Journal of Information and Technology

Using Machine Learning Algorithms to Predict Students' Performance and Improve Learning Outcome: A Literature Based Review

Francis Ofori, Dr. Elizaphan Maina and Dr. Rhoda Gitonga

ISSN: 2617-3573

Stratford Peer Reviewed Journals and Book Publishing
Journal of Information and Technology
Volume 4||Issue 1||Page 23-45||March||2020|
Email: info@stratfordjournals.org ISSN: 2616-3573

Using Machine Learning Algorithms to Predict Students' Performance and Improve Learning Outcome: A Literature Based Review

[1*]Francis Ofori, [2]Dr. Elizaphan Maina & [3]Dr. Rhoda Gitonga

[1]PhD Candidate, Kenyatta University

[2]Lecturer, Kenyatta University

[3]Lecturer, Kenyatta University

***Email of the Corresponding Author: apoe12@yahoo.com**

***How to cite this article*: Ofori, F., Maina, E. & Gitonga, R. (2020). Using Machine Learning Algorithms to Predict Students' Performance and Improve Learning Outcome: A Literature Based Review, *Journal of Information and Technology,* Vol. 4(1), 23-45.**

Abstract

The application of machine learning techniques in predicting students' performance, based on their background and their in-term performance has proved to be a helpful tool for foreseeing poor and good performances in various levels of education. Early prediction of students' performance is useful in taking early action of improving learning outcome. The prediction of the student's academic performance is important as it helps increase graduation rates by appropriately guiding students, guiding changes in university academic policies, informing instructional practices, examining efficiency and effectiveness of learning, providing meaningful feedback for teachers and learners and modifying learning environments. A high prediction accuracy of the students' performance is helpful to identify the low performance students at the beginning of the learning process. However, to achieve these objectives, large volume of student data must be analyzed and predicted using various machine learning models. Moreover, it is not clear which model is best in predicting performance and which machine learning model is appropriate in improving learning in among students. The paper through intensive literature review attempts to identify best machine learning model in predicting student performance and appropriate machine learning model in improving learning. The empirical review indicated contentious results on machine learning model that best predicts students' performance. Moreover, it is not clear among the various machine learning algorithms which one derives the best approach in predicting students' performance while improving learning outcome. The varying prediction level by various machine learning models may be as a result of differences in socioeconomic. It may also be important to note that student's academic performances are affected by many factors, like socioeconomic factors of students like family income, parental

Stratford Peer Reviewed Journals and Book Publishing
Journal of Information and Technology
Volume 4||Issue 1||Page 23-45||March||2020|
Email: info@stratfordjournals.org ISSN: 2616-3573

level of education and employment status of students or parents but are not considered when testing the accuracy of various machine learning models in predicting students' performance. Moreover, the various machine learning models did not identify the most appropriate machine learning model in improving students' outcome. Most models focused largely in predicting students' performance without considering mechanisms to improve learning outcome of students. As a result, it is important to test the accuracy of various machine learning models that best predicts students' performance and the one that is most appropriate in improve learning outcome while considering socio economic and demographic factors of the students. The study makes a conclusion that predicting students' performance is of the highest priority for any learning institution across the globe. Using various machine learning methods to accurately predict student's performance would be highly required. It is important to accurately rank machine models based on their prediction capabilities in predicting students' performance and in improving learning outcome.

Key words: *Machine learning algorithms, students' performance, learning outcome*

1. Introduction

Early prediction of students' performance is useful in taking early action of improving learning outcome. Predicting a student's performance from past academic data is one of the most popular applications of educational data mining and, therefore, it is a valuable source of information that can be used to improve students' performance (Buenaño-Fernández, Gil & Luján-Mora, 2019). Prediction of the student's success helps education institutions improve learning and teaching methodologies by identifying instructional methods that suit students from varied background information (Belachew & Gobena, 2017). A high prediction accuracy of the students' performance is helpful to identify the low performance students at the beginning of the learning process.

Excellent student performance is one of indicators of course learning outcome in a learning institution. Educational institutions required the result of the predicate process to improve student's academic achievement and also to improve learning (Keshtkar, 2018). Examinations, assignments and course projects are commonly used as course assessments to evaluate the students understanding and knowledge gain. Analyzing students' performances is challenging task because of the voluminous educational data that must be considered (Pojon, 2017). Therefore, the need of having effective tools to process these student data has risen.

The ability to timely predict the academic performance of students is very important in learning institutions (Koutina & Kermanidis, 2011). The student's performance prediction is an important area as it can help teachers identify students that need additional academic assistance (Hussain, Zhu, Zhang, Abidi & Ali, 2019). Prediction of student academic performance helps instructors develop a good understanding of how well or how poorly the students in their classes will perform, so instructors can take proactive measures to improve student learning. Accurately predicting students' future performance based on their ongoing academic records is crucial for effectively carrying out necessary pedagogical interventions to ensure students' on-time and satisfactory course completion (Belachew & Gobena, 2017). However, to achieve these

objectives, large volume of student data must be analyzed and predicted using various machine learning models. Moreover, it is not clear which model is best in predicting performance and which model is best in improving learning outcome among students. The paper through intensive literature review attempts to identify best model in predicting student performance and best model in improving learning.

The empirical review indicated contentious results on model that best predicts students' performance. Moreover, it is not clear among the various machine learning algorithms which one derives the best approach in predicting students' performance while improving learning outcome. Hussain, Muhsin, Salal, Theodorou, Kurtoğlu and Hazarika (2019) indicated that Artificial Neural Network produces the highest classification accuracy in predicting students' performance at 95.34%. However, according to Belachew and Gobena (2017), Naive Bayesian compared to Support Vector Machine, multi-layer perceptron network and neural networks was more accurate in predicting students' performance in the department of Information Technology at 95.7% prediction accuracy. Jayaprakash, Balamurugan and Chandar (2018) further indicated that Naive Bayes was best in predicting students' academic performance with accuracy prediction of 92.2%. Obsie and Adem (2018) noted that Linear Regression and Support Vector Regression performed better than Neural Network in predicting student academic performance where prediction accuracy for neural network was 97.63%, Support Vector Regression was 98.05% and Linear Regression 98.05%. In terms of time to predict results, Support Vector Regression was best at 0.03 seconds, followed by 0.05 seconds for Linear Regression and 0.78 seconds for Support Vector Regression.

Moreover, a substantial amount of literature focuses on predicting student performance in solving problems or completing courses (Yadav & Pal, 2012). Many machine learning techniques, such as decision trees, artificial neural networks, matrix factorization, collaborative filters and probabilistic graphical models have been applied to develop prediction algorithms (Lakkaraju, *et al.*, 2015). However, it is not clear which among the various machine learning models accurately predicts students' performance as various authors have presented conflicting results regarding the prediction accuracy of the models. Moreover, the studies conducted by other authors have not identified a machine learning model that can be employed to improve learning outcome of students.

