drexel, ed.). [https://doi.org/10.1016/S0738-3991\(02\)00243-4](https://doi.org/10.1016/S0738-3991(02)00243-4)
- Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A Review on Predicting Student's Performance Using Data Mining Techniques. *Procedia Computer Science*, 72, 414–422. <https://doi.org/10.1016/j.procs.2015.12.157>
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. <https://doi.org/10.1017/CBO9781107298019>
- Simeunović, V., & Preradović, L. (2014). Using Data Mining to Predict Success in Studying. *Croatian Journal of Education*, 16(December 2012), 491–523.
- Solomon, D., Patil, P. seem., & Agrawal, P. proov. (2018). Predicting Performance and Potential Difficulties of University Student using Classification : Survey Paper. *International Journal of Pure and Applied Mathematics*, 118(18), 2703–2707.
- Strecht, P., Cruz, L., Soares, C., Mendes-Moreira, J., & Abreu, R. (2015). A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance. *Proceedings of the 8th International Conference on Educational Data Mining*, 392–395. <https://doi.org/10.1038/417671a>
- Tekin, A. (2014). Early Prediction of Students' Grade Point Averages at Graduation: A Data Mining Approach. *Eurasian Journal of Educational Research*, 14(54), 207–226. <https://doi.org/10.14689/ejer.2014.54.12>
- Tinto, V. (2006). Research and Practice of Student Retention: What Next? *Journal of College Student Retention: Research, Theory and Practice*, 8(1), 1–19. <https://doi.org/10.2190/C0C4-EFT9-EG7W-PWP4>
- Verma, K., Singh, A., & Verma, P. (2016). A review on predicting student performance using data mining method. *International Journal of Current Engineering and Scientific Research (IJCESR)*, 3(1), 127–132.
- Witten, I. H., & Frank, E. (2005). *Data Mining Practical Machine Learning Tools and Techniques* (Second). <https://doi.org/0120884070,9780120884070>
- Yadav, S. K., Bharadwaj, B., & Pal, S. (2011). Data Mining Applications: A comparative Study for Predicting Student's performance. *International Journal of Innovative Technology & Creative Engineering*, 1(12), 13–19.
- Yadav, S. K., & Pal, S. (2012). Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification. *World of Computer Science and Information Technology Journal (WCSIT)*, 2(2), 51–56.

- Yang, S. J. H., Lu, O. H., Huang, A. Y. Q., Huang, J. C. H., Ogata, H., & Lin, A. J. Q. (2018). Predicting Students ' Academic Performance Using Multiple Linear Regression and Principal. *Journal of Information Processing*, 26, 170–176. <https://doi.org/10.2197/ipsjjip.26.170>
- Yildiz, O.; Bal., A.; Gulsecen, S. & Kentil, F. D. (2012). A Genetic-Fuzzy BAsed Mathematical Model to Evaluate the Distance Education Students' Academic Performance. *Procedia, Social and Behavioural Sciences. International Conference on New Horizons in Education. (INTE2012)*. Elsevier Ltd.