International Journal for Modern Trends in Science and Technology, 8(06): 301-308, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0806052
Available online at: http://www.ijmtst.com/vol8issue06.html

Student Performance Prediction Using Machine Learning Algorithms

P. Pavani[1] | Dr. G.N.V.G. Sirisha[1] | R. Krishna Chaitanya[2]

[1]Department of Computer Science and Engineering, S.R.K.R Engineering college, Bhimavaram, Andhra pradesh, India.
[2] Department of Electronics and Communication Engineering, S.R.K.R Engineering college, Bhimavaram, Andhra pradesh, India.
Corresponding Author Email Id: peddapavani@gmail.com

To Cite this Article
P. Pavani, Dr. G.N.V.G. Sirisha and R. Krishna Chaitanya. Student Performance Prediction Using Machine Learning Algorithms. International Journal for Modern Trends in Science and Technology 2022, 8(06), pp. 301-308. https://doi.org/10.46501/IJMTST0806052

Article Info
Received: 16 May 2022; Accepted: 12 June 2022; Published: 16 June 2022.

ABSTRACT

In the educational domain, the use of machine learning techniques aids in the discovering of hidden knowledge and patterns, like student performance. Students performance can be predicted in advance with the use of supervised learning models. Early analysis of students' performance based on personal traits, demographical Features, academic background Features, behavioral Features, and past academic performance helps both students and teachers. This paper aims at building models to predict student performance based on personal traits, demographical features, academic background features, behavioral features. The performance of different supervised learning algorithms like Random Forest Classifier, XGBoost Classifier, KNN, Multilayer Perceptron, Gaussian Naive Bayes Classifier, Decision Tree Classifier, Support Vector Machines, Logistic Regression, Perceptron, AdaBoost Algorithm, Gradient Boosting Algorithm when applied on Kalboard 360 dataset are compared. XGBoost Classifier performance is found to be better than other classification algorithms with an F1-Score of 0.82.

KEYWORDS: Student Academic Performance, Machine Learning, Ensemble Methods, Support Vector Machines.

1. INTRODUCTION

In the developnment of a country, education is critical. Though many people get enrolled into schools and college, with time there are many dropouts. Early analysis of students' performance in schools and colleges can help in reducing dropouts. This will be helpful for both students and teachers.

Student academic improvement and learning performance in online learning platforms are affected by numerous factors like Visited Resources, Raised Hands, Announcements Viewed, Discussion, and Categorical factors like Gender, Relation, Nationality, Grade Id, Place of Birth, Semester, Stage Id, Topic, Parent School Satisfaction, Parent Answering Survey, Class, and Student Absence Days. Much research has been conducted in the area of student achievement, and these studies have identified and evaluated a variety of elements that influence a student's academic performance at the school, college, and university levels. Student academic performance can be predicted using Academic Background Features, Demographical Features, Behavioral Features, and Parents Participation Features.

Machine Learning (ML) is a branch of Artifical Intelligence (AI). Machine Learning assists in the development of systems and applications that improve over time. Supervised, Unsupervised, and Reinforcement Learning are the three types of machine

learning tasks. Predicting student performance involves predicting their outcomes. Much research has been conducted in which classification is the most common approach of prediction. For this purpose, researchers have used Machine Learning models like Decision Tree (DT), Support Vector Machines (SVM), Naïve Bayes (NB), Random Forest (RF), Logistic Regression, Multi-Layer Perceptron (MLP), Perceptron, XGBoost, KNN, AdaBoost, Gradient Boosting Algorithm. Student academic performance prediction is a supervised learning task.

1.1. Problem Statement

Early prediction of student performace based on personal traits, demographical features, academic backgroud features, behavioral features, and past academic performance helps both students and teachers.

Teachers: Teachers can focus on students who are anticipated to attain low grades. Student benefit from early communication and teacher assistance in improving their academic performance.

Students: This early prediction also help students to improve their knowledge by focusing more and requesting teachers' or peers' help when required.