It is evident that different machine learning reveals different prediction accuracy of students' performance. No clear model has been proved to be the best in accurately predicting students' performance. The varying prediction level by various machine learning models may be as a result of differences in socioeconomic factors of students like family income, parental level of education and employment status of students or parents. It may also be important to consider student's academic performance affected by many factors, like personal, socio-economic and other environmental variable when testing the accuracy of various machine learning models in predicting student performance. Moreover, the various machine learning models did not identify the best model in improving student's outcome. It is not clear which model is best in predicting performance and at the same time best in improving learning in among students. Most models

focused largely in predicting students' performance without considering mechanisms to improve learning experience by students. Students' performances is influenced by socioeconomic factors of students like family income, parental level of education and employment status of students or parents but are not considered when testing the accuracy of various machine learning models in predicting students' performance. As a result, it is important to test the accuracy of various machine learning models that accurately and precisely predicts students' performance and the one that is most appropriate in improve learning outcome while considering effects of socio economic and demographic factors of students on academic performance.

2. Machine Learning Models

The term machine learning is often referred to an analytic process designed that discovers data patterns and relationships between data variables. Moreover, a key feature of machine learning is the capacity to analyze complex non- linear relationships, given that complex input variables are expected (Yadav & Pal, 2012). Many machine learning models can be adapted to analyze the data such as classification, clustering and association rules mining depending on the suitability data collection and objectives of data analytical process. According to Hussain, Muhsin, Salal, Theodorou, Kurtoğlu and Hazarika (2019), machine learning is useful in monitoring and analyzing learning process in schools, predicting learners' performance by offering required academic assistance, academic guiding and advice mentoring, examining efficiency and effectiveness of learning methods, providing meaningful feedback for teachers and learners and modifying learning environments to the benefit of students.

The application of machine learning techniques to predicting students' performance, based on their background information and their in-term performance has proved to be a helpful tool for foreseeing poor and good performances in various levels of education (Soni, Kumar, Kaur & Hemavath, 2018). Machine learning offers an advantage over traditional forms of statistical analysis, placing emphasis on predictive performance over provable theoretical properties and priori super-population assumptions. Thereby tutors are enabled to timely help the weakest ones, but also, to promote the strongest thus improving learning. Machine learning is used to attain this objective. Machine learning techniques are used to discover models or patterns of data, and it is helpful in the decision-making (Hussain, Zhu, Zhang & Abidi, 2018). The ability to predict performance of students is very crucial in our present education system. However, it is not evident which machine learning model is best in predicting student performance and which one is best in improving learning outcome (Kumar, Singh & Handa, 2017). There are several data mining methods that are used to obtain hidden knowledge from vast amount of data. Some of the machine learning models includes decision trees, neural network, bayesian classifier-nearest neighbor, support vector machine, random forest, logistic regression, linear discriminant analysis, multiple regression and self-organised map.

2.1 Decision Tree

Decision Tree a flow-chart-like tree structure is one of a popular technique for prediction and decision-making. Decision Tree has been used extensively by most of researchers because of its simplicity and comprehensibility to uncover small or large data structure and predict the value. Decision Tree classifiers are used in data mining to produce trees after studying the training set and will be used to create predictions (Swamy & Hanumanthappa, 2012). Decision tree classifiers are one of the admired and influential tools for classification. Normally, decision tree classifiers have a tree-like structure which starts from root attributes, and ends with leaf nodes. It also has several branches consisting of dissimilar attributes, the leaf node on each branch representing a class or a kind of class distribution (Pandey & Sharma, 2013). Decision tree algorithms explain the relationship with attributes, and the comparative significance of attributes. The benefit of decision trees are that they characterize rules which could simply be understood and interpreted by users, do not need complex data preparation, and do well for numerical and categorical variables (Olaniyi, Kayode, Abiola, Tosin & Babatunde, 2017). The core algorithm for constructing decision trees is called ID3.

2.2 Artificial Neural Network

Artificial Neural Network (ANN) is arguably one of the popular techniques used in educational data mining. Neural network receives signals through synapses in the dendrites. As per the ANN approach, when the received signals are strong enough (over threshold), the neuron is activated and the signal on the axon is generated. This signal can be sent to other synapses and may activate other neurons (Sekeroglu, Dimililer & Tuncal, 2019). Artificial Neural Network is generally composed from input (synapses) that are multiplied by weight (the strength of each signal) then calculated by a mathematical function that determines the activation of the neuron to produce an output. The advantage of neural network is that it has the potential to detect all possible interactions between predictor variables (Bassi, Dada, Hamidu & Dauda, 2019). Neural network could also do a complete detection without having any doubt even in complex nonlinear relationship between dependent and independent variables. Therefore, neural network technique is selected as one of the best prediction method.

2.3 Logistic regression

Logistic regression as another method in machine learning that is used to determine the effect of one or more independent variables(x) to one dependent variable(y) that is a binary in nature (Soule, 2017). Criteria variables are used: -The dependent variable is a dummy variable that should have only two alternatives, such as yes or no, 1 or 0.

2.4 Naive Bayes

The Naive Bayes algorithm is a simple probabilistic classifier which is based on Bayes theorem with strong and naive independence assumptions. Naive Bayes is one of the inductive learning algorithms that is effective and efficient for machine learning and data mining. These algorithms

use the Bayesian probability of predicting the probability of the future based on past experience. It is one of the most basic classification techniques with various applications in email spam detection, personal email sorting, document categorization, sexually explicit content detection, language detection and sentiment detection. Despite the naive design and oversimplified assumptions that this technique uses, Naive Bayes performs well in many complex real-world problems. According to Jayaprakash, Balamurugan and Chandar, (2018), a Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods. Naive Bayes algorithm is highly scalable and requires a number of parameters linear in the number of variables (Shaziya, Zaheer & Kavitha, 2015). A Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. Several colleges and universities have adopted feedback analysis system using various models in data-mining to improve student retention and to channel students to courses and programs that the institutions judge most appropriate.

2.5 Support Vector Machine

Support Vector Regression is the prediction type of Support Vector Machine which assigns support vectors in order to separate features. SVMs are described as a set of related supervised learning techniques used for classification and regression (Oloruntoba & Akinode, 2017). They are member of a family of generalized linear classification. An important property of SVM is, SVM simultaneously minimize the empirical classification error and maximize the geometric margin. Thus, SVM is also known as a Maximum Margin Classifiers. SVM is based on the Structural risk Minimization (SRM). SVM map input vector to a higher dimensional space where a maximal separating hyper plane is constructed. Two parallel hyper planes are constructed on each side of the hyperplane that separate the data. The separating hyperplane is the hyperplane that maximize the distance between the two parallel hyperplanes. In classification phase, increment of classes may cause the reduction of success rate in SVM. However, it can be used effectively for 2-class problems.

2.6 Random Forest

Random forest is a collection of decision trees built up with some element of random choice (Ulinnuha, Sa'Dyah, 2017). Random forest is an ensemble learning method for classification, in which it constructs multiple of unpruned classification trees in the training phase, by bootstrap sampling method on the training data. Several interesting problem have been examined using random forest and it is evident that this technique has significant potential in providing useful classification model (Sultana, Rani & Farquad, 2019). The final predicted output for a random selected feature is given by finding the mean from all unpruned classification trees in the testing phase [Rodriguez-Galiano, Ghimire, Rogan, Chica-Olmo & Rigol-Sanchez, 2012).

2.7 K-Nearest Neighbor

The k-Nearest Neighbor algorithms (k-NN) organize objects based on the neighboring training examples in the feature space. K-NN is a kind of instance-based learning, or lazy learning, where the function is only approximated nearby and the entire calculation is delayed in anticipation of classification. The main problem of k-NN algorithm is that its accuracy can be strictly ruined by the existence of loud or inappropriate features. Likewise, its accuracy becomes unfortunate if the feature balance are not reliable with their importance

2.8 Multiple regression

Regression is a statistical method to identify the relationship between the variables present in the data. It mainly focuses on the relationship between the dependent variable and independent variables which is otherwise called as predictors (Rajalaxmi, Natesan, Krishnamoorthy, Ponni, 2019). It helps to understand the changes occur in the value of dependent variable when anyone of the independent variables is changed. By using the value of the independent variable, an equation is formulated which contains the independent variables along with some coefficients and the slope value (Oyerinde & Chia, 2017). There are lot many types of regression techniques. One such is linear regression technique which is mainly used for prediction. The linear regression is used to examine the relationship between one dependent variable and one or more independent variables.

2.9 Stochastic Gradient Descent binary classifiers

Stochastic Gradient Descent (SGD) has gained popularity for solving large scale supervised machine learning problems. It provides a rapid method for minimizing a number of loss functions and is applicable to Support Vector Machine (SVM) and Logistic optimizations. However SGD does not provide a convenient stopping criterion. Generally an optimal number of iterations over the data may be determined using held out data. SGD has proved to be a very effective method of training machine learning algorithms (Wilbur & Kim, 2014). It has generally been found to confer a significant decrease in training time without sacrificing accuracy (Lau, T., Sun & Yang, 2019). SGD can be applied to standard convex loss functions with regularization terms with good effect, but Diab (2019) suggested using SGD without the usual regularization term and performing the regularization with early stopping. This has become a widely practiced approach and is implemented by dividing the training set into disjoint pieces consisting of a new training set and a validation set. The advantage of Stochastic Gradient Descent is that it is efficient and easy to implement compared to other machine learning models like deep learning. However, the demerits of Stochastic Gradient Descent are that; requires a number of hyperparameters such as the regularization parameter and the number of iterations and is sensitive to feature scaling (Allen-Zhu, 2017).

3. Empirical Literature

Buenaño-Fernández, Gil and Luján-Mora (2019) conducted a study on application of machine learning in predicting performance for computer engineering students. The results showed the effectiveness of machine learning techniques to predict the performance of students. Nevertheless, several studies related to the prediction of academic performance. The study focused in predicting students without suggesting how to improve model. Further, little or nothing has been done to show how machine learning has been used to improve learning outcome presenting conceptual gap. Moreover, the study only suggested that decision trees methodology is the best model to predict students' performance, whereas empirical literature presents conflicting results on which machine model is most accurate in predicting students' performance.

Hussain, Muhsin, Salal, Theodorou, Kurtoğlu and Hazarika (2019) undertook a study on prediction model on student performance based on internal assessment using Artificial Neural Network. The highest classification accuracy achieved in this study was 95.34% produced by Artificial Neural Network. The Precision, Recall, F-Score, Accuracy, and Kappa Statistics Performance were calculated as a statistical decision to find the best classification methods. Artificial Neural Network has been largely employed in predicting students' performance. The use of Artificial Neural Network in improving learning outcome has not been explored presenting a conceptual gap. Conflicting empirical results on which machine model is most accurate in predicting students' performance is evident and comparative study using various machine models is required.

Belachew and Gobena (2017) investigated student performance prediction model using machine learning approach, the case of Wolkite University. The dataset used in their study was taken from the Wolkite university registries office for college of computing and informatics from 2004 up to 2007 E.C with respect to each department. The study collected student's transcript data that included their final GPA and their grades in all courses. After pre-processing the data, they applied the machine learning methods, neural networks, Naive Bayesian and Support Vector Machine (SMO). Under the department of Information Technology had higher performance using Naïve Bayesian method (95.7%), implying that Naive Bayesian had higher performance when compared to Support Vector Machine and multi-layer perceptron network. Despite the neural networks, Naive Bayesian and Support Vector Machine being useful in predicting students' performance, the models have not been employed to improve learning outcome of students presenting both conceptual and methodological gaps.

Jayaprakash, Balamurugan and Chandar (2018) conducted a study to predict students' academic performance using Naive Bayes Algorithm. The data herein was collected by means of feedback rating-scale questionnaire. The questions in the questionnaire were measured with a scale value of 1 to 5. Then, the data was collected from 700 students in various departments of Blue-Crest College, Accra, Ghana in the academic year 2014 with the internal examination score. The

results show Naive Bayes Algorithm predicts students' performance within 2 seconds with 92.2% prediction accuracy. The comparison between feedback and internal examination marks Navie Bayes algorithm gave the better prediction result and it was measured using confusion matrix with accuracy prediction of 92.2%. The results were predicted within 2 seconds. This simple analysis works showed that the proper data mining application on student's performance data can be efficiently used for vital hidden knowledge/information retrieval from the vast data, which can be used for the process of decision making by the management of an educational institution. It helps the institutions to identify the weaker students in advance and they can arrange special measures to get good score. This paper also concludes with that for data mining application, effective and faster results prediction, classification and clustering, the institutions can improve their quality based on the analysis to conduct the special training to their students. The study identified Navie Bayes algorithm as the most accurate machine learning model in predicting students' performance. The best model in improving learning outcome was not tested and identified thus highlighting the need to test the ability of other machine learning models in improving students' outcome. Contentious results on model that accurately best predicts students' performance and improves learning outcome is evident among various empirical studies highlighted in this paper.