2. LITERATURE SURVEY

Mishra, et al., have used the J48, and Random Forest algorithm to predict the performance of MCA students [9]. Based on student academic integration, social integration, and various emotional skills,classification techniques are used to predict student performance. These methods are applied to a datasetconsisting of 10330 students, 14 attributes. The algorithms are evaluated using True positive rate, Precision, and Recall as metrics. Random Forest outperformed other methods with an accuracy of 94%.

Kaunang, et al., have used random forest and decision tree to predict student academic performance by considering computer science student's data [8]. Data consists of students' demographics, previous GPA, family background. The dataset consists of 249 records with different classes (poor, average, good). The performance of decision tree is better than random forest in terms of classifier accracy. The reported accuracy of decision tree is 66.9%.

Devasia, et al., have used Naive Bayesian to predict students' performance. Student admission details and course details are used as features [6]. These methods are applied to a dataset consisting of 700 students, 19 attributes. The input is the student's academic history, and the output is the student's performance in the upcoming semester. The Naive Bayesian algorithm outperformed other methods like Regression, Neural Network, Decision Tree, etc with an accuracy of 89%. The end-of-semester performance is predicted using the class test, seminar grades, and assignment grades.

Denny, et al., have used Intelligence Quotient (IQ), Emotional Quotient (EQ) for predicting the academic performance of students [5]. The IQ and EQ levels of students have an impact on their learning and academic achievement. The proposed technique divides the students into clusters during the prediction phase. Students' ability to pass or fail is determined by clusters. Academic learning, academic grades have both been conducted to EQ. When it comes to predicting a students academic performance, both EQ and IQ are equally important.

Samuel, et al., have assessed the role of different feature selection methods in predicting student performance [1]. Discriminant Analysis (DISC), K-Nearest Neighbor (KNN), Decision Tree (DT), and Naive Bayes (NB), are some of the classification techniques used for prediction. These methods are used on a dataset of 500 students and 16 features. Correlation Feature Selection (CFS), Relief F, Sequential Forward Selection (SFS), Kullback-Leibler Divergence, Sequential Backward Selection (SBS), and Differential Evolution are the feature selection methods used. Recall, Precision, and F-measure are the metrics used. In terms of which performance measure KNN is best and how much value is obtained.

Masna, et al., have used a supervised model to assist the instructor in making future recommendations and performance evaluations [14]. MOORA (Multi-Objective Optimization by Ratio Analysis) and SMART (Simple Multi-Attribute Rating) are two simple techniques that use entropy and gain to determine criteria weight and non-criteria weight (or) sub-criteria weight. These methods are tested using a 450-student academic dataset, with 95 students graduating quickly, 98 students graduating moderately, and 257 students graduating slowly. Multi-criteria decision-making can

be used to calculate the weighted performance of student criteria. The accuracy and precision of weight as a result of entropy and weight are 60.9 percent and 52.7 % to 53.01 %, respectively.

Roy, et al., have used education data mining to predict the academic performance of students [11]. Among the data mining techniques used are Naive Bayes, Neural Networks, and decision trees. These methods are applied on datasets consisting of 1400 students, 33 attributes. All algorithms' performance is assessed using the False positive rate, True positive rate, Precision, and Recall. Different types of attributes like Social, school-related factors, and Demographic, can influence student performance. MLP, Naive Bayes, and J48 Decision Tree are used for classification. It helps in reducing students' strengths and weaknesses. Naive Bayes has the highest accuracy of 68.6%.

Burman, et al., have used an SVM to predict the student's academic performance. Students are categorized as having low, high, average academic scores [2]. SVM performance is evaluated using Sensitively, Specificity, and Accuracy. On the training data set, the performance of SVM with different kernels like Linear Kernel(LK) and Radial Basis Kernel(RBF) is shown. When compared to the LK, RBF produces better results. These methods are applied to a 1,000-record dataset based on 29 non-intellectual (short-term or long-term goal) constructs of students which have a major effect on academic growth and study. With a % accuracy, RBF consistently outperforms LK.