Obsie and Adem (2018) conducted a study on prediction of student academic performance using Neural Network (NN), Linear Regression (LR) and Support Vector Regression (SVR). The dataset used in this study was collected from Student Information System (SIS) of Hawassa University for the School of Computer Science. The dataset comprised 134 undergraduate degree students graduated from the university in the year 2015, 2016 and 2017 which consisted of 52(38.81%), 39(29.10%) and 43(32.09%) students respectively that managed to reach the final at semester eight. The collected data was organized in Microsoft Excel sheet. Results revealed that time prediction for neural network was 0.9763, SVR was 0.9805 and Linear Regression 0.9805. Accurate neural network was 0.78 seconds; SVR was 0.03 seconds and Linear Regression 0.05 seconds. Prediction overall, the least accurate prediction result for all scenarios was obtained by the NN method. The study verified that data mining techniques can be used in predicting students' academic performance in higher educational institutions. All the experiments gave valid results and can be used to predict graduation CGPA. Comparisons of the experiments were done to determine which approaches perform better than others. Generally, SVR and LR methods performed better than NN. The study recommended the adoption of SVR and LR methods to predict final CGPA8, and the models can also be used to implement Student Performance Prediction System (SPPS) in a university. Thus, the study has used the models from SVR and LR methods for designing an application to do the prediction task. The various machine learning models have not been employed to improve learning outcome presenting conceptual gap. Moreover, conflicting results on model that accurately best predicts students' performance is evident as highlighted by empirical results in this paper.

Oyerinde and Chia (2017) conducted a study in predicting students' academic performances-a learning analytics approach using multiple linear regression. Linear Regression was used with the aid of the Statistical Package for Social Sciences (SPSS) analysis tool. Statistical Hypothesis testing was then used to validate the model with a 5% level of significance. With these results, it

can be concluded that the mathematics courses have strong predictive powers, as these variables accounts for about 89.0% of the variation in CS201 performance. Contentious results on model that best predicts students' performance and improves learning outcome is evident presenting a conceptual gap. Moreover, the accuracy of multiple linear regression in predicting students' performance is limited on the size of the sample thus a study with small sample size yields inadequate and inconclusive results an evident of methodological gap.

Rajalaxmi, Natesan, Krishnamoorthy and Ponni (2019) undertook a study on regression model for predicting engineering students' academic performance. Based on the proposed study, the questionnaire was prepared to gather information from the students. The input data on student performance in academics was collected from students of 150 undergraduate engineering disciplines. Multiple measures were used to calculate and corroborate the models that were predicted along with the percentage of good predictions. The results showed that the regression model gives the better accuracy in prediction. From the above, the educators can analyze the performance of the class and can also improvise the teaching techniques used based on the result of each category of the engineering students. However, the study failed to show whether regression model can be used to improve learning outcome. Moreover, multiple regression model is not accurate when the sample size is small presenting methodological gap.

Oloruntoba and Akinode (2017) conducted a study on student academic performance prediction using support vector machine. Data Sample of students in one of the Federal Polytechnic in south West part of Nigeria was used. The academic performance was defined using student's Grade Point Average (GPA). This research focused on using data mining technique to develop a model for predicting student performance based on 'O' level results and their first 3 semester at each semester. Data preprocessing was done to remove the results of rusticated and expelled student. Results obtained by comparing SVM with other ML techniques such as KNN, decision trees, linear Regression showed that SVM outperforms other ML algorithms. The parameters of the SVM algorithm (kernel) were also tuned to improve its accuracy and result obtained showed that the RBF kernel with penalty (C=100) performed best. SVM and RBF gave the highest training accuracy of 94% and 97% predicting accuracy which outperforms other state of the art ML technique like KNN and decision trees.

Gerritsen (2017) undertook a study to predict student performance with Neural Network. The dataset used for this study was a Moodle log file containing log information about 4601 students over 17 undergraduate courses. To assess the applicability of Neural Networks, the study compared their predictive performance against six other classifiers on this dataset. These classifiers were Naive Bayes, k-Nearest Neighbors, Decision Tree, Random Forest, Support Vector Machine and Logistic Regression and were trained on data obtained during each course. The features used for training originated from LMS data obtained during the length of each course, and ranged from data on time spent on each course page to grades obtained for course assignments. The Neural Network outperformed all other classifiers when it came to accuracy with a score of 66.1% followed by Logistic Regression which had an accuracy score that is 3.5% lower than the Neural Network, at 62.4%. The study concluded that Neural Networks are applicable to student performance prediction and outperform classifiers like k-Nearest

Neighbors, Naive Bayes, Support Vector Machine, Logistic Regression, Decision Tree and Random Forests when general training is used. Prediction accuracy was that Naive Bayes at 0.571, SVM at 0.597, Logistic Regression at 0.624, Decision Tree at 0.528, Random Forest at 0.568 and Neural Network at 0.661. The models dwelled in predicting students' performance. However, the best model in improving learning outcome has not tested thus highlighting the need to test the ability of other machine learning models in improving students' outcome. Contentious results on model that accurately best predicts students' performance and improves learning outcome is evident among various empirical studies highlighted in this paper.

Vinod and Bhatt (2019) investigated performance prediction for post graduate students using Artificial Neural Network. The paper presented a comprehensive study on predicting student performance in R Programming for postgraduate students using deep learning (which is a small part of the artificial neural network). The study ran the experiments on the 4 GB RAM PC, with 1.90GHz of Intel i3 Processor. In evaluating the Artificial Neural Network (Deep Learning), we used R Programming. In order to improve the performance of the model i.e. the accuracy of prediction, Artificial Neural Method was used. Comparing the accuracy of various methods like Linear Regression, Random Forest and Deep Leaning (ANN) we learn that Linear Regression could only produce an Accuracy of 12.339%. Random Forest produced a slightly higher accuracy of 28.101%, while on the other hand, Deep Leaning produced an Accuracy of 97.429% on Total Dataset. Based on the prediction accuracy, it can be stated in this paper that Artificial Neural Networks exhibit more consistent behavior and illustrate better classification outcomes than other traditional classifiers. Since it is very evident that Deep Learning/ Artificial Neural Network can predict the final grades (G3) with an accuracy rate of 97.749 percent on the test dataset that was not used to prepare the model, another dataset with approximately the same accuracy can also be predicted. The technique of ANN modeling has many favorable features such as efficiency, generalization, and simplicity. These features make ANN an attractive choice for more accurate modeling of complex systems. The models dwelled in predicting students' performance. However, the best model in improving learning outcome has not tested thus highlighting the need to test the ability of other machine learning models in improving students' outcome. Contentious results on model that accurately best predicts students' performance and improves learning outcome is evident among various empirical studies highlighted in this paper.

Agrawal and Vishwakarma and Sharma (2017) conducted a study; using data mining classifier for predicting student's performance in UG level. The performance of students is evaluated using four distinct classifiers named as decision tree, random forest, Naive Bayes and rule induction. Different classifiers show different accuracy depending on different algorithms used in it. These analyzed results are explicitly used to predict the upcoming grades of the students and the relevant features (like access to the Internet, study time, etc.) which affect the academic performance of the students. The results revealed that student accuracy prediction of student performance with 90.00% prediction accuracy by decision trees, 84.00% prediction accuracy by Naive bayes, 85.00% prediction accuracy by random forest and 82.00% prediction accuracy by induction rule. The model that can be used to improve learning outcome was never tested presenting a conceptual gap. Moreover, the contentious results on model that accurately best

predicts students' performance and improves learning outcome as evidenced by prediction accuracy of the machine learning models warrant further investigation.

Al-Shabandar, Hussain, Laws, Keight, Lunn and Radi (2017) studied machine learning approaches to predict learning outcomes in Massive open online courses. Exploratory Data Analysis demonstrates that there is strong correlation between click stream actions and successful learner outcomes. Various Machine Learning algorithms have been applied to enhance the accuracy of classifier models. Simulation results from our investigation have shown that Random Forest achieved viable performance for our prediction problem, obtaining the highest performance of the models tested. Conversely, Linear Discriminant Analysis achieved the lowest relative performance, though represented only a marginal reduction in performance relative to the Random Forest. The simulation results in both experiments indicate that RF and SVM achieved ideal performance, with the accuracy values of 0.9881 and 0.9851 respectively. Other classifier models gave lower performance, for instance NB showed a value of accuracy 0.9794, and 0.9621 for both set of experiments. The results show that machine learning is a viable approach to our problem, providing an exceptional capability to distinguish between success and failure outcomes. The model that can be used to improve learning outcome has never been suggested presenting a conceptual gap. Moreover, the contentious results on model that accurately best predict students' performance and improve learning outcome need to be studied.

Sekeroglu, Dimililer and Tuncal (2019) conducted a study on student performance prediction and classification using machine learning algorithms. Two datasets; Student Performance Dataset (SPD) and Students Academic Performance Dataset (SAPD), are considered in this research. Analysis of educational data especially the effect of social environment and family on the students' performance is highly important to improve the quality of education for future generations by enhancing the factors. For this reason, analysis of different and varied datasets in order to predict and classify the behaviour of students in related courses and provide early intervention to increase the performances has vital importance. Different machine learning algorithms are useful and effective for different kinds of problems as prediction and classification. Data fed to algorithms without any data selection algorithm and significant results are obtained. In prediction, minimum Mean Square Value and highest R2 and EV Scores are obtained by SVR. Even Backpropagation (BP) produced lowest prediction rates; it was superior to other classification algorithms in classification experiments with 87.78%. The model that can be used to improve learning outcome has never been suggested presenting a conceptual gap. Moreover, the contentious results on machine learning models that accurately best predicts students' performance and improves learning outcome warrant further investigation.

Nohuddina and Zuraini (2018) undertook a study on monitoring students' performance using Self Organizing Map Trend Clustering. The analysis of relation between student performance and other variables in education setting is often useful in identifying influential factors on performance. Data mining is referred to an analytic process designed that discovers data patterns and relationships between datasets. In this study, clustering is used to cluster student grade datasets to generate trend line clusters. The study deduced that from the SOM Cluster and Trend maps, lecturers and teachers are able to investigate the trends of student performances throughout

the semester. With the generated maps, lecturers and teachers are able to identify types and numbers of trend lines represent student performance based in their academic assessments. The model that can be used to improve learning outcome has never been suggested presenting a conceptual gap. Moreover, the contentious results on machine learning models that accurately best predicts students' performance and improves learning outcome warrant further investigation.

Pojon (2017) undertook a study; using machine learning to predict student performance. Three different machine learning methods were used in this thesis. They are linear regression, decision trees, and naïve Bayes classification. Feature engineering, the process of modification and selection of the features of a data set, was used to improve predictions made by these learning algorithms. Two different data sets containing records of student information were used. The machine learning methods were applied to both the raw version and the feature engineered version of the data sets, to predict the student's success. The results show that it is possible to predict student performance successfully by using machine learning. The best algorithm was naïve Bayes classification for the first data set, with 98 percent accuracy, and decision trees for the second data set, with 78 percent accuracy. Feature engineering was found to be more important factor in prediction performance than method selection in the data used in this study. However, the model that can be used to improve learning was not identified. The model that can be used to improve learning outcome has never been suggested presenting a conceptual gap. Moreover, the contentious results on machine learning models that accurately best predicts students' performance and improves learning outcome warrant further investigation.

Kim (2019) conducted a study on Artificial Neural Network to predict student outcomes. In particular, the study introduced GritNet architecture and developed an unsupervised domain adaptation method to transfer a GritNet trained on a past course to a new course without any student outcome label. Our results for real Udacity student graduation predictions show that the GritNet not only generalizes well from one course to another across different Nanodegree programs, but also enhances real-time predictions explicitly in the first few weeks when accurate predictions are most challenging. This method is effective in the sense that it works across different courses varying in lengths, format and contents and does not require custom feature engineering or additional target-course data or labels.

Keshtkar (2018) undertook a study; predicting Risk of Failure in Online Learning Platforms Using Machine Learning Algorithms for Modeling Students' Academic Performance. The dataset are extracted into online and offline courses. These datasets were used to create classification models for each course using different algorithm. The results showed that in the classification methods, our algorithms are finding useful patterns that we can use to predict risk of failure in students' outcome. The results on a classification model, where the goal is to determine the grade of the student in an online course, were not satisfactory compared to offline courses but beat the baseline in the first two sessions. However, our results show that using Logistic Model Tree (LMT) can help us to predict the negative learning performance; doing so, we are able to detect which students are at risk of failure in offline courses.

Zohair (2019) conducted a study prediction of Student's performance by modelling small dataset size. Prediction of student's performance became an urgent desire in most of educational entities and institutes. This research explores as well the possibility of identifying the key indicators in the small dataset, which will be utilized in creating the prediction model, using visualization and clustering algorithms. Best indicators were fed into multiple machine learning algorithms to evaluate them for the most accurate model. Among the selected algorithms, the results proved the ability of clustering algorithm in identifying key indicators in small datasets. The main outcomes of this study have proved the efficiency of support vector machine and learning discriminant analysis algorithms in training small dataset size and in producing an acceptable classification's accuracy and reliability test rates. The model that can be used to improve learning outcome has never been suggested presenting a conceptual gap. Moreover, the conflicting results on machine learning models that accurately best predicts students' performance and improves learning outcome warrant further study.

Lau, Sun and Yang (2019) conducted a study on modelling, prediction and classification of student academic performance using artificial neural networks. This paper presents an approach with conventional statistical analysis and neural network modelling/prediction of students' performance. Conventional statistical evaluations are used to identify the factors that likely affect the students' performance. The neural network is modelled with 11 input variables, two layers of hidden neurons, and one output layer. Levenberg–Marquardt algorithm is employed as the backpropagation training rule. The performance of neural network model is evaluated through the error performance, regression, error histogram, confusion matrix and area under the receiver operating characteristics curve. Overall, the neural network model has achieved a good prediction accuracy of 84.8%, along with limitations. However, it is not clear whether neural network model can be employed in improving learning outcome of students.