Dangen, et al., have employeed the Naive Bayes Classifier (NBC) to access student academic evaluation techniques [4]. This dataset contains 279 samples (2014 to 2017), including the total GPA (TGPA) and progress GPA (PGPA). The course subjects' scores are used to calculate the PGAP and TGPA. Over the span of 4 semesters, we focus on 3 graduation criteria (fast, on, and delay times). According to the data, the algorithm's accuracy (AC) is 76.9% and its true-positive rate (TP) is 44.6 %.

Hamsa, et al., have used the Fuzzy Genetic algorithm and Decision tree to predict student's academic performance by considering Master's degree and Bachelor's student's data in computer science, Electronics, and communication streams [7]. Internal marks, session marks, admission scores were used in the research work. These methods are applied on a dataset consisting of 120 students, 3 attributes. The Fuzzy Genetic Algorithm has higher accuracy.

3. ARCHITECTURE

Different classification models based on Academic Background features, Demographic features, and Behavioural features are used to predict student academic performance in this study. The proposed methodology's main steps are depicted in Figure 1. In this process, data from Kaggle is collected first. The dataset is collected from the Kalboard 360 E-Learning system. After that, there is a data preprocessing step, which involves converting the obtained data into a format that can be used.

The techniques used to assess students performance are Decision Tree, Support Vector Machine, Random Forest, Multi-Layer Perceptron, Logistic Regression, Perceptron, Gaussian Naive Bayes, XGBoost, KNN, AdaBoost, and Gradient Boosting Algorithm.

3.1 Dataset Details

The educational dataset used in this study was collected from the Kalboard 360 LMS. There are 480 tuples and 16 attributes in it. Academic Features, Demographic Features, parents participation Features and Behavioral Features are the 4 types of features. Based on their total marks or grades, the student's grades are classified into three categories: High-0, Low-1, and Medium-2.

3.2 Data Pre-Processing

It's a crucially important step in improving the quality of data before it's fed into machine learning algorithms. This process involves phases such as data cleaning, data transformation, and data reduction.

Data preprocessing is an important step. They are import the libraries, import the dataset, handling the missing data, handling of categorical data, and feature scaling or Standard Scalar or Normalization.

3.3 Machine Learning Models

3.3.1 Traditional Methods

Logistic Regression

It's used to predict a categorical dependent variable's result. The value of the variable is the output or prediction of linear regression, but the probability of occurrence is the output of logistic regression. It solves classification problems. It uses sigmoid function which is an S-shaped curve. Advantages are Logistic Regression is easy to understand, model training and

prediction are fast. It shows good accuracy for simple datasets and is resistant to overfitting. Disadvantages are sometimes simple to capture complex relationships between variables.

Decision Tree

A decision tree is a graphic representation of all the different solutions to a problem based on a set of criteria. Advantages of the decision tree are easy to understand and implement, useful in data exploration, handle outliers, where data type is not a constraint. Disadvantages are Decision tree is not often used for prediction because it's also often too simple and not powerful enough for complex data, and may lead to overfitting.

Navie Bayes

Based on the Bayes theorem, it is a classification technique. When the dataset size is insufficient, Naive Bayes is a simple and easy algorithm that may outperform a more complex model due to its simplicity.

Types of Naive Bayes

Gaussian: It's used to categorize a variable that follows a normal distribution.

Multinomial: It is used to categorize discrete valued attributes.

Advantages are naive bayes is easy to understand and implement, and better performance compares to other models. Disadvantages are unable to predict in cases where the features are correlated.

KNN

The KNN algorithm, also known as the k-nearest neighbor algorithm, keeps track of all variable examples and uses a similarity metric to classify new cases. The "K" in the KNN algorithm stands for the number of closest neighbors from whom votes are taken. The advantages of KNN is very simple (basic) and easy to implement, it learns a non-linear decision boundary, and there is no training required. Disadvantages of KNN are it does not work well with high dimension, it does not handle categorical Features well, and needs a large number of samples for accuracy.

Support Vector Machine

SVM is a widely used algorithm for categorizing data. A hyperplane is used as a decision boundary between the various classes. In general, SVM can be used to generate a large number of separating hyperplanes to divide data into segments. Support vector machines have the advantage of being able to solve classification and regression issues at the same time. The SVR (support vector regressor) is used for regression problems when it is best known for classification. High memory usage, high processing costs, time-consuming training, and difficulty understanding the algorithm's structure are all drawbacks.

3.3.2 Neural Networks

Perceptron

A perceptron is a neural network. It is composed of two layers. There are many two layers: input and output. It is a linear classifier that can only be used on linearly separable datasets.

Multi-Layer Perceptron

It is a neural network with several layers. In a multi-layer perceptron, there must be one or more hidden layers. There must be a minimum of three layers since we have one input, one output and then you must have a hidden layer.

Back propagation

Back propogation is a neural network training technique that is only employed during the training phase. There are two parts to the backpropagation algorithm:

Forward pass: The expected output that corresponds to the given inputs is evaluated in the forward pass

Backward pass: The network is updated using partial derivatives of the cost function for different parameters.

3.3.3 Enemble Methods

Ensemble Methods improves overall performance by combining the predictions of several base estimators. Ensemble methods can be divided into 2 different groups:

Parallel ensemble methods [Bagging]

In parallel, base learners are generated. The basic motivation is to make use of the base learners' independence.

Eg: Random Forest.

Sequential ensemble methods [Boosting]

In sequential order base learners are generated. The base learners' dependence will be capitalized.

Eg: AdaBoost, Gradient Boosting, XGBoost.