Ulinnuha, Sa'Dyah and Rahardjo (2017) conducted a study in predicting academic performance using Random Forest, Artificial Neural Network, Naïve Bayesian and Logistic. There are several techniques that have been studied in order to solve admission problem, however the study chose to compare the four models. In the present study, the study compared random forest algorithm with artificial neural network, naïve Bayesian and logistic regression. The study found that the composition of training dataset is important. The study also found that random forest is often over fit and in accuracy is not better than other methods for dataset with small fitur. Random Forest in accuracy is not better than other methods for dataset with small fitur. Neural network still give better performance than others in several testing set. In addition, naïve bayes and logistic regression have same performance for this dataset.

Reiter-Haas, Slawicek and Lacic (2017) undertook a study on studo Jobs, enriching data with predicted job labels. The study performed an exhaustive comparison study of state-of-the-art classifiers to be used for label prediction in the job domain. Results of the study revealed that in most cases an SVM based approach using stochastic gradient descent performs best on the textual content of job advertisements in terms of Accuracy, F1-measure and AUC. Few or none empirical studies have employed to predict students' performance. It may be practical to test accuracy of stochastic gradient descent in predicting students' performance.

Anderson, Boodhwani and Baker (2019) conducted a study to predict graduation at a public R1 university using linear support vector machines, decision trees, logistic regression, and stochastic gradient descent binary classifiers. The study used a data set of over 14,000 students from six Fall cohorts, containing 104 features, drawn from pre-existing university data. The prediction accuracy was as follows; decision tree 0.786, linear SVM 0.801, logistic regression, 0.810 and stochastic gradient descent binary classifiers 0.824. In the context of this, study stochastic gradient descent binary classifiers is best in predicting students' graduation chances, however stochastic gradient descent binary classifiers may not be the best in predicting students' performance. Moreover, it is evident whether SGD Classifier can employed to improve learning outcome hence need to undertake further study on various machine learning models and prediction of students' performance.

Diab (2019) undertook a study; optimizing Stochastic Gradient Descent in Text Classification Based on Fine-Tuning Hyper-Parameters Approach by Studying on Automatic Classification of Global Terrorist attacks. The study explored different settings for representation, transformation and weighting features from the summary description of terrorist attacks incidents obtained from the Global Terrorism Database as a pre-classification step, and validated SGD learning on Support Vector Machine (SVM), Logistic Regression and Perceptron classifiers by stratified 10-K-fold cross validation to compare the performance of different classifiers embedded in SGD algorithm. The research indicated that using a grid-search to find the hyperparameters optimize SGD classification, not in the preclassification settings only, but also in the performance of the classifiers in terms of accuracy and execution time. Using SGD learning with hyper-parameters improves the accuracy and the execution time of SGD on (SVM, Logistic Regression, and Perceptron) classifiers from SVM (0.87829), Logistic Regression (0.86872), and Perceptron (0.85306) without using SGD learning to SVM (0.87946), Logistic Regression (0.87192), and Perceptron (0.87378) after using SGD learning with hyper-parameters. There may be need to test Stochastic Gradient Descent as a method to predict students' performance by comparing it with other machine learning approaches. Table 1, summarizes various machine Learning Models, architecture, algorithm and prediction accuracy of student performance.

Stratford Peer Reviewed Journals and Book Publishing
Journal of Information and Technology
Volume 4||Issue 1||Page 23-45||March||2020|
Email: info@stratfordjournals.org ISSN: 2616-3573

Table 1: Various machine Learning Models and prediction accuracy

Model	Architecture	Type	Algorithm	Author	Prediction Accuracy
Decision Tree	Recursive partition Decision rules	Nonlinear	C4.5 algorithm	Al-Shabandar et al. (2017)	98.3%
				Pojon, (2017)	78.0%
				Agrawal, S., Vishwakarma, S. K., & Sharma, A. K. (2017)	90.0%
				Stapel, M., Zheng, Z., & Pinkwart, N. (2016)	71.5%
				Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016)	75.8%
Random Forest	Ensemble DT	Nonlinear	Random subset Features Bootstrap	Al-Shabandar et al. (2017)	98.81
				Ulinnuha,N., Sa'Dyah, H., Rahardjo, M. (2017)	68.27%
				Agrawal, S., Vishwakarma, S. K., & Sharma, A. K. (2017)	85.0%
				Stapel, M., Zheng, Z., & Pinkwart, N. (2016)	67.9%
Support Vector Machine	Hyperplane kernel trick	Nonlinear	Quadratic Optimisation	Al-Shabandar et al. (2017)	98.44%
				Sekeroglu, B., Dimililer, K., & Tuncal, K. (2019)	79.38%
				Zohair, L. M. A. (2019)	76.3
				Oloruntoba, S. A., & Akinode, J. L. (2017).	94.00%
Naive Bayes	Bayesian Decision Rule	Linear	Maximum Likelihood Estimation	Al-Shabandar et al. (2017)	97.94%
				Acharya and Sinha (2014)	66.0%
				Pojon, (2017)	98.0%
				Agrawal, S., Vishwakarma, S. K., & Sharma, A. K. (2017)	84.0%
				Stapel, M., Zheng, Z., & Pinkwart, N. (2016)	65.4%
				Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016)	67.7%
Feedforward Neural Network	Units 14-3-2	Nonlinear	Back propagat ion	Al-Shabandar et al. (2017)	98.56%
				De Albuquerque et al. (2015)	85.0%
				Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016)	79.2%
				Vinod and Bhatt, (2019)	97.43%
Logistic regression	Generalised Linear Model	Linear	Maximum Likelihood	Al-Shabandar et al. (2017)	97.54%
				Keshtkar, F. (2018)	86.2%

			Estimation	Ulinnuha,N., Sa'Dyah, H., Rahardjo, M. (2017)	75.96%
				Stapel, M., Zheng, Z., & Pinkwart, N. (2016)	68.2%
Linear Discriminant Analysis	Generalized Linear Model	Linear	Maximum Likelihood Estimation	Al-Shabandar et al. (2017)	96.56
				Zohair, L. M. A. (2019)	71.1%
Self-Organised Map	Unit 25-3-2	Nonlinear	Competitive learning	Al-Shabandar *et al.* (2017)	9765%
				Kurdthongmee, W. (2008)	67.3%
Stochastic Gradient Descent Binary Classifiers	Adam	Linear	AdaGrad and RMSProp algorithms	Boodhwani and Baker (2019)	82.4%
				Diab, S. (2019).	SGD improves prediction accuracy of SVM (87.83%), Logistic Regression (86.87%), and Perceptron (85.3%) without using SGD learning and SVM (87.95%), Logistic Regression (87.19%), and Perceptron (87.38%) after using SGD learning with hyper-parameters

4. Research Gaps

The empirical review indicated contentious results on model that best predicts students' performance. Moreover, it is not clear among the various machine learning algorithms which one derives the best approach in predicting students' performance while improving learning outcome. Hussain, Muhsin, Salal, Theodorou, Kurtoğlu and Hazarika (2019) indicated that Artificial Neural Network produces the highest classification accuracy in predicting students' performance at 95.34%. However, according to Belachew and Gobena (2017), Naive Bayesian compared to Support Vector Machine, multi-layer perceptron network and neural networks was more accurate in predicting students' performance in the department of Information Technology at 95.7% prediction accuracy. Jayaprakash, Balamurugan and Chandar (2018) further indicated that Naive Bayes was best in predicting students' academic performance with accuracy prediction of 92.2%. Obsie and Adem (2018) noted that Linear Regression and Support Vector Regression performed better than Neural Network in predicting student academic performance where prediction accuracy for neural network was 97.63%, Support Vector Regression was 98.05% and Linear Regression 98.05%. In terms of time to predict results, Support Vector Regression was best at 0.03 seconds, followed by 0.03 seconds for Linear Regression 0.05 seconds and 0.78 seconds for Support Vector Regression.

Employing, linear regression to predict students' academic performances, Oyerinde and Chia (2017) indicated that linear Regression accurately predicts students' performance at 89.0% prediction accuracy. Gerritsen (2017) while predicting student performance using machine learning revealed that Neural Network outperforms all other classifiers when it comes to accuracy with a score of 66.1% followed by Logistic Regression which has an accuracy score that is 3.5% lower than the Neural Network, at 62.4%. The study concluded that Neural Networks are applicable to student performance prediction and outperform classifiers like k-Nearest Neighbors, Naive Bayes, Support Vector Machine, Logistic Regression, Decision Tree and Random Forests when general training is used. Prediction accuracy was that Naive Bayes at 0.571, SVM at 0.597, Logistic Regression at 0.624, Decision Tree at 0.528, Random Forest at 0.568 and Neural Network at 0.661. The best model to improve learning outcome of students was not identified presenting a conceptual gap. Vinod and Bhatt (2019) also indicated that Artificial Neural Network is best in predicting students' performance at 97.429% compared to Linear Regression at 12.339% and Random Forest at 28.101%. Agrawal and Vishwakarma and Sharma (2017) while predicting student's performance revealed 90.00% prediction accuracy by decision trees, 84.00% prediction accuracy by Naive Bayes, 85.00% prediction accuracy by random forest and 82.00% prediction accuracy by induction rule.

Moreover, a substantial amount of literature focuses on predicting student performance in solving problems or completing courses (Yadav & Pal, 2012). Many machine learning techniques, such as decision trees, artificial neural networks, matrix factorization, collaborative filters and probabilistic graphical models have been applied to develop prediction algorithms (Lakkaraju, *et al.,* 2015). However, it is not clear which among the various machine learning models accurately predicts students' performance as various authors have presented conflicting results regarding the prediction accuracy of the models. Moreover, the studies conducted by

other authors have not identified a machine learning model that can be employed to improve learning outcome of students.

It is evident that different machine learning reveals different prediction accuracy of students' performance. No clear model has been proved to be the best in accurately predicting students' performance. The varying prediction level by various machine learning models may be as a result of differences in socioeconomic factors of students like family income, parental level of education and employment status of students or parents. It may also be important to consider student's academic performance affected by many factors, like personal, socio-economic and other environmental variable when testing the accuracy of various machine learning models in predicting student performance. Moreover, the various machine learning models did not identify the best model in improving student's outcome. It is not clear which model is best in predicting performance and at the same time best in improving learning in among students. Most models focused largely in predicting students' performance without considering mechanisms to improve learning experience by students. Students' performances is influenced by socioeconomic factors of students like family income, parental level of education and employment status of students or parents but are not considered when testing the accuracy of various machine learning models in predicting students' performance. As a result, it is important to test the accuracy of various machine learning models that accurately and precisely predicts students' performance and the one that is most appropriate in improve learning outcome while considering effects of socio economic and demographic factors of students on academic performance.

5. Conclusions

The study makes a conclusion that predicting students' performance of a student is of the highest priority for any learning institution across the globe. The application of Machine Learning Techniques in predicting students' performance proved to be helpful for identifying poor performers and it can enable tutors to take remedial measures at an earlier stage, even from the very beginning of an academic year using only students' internal assessment data of previous semesters, in order to provide additional help to the groups at risk. Using various methods to predict the performance of the student accurately would be highly required. Predicting the performance would also enable the institutions to focus more on students having more probability of performing lower in order to improve their performance. The prediction of the student's success helps the education organization to supply the student with additional assessment; this process also enhances the development of the education system in educational institutions.

From empirical results, by going through the internal assessment of a candidate the result of the candidate might be predicted. So, such candidates with poor internal assessment marks, the tutors may devote some extra time to improve the final examinations marks. The students whose performance is poor, predictive model is used as a signal to students and parents, it might be applied to improving the marks of such candidates. The instructors might intervene in real time by looking at the internal assessment marks secured by the students. So, the internal assessment may be continued as an integral part of a particular course. However, basing on empirical

literature, there is contentious results on which is the best machine learning model in predicting students' performance. It is important to accurately rank machine models based on their prediction capabilities in predicting students' performance prediction and subsequent decision making. Moreover, most machine learning models dwelled much in studying students' performance prediction but failed to identify the best model in improving students' outcome. Machine learning model (s) with capability to improve students learning outcome should be identified. Besides, educational research shows that some socioeconomic, psychological factors, such as learning style, self-efficacy, motivation and interest, and teaching and learning environment, also play a role in student learning and thus affect student achievement.

References

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.

Allen-Zhu, Z. (2017). Katyusha: The first direct acceleration of stochastic gradient methods. *The Journal of Machine Learning Research*, *18*(1), 8194-8244.

Al-Shabandar, R., Hussain, A., Laws, A., Keight, R., Lunn, J., & Radi, N. (2017, May). Machine learning approaches to predict learning outcomes in Massive open online courses. In *2017 International Joint Conference on Neural Networks (IJCNN)* (pp. 713-720). 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.

Anderson, H., Boodhwani, A., & Baker, R. (2019). Predicting graduation at a public R1 university. In *Proceedings of the 9th International Learning Analytics and Knowledge Conference*.

Bassi,J. S., Dada,E. G., Hamidu, A., A. & Dauda, M., E. (2019). Students Graduation on Time Prediction Model Using Artificial Neural Network, *Journal of Computer Engineering, 21*(3), 28-35.

Belachew, E. B., & Gobena, F. A. (2017). Student Performance Prediction Model using Machine Learning Approach: The case of Wolkite University. *International Journal if Advanced Research in Computer Science and Software Engineering*, *7*(2), 46-50.