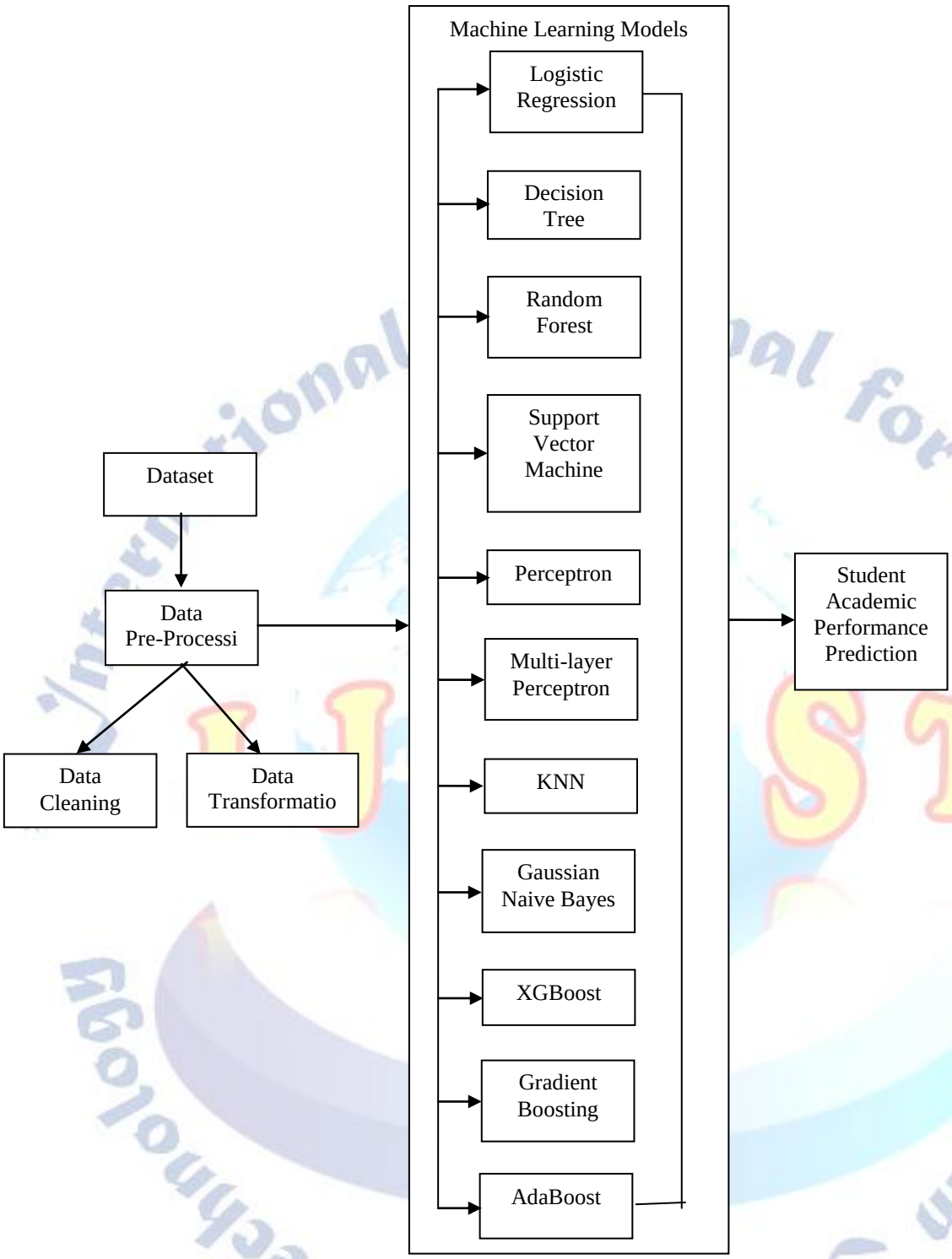

Figure1. Students' Performance Prediction Model

AdaBoost

AdaBoost stands for Adaptive Boosting. It is used for both classification and regression. AdaBoost is the first boosting algorithm to weight various inputs using an ensemble learning approach. It uses multiple weak classifiers to construct one strong classifier. The advantage of Adaboost is the excellent use of weak classifiers for cascading (large quantities). Various classification algorithms can be used as weak classifiers. AdaBoost is a high-precision algorithm. Data imbalance leads to the decreased classification accuracy of the Adaboost algorithm.

Gradient Boosting Algorithm

Gradient Boosting can be used for both classification and regression. It is a collection of many weak models such as decision trees. The advantages of gradient boosting will provide high prediction accuracy, and it requires minimal data pre-processing and it works well great with categorical and numerical features. Disadvantages of gradient boosting may are it sometimes leads to overfitting and it is difficult to interpret the model predictions.

Gradient boosting algorithm takes a long time to generate output.

XGBoost Algorithm

Extreme Gradient Boosting is referred to as XGBoost. It is a fast and efficient implementation of a gradient boosted decision tree. It resembles the gradient boosting framework in appearance, but it is more efficient. It's a tree-based learning system. XGBoost is at least 10 times faster than other gradient boosting implementations. Only a few of the objective functions supported include regression, classification, and ranking.

Advantages

Regularization: L1, L2 regularization is a built-in XGBoost, which prevents the model from overfitting.

Parallel processing: XGBoost takes advantage of parallel processing's benefits. The model is run on multiple CPU cores.

Disadvantages

It is vulnerable to uniform noise and sometimes overfits.

Random Forest

Both regression and classification are possible using the Random Forest algorithm. It's an ensemble learning method. It's a common approach to predictive modeling. The advantages of random forest are high-quality results, fast to train, easy to parallelize, and it works on high-dimensional data. The disadvantages of random forest can be slow to output predictions relative to other algorithms, and models can get very large.

4. Dataset and Exploratory Data Analysis

In this paper, dataset is collected from Kaggle which belongs to Kalboard 360 E-Learning system. The goal of this research is to predict student performance using machine learning models. It also analyzes the accuracy with which current machine learning algorithms predict student performance. Dataset consists of 480 student tuples and 16 attributes. The Features are categorized into three main categories: (1) Demographic Features such as Nationality, Place of Birth, Gender, Relation. (2) Academic background Features such as Stage ID, Grade ID, Section ID, Student Absence Days, Semester, Topic. (3) Parents Participation like Parent School Satisfaction Parent Answering Survey, (4) Behavioral factors such as a Raise hand on Class, Opening or visited resources, Discussion groups, Viewing announcements.

4.1 Data pre-processing

It's a crucially important step in improving the quality of data before it's fed into machine learning algorithms. This process involves phases such as data cleaning, data transformation, and data reduction.

4.1.1 Data Transformation

To represent class labels, data transformation is used to encode data from nominal values into numerical values for classification. Based on a student's grade or marks, different class intervals are taken in the dataset into the following class intervals in Table 1: Low, Medium, and High.

Table 1. Classes Based On Their Numerical Values

Classes	
Interval Value	**Class**
0	High
1	Low
2	Medium

4.1.2 Feature Selection

Feature Selection is regarded as one of the most critical tasks in data preprocessing. The purpose of this step is to decrease the number of available attributes in

the algorithm to a few critical ones lowering the proportion of feature area and removing repeated and inappropriate data. In this way, feature selection improves the data quality and thus the effectiveness of the learning algorithm. There are two types of feature selection approaches: (1) Wrapper-Based methods and (2) Filter-Based methods. The filter method approach is used to identify the relevant traits while ignoring the rest. These methods use variable ranking algorithms to rank the features, allowing highly ranked features to be picked and applied to the learning process. They get different features gain ratio, a filter-based approach, to identify the most proment features for building students' performance models.

As shown in Figure 2, StudentAbsenceDays got the highest rank followed by parent's participation, RaisedHands, VisitedResources, and so on. It's been seen that a significant subset of traits is chosen, while others are excluded. As the result, the features studied in this study received highest ranking, implying that student punctuality and their parents' participation throughout the educational process have a significant impact on their academic achievement (or) performance.

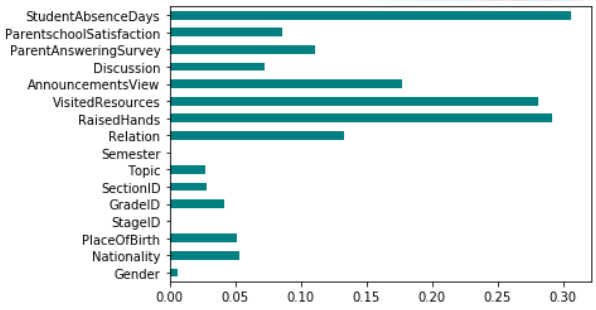

Figure 2. Highly Ranked features applying Filter-Based Method using gain ratio

5. RESULT ANALYSIS

Algorithm performance is measured using metrics like Precision, Recall, Accuracy, and F1-score. Different data pre-processing techniques are applied to improve the quality of data. Generally, Precision, Recall, Accuracy, and F1-score are used for assessing the classifier's performance. High precision shows that the algorithm returns more relevant results than irrelevant ones, while a high recall indicates it returns the great majority of relevant ones.

After preprocessing and feature selection, standard datasets are obtained and apply Decision Tree, Random Forest, Logistic Regression, Support Vector Machine with 'RBF' kernel, Perceptron, Multilayer Perceptron, KNN, Gaussian Naive Bayes, Gradient Boosting Classifier, XGBoost, and Adaboost models to the dataset.

All categorical attributes are encoded using 1,2,3, and so on based on how many distinct values an attribute have. The F1-Score of these algorithms is given in Table 2.

Table 2. Classification F1-Score of different machine learning algorithms

Algorithm	F1-Score
Logistic Regression	0.74
Decision Tree Classifier	0.79
Random Forest Classifier	0.79
Support vector machine with 'RBF' kernel	0.52
Perceptron	0.48
Multilayer Perceptron	0.71
KNN	0.63
Gaussian Naïve Bayes	0.81
Gradient Boosting Classifier	0.76
XGBoost Classifier	0.82
AdaBoost	0.72

From the above table, it is evident that XGBoost Classifier has the highest F1-score of 0.82.