Buenaño-Fernández, D., Gil, D., & Luján-Mora, S. (2019). Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study. *Sustainability*, *11*(10), 2833-2851.

Diab, S. (2019). Optimizing Stochastic Gradient Descent in Text Classification Based on Fine-Tuning Hyper-Parameters Approach. A Case Study on Automatic Classification of Global Terrorist Attacks. *arXiv preprint arXiv:1902.06542*.

Gerritsen, L. (2017). Predicting student performance with Neural Network, Tilburg University, Netherlands.

Hussain, M., Zhu, W., Zhang, W., & Abidi, S. M. R. (2018). Student engagement predictions in an e-learning system and their impact on student course assessment scores. *Computational intelligence and neuroscience*, Article ID 6347186, 21 pages https://doi.org/10.1155/2018/6347186

Hussain, S., Muhsin, Z. F., Salal, Y. K., Theodorou, P., Kurtoğlu, F., & Hazarika, G. C. (2019). Prediction Model on Student Performance based on Internal Assessment using Deep Learning. *International Journal of Emerging Technologies in Learning*, *14*(8), 4-22.

Jayaprakash, S., Balamurugan E. & Chandar, V. (2018). Predicting Students Academic Performance using Naive Bayes Algorithm**,** BlueCrest College Accra, Ghana.

Keshtkar, F. (2018). Predicting Risk of Failure in Online Learning Platforms Using Machine Learning Algorithms for Modeling Students' Academic Performance, Southeast Missouri State University.

Kim, B. (2019). Deep Learning To Predict Student Outcomes, Presented at the ICLR AI for social good workshop 2019.

Koutina, M., & Kermanidis, K. L. (2011). Predicting postgraduate students' performance using machine learning techniques. In *Artificial intelligence applications and innovations* (pp. 159-168). Springer, Berlin, Heidelberg.

Kumar, M., Singh, A. J., & Handa, D. (2017). Literature survey on educational dropout prediction. *International Journal of Education and Management Engineering*, *7*(2), 8-19.

Kurdthongmee, W. (2008). Utilization of a Self Organizing Map as a Tool to Study and Predict the Success of Engineering Students at Walailak University. *Walailak Journal of Science and Technology*, *5*(1), 111-123.

Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., & Addison, K. L. (2015, August). A machine learning framework to identify students at risk of adverse academic outcomes. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1909-1918).

Lau, E. T., Sun, L., & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Applied Sciences*, *1*(9), 982-992.

Nohuddina,P.N. E. & Zuraini Z. A. N. (2018). Monitoring Students Performance using Self Organizing Map Trend Clustering, *ZULFAQAR International Journal of Defence Science, Engineering & Technology*, 1 (1) 50-56.

Obsie, E., Y. & Adem , S. A. (2018). Prediction of Student Academic Performance using Neural Network, Linear Regression and Support Vector Regression: A Case Study, *International Journal of Computer Applications*, 180 (40), 39-47.

Olaniyi, A. S., Kayode, S. Y., Abiola, H. M., Tosin, S. I. T., & Babatunde, A. N. (2017). Student's Performance Analysis Using Decision Tree Algorithms. *Annals. Computer Science Series*, *15*(1), 55-62.

Oloruntoba, S. A., & Akinode, J. L. (2017). Student academic performance prediction using support vector machine. *International Journal of Engineering Sciences and Research Technology*, *6*(12), 588-597.

Oyerinde, O. D., & Chia, P. A. (2017). Predicting students' academic performances–A learning analytics approach using multiple linear regression. *International Journal of Computer Applications*, 157 (4), 437-44.

Pandey, M., & Sharma, V. K. (2013). A decision tree algorithm pertaining to the student performance analysis and prediction. *International Journal of Computer Applications*, *61*(13), 1-5.

Pojon, M. (2017). Using machine learning to predict student performance (Master's thesis). University of Tampere, Finland.

Rajalaxmi, R. R. Natesan, P. Krishnamoorthy, N. Ponni, S. (2019). Regression Model for Predicting Engineering Students Academic Performance, *International Journal of Recent Technology and Engineering*, 7(6S3), 71-75.

Reiter-Haas, M., Slawicek, V., & Lacic, E. (2017). Studo Jobs: Enriching Data With Predicted Job Labels. In *Workshop on Recommender Systems and Social Network Analysis co-located with i-KNOW*.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67(1)*, 93-104.

Sekeroglu, B., Dimililer, K., & Tuncal, K. (2019, March). Student performance prediction and classification using machine learning algorithms. In *Proceedings of the 2019 8th International Conference on Educational and Information Technology* (pp. 7-11). ACM.

Shaziya, H., Zaheer, R., & Kavitha, G. (2015). Prediction of students performance in semester exams using a Naïve Bayes classifier. *International Journal of Innovative Research in Science, Engineering and Technology*, *4*(10), 9823-9829.

Soni, A., Kumar, V., Kaur, R., & Hemavath, D. (2018). Predicting student performance using data mining techniques. *International Journal of Pure and applied Mathematics*, *119*(12), 221-227.

Soule, P. (2017). Predicting Student Success: A Logistic Regression Analysis of Data From Multiple SIU-C Courses.

Stapel, M., Zheng, Z., & Pinkwart, N. (2016). An ensemble method to predict student performance in an online math learning environment. In Proceedings of the 9th international conference on educational data mining, international educational data mining society (pp. 231–238).

Sultana, M., J. Rani,U. & Farquad, M.A.H. (2019). Student's Performance Prediction using Deep Learning and Data Mining Methods, *International Journal of Recent Technology and Engineering*, 8(1S4), 1018-1021.

Swamy, M. N., & Hanumanthappa, M. (2012). Predicting academic success from student enrolment data using decision tree technique. *Int. J. Appl. Inf. Syst*, *4*(3), 1-6.

Ulinnuha,N., Sa'Dyah, H., Rahardjo, M. (2017). A Study of Academic Performance using Random Forest, Artificial Neural Network, Naïve Bayesian and Logistic Regression, Institut Teknologi Sepuluh Nopember, Indonesia.

Vinod K. P. & Bhatt, V. K. K. (2019). Performance Prediction for Post Graduate Students using Artificial Neural Network, *International Journal of Innovative Technology and Exploring Engineering*, 8(7S2),446-454.

Wilbur, W. J., & Kim, W. (2014). Stochastic gradient descent and the prediction of MeSH for PubMed records. In *AMIA Annual Symposium Proceedings* (Vol. 2014, p. 1198). American Medical Informatics Association.

Yadav, S. K., & Pal, S. (2012). Data mining: A prediction for performance improvement of engineering students using classification. *arXiv preprint arXiv:1203.3832*.

Zohair, L. M. A. (2019). Prediction of Student's performance by modelling small dataset size. *International Journal of Educational Technology in Higher Education*, *16*(1), 27-45.