6. CONCLUSION

This paper focuses on the early prediction of student performance. The machine learning models developed may help the students and teachers in taking necessary action before final exams. A review of recent literature on 'students performance prediction systems' revealed that ML algorithms like Naïve Bayes, Random Forest, Decision Tree, Perceptron, Support Vector Machines, Multilayer Perceptron, AdaBoost, Gradient Boosting algorithm, Logistic Regression, XGBoost, and KNN are applied for predicting student's academic performance. None of the papers that are reviewed in this paper applied all these algorithms on a single dataset. To test the performance of these algorithms on a single dataset,

these algorithms are applied to Kalboard 360 dataset. The results show that XGBoost Classifier performance is better with an F1-Score of 0.82 when compared to other classification algorithms.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Ajibade, S. S. M., Ahmad, N. B., & Shamsuddin, S. M. (2019, August). An Heuristic Feature Selection Algorithm to Evaluate Academic Performance of Students. In 2019 IEEE 10th Control and System Graduate Research Colloquium (ICSGRC) (pp. 110-114). IEEE.

[2] Burman, I., & Som, S. (2019, February). Predicting students academic performance using support vector machine. In 2019 Amity International Conference on Artificial Intelligence (AICAI) (pp. 756-759). IEEE.

[3] Caddell, J. D., & Newell, D. (2019, April). Evaluating Teacher Impact on Student Performance: A Case Study at the United States Military Academy. In 2019 IEEE International Systems Conference (SysCon) (pp. 1-5). IEEE.

[4] Dengen, N., Budiman, E., Wati, M., & Hairah, U. (2018, November). Student Academic Evaluation using Naïve Bayes Classifier Algorithm. In 2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT) (pp. 104-107). IEEE.

[5] Denny, J., Rubeena, M. M., & Denny, J. K. (2019, February). A Noval Approach For Predicting The Academic Performance Of Student. In 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT) (pp. 1-5). IEEE.

[6] Devasia, T., Vinushree, T. P., & Hegde, V. (2016, March). Prediction of students performance using Educational Data Mining. In 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE) (pp. 91-95). IEEE.

[7] 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.

[8] 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.

[9] Mishra, T., Kumar, D., & Gupta, S. (2014, February). Mining students' data for prediction performance. In 2014 Fourth International Conference on Advanced Computing & Communication Technologies (pp. 255-262). IEEE.

[10] Mutanu, L., & Machoka, P. (2019, August). Enhancing Computer Students' Academic Performance through Predictive Modelling-A Proactive Approach. In 2019 14th International Conference on Computer Science & Education (ICCSE) (pp. 97-102). IEEE.

[11] Roy, S., & Garg, A. (2017, October). Predicting academic performance of student using classification techniques. In 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON) (pp. 568-572). IEEE.

[12] (Student Performance DataSet, 2019) https://www.kaggle.com /aljarah/xAPI-Edu-Data

[13] Uskov, V. L., Bakken, J. P., Byerly, A., & Shah, A. (2019, April). Machine Learning-based Predictive Analytics of Student Academic Performance in STEM Education. In 2019 IEEE Global Engineering Education Conference (EDUCON) (pp. 1370-1376). IEEE.

[14] Wati, M., Novirasari, N., & Budiman, E. (2018, October). Multi-Criteria Decision-Making for Evaluation of Student Academic Performance Based on Objective Weights. In 2018 Third International Conference on Informatics and Computing (ICIC) (pp. 1-5). IEEE.

[15] Zhang, X., Xue, R., Liu, B., Lu, W., & Zhang, Y. (2018, July). Grade Prediction of Student Academic Performance with Multiple Classification Models. In 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) (pp. 1086-1090). IEEE.