I.J. Modern Education and Computer Science, 2021, 1, 1-21
Published Online February 2021 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2021.01.01

Comparative Study of Supervised Algorithms for Prediction of Students' Performance

Madhuri T. Sathe
Department of Computer Engineering, Rajarambapu Institute of Technology, Rajaramnagar, MS, India
Email: madhurisathe94@gmail.com

Amol C. Adamuthe
Department of Computer Science and Information Technology, Rajarambapu Institute of Technology, Rajaramnagar, MS, India
Email: amol.adamuthe@gmail.com

Received: 08 May 2020; Accepted: 28 June 2020; Published: 08 February 2021

Abstract: Predicting academic performance of the student is crucial task as it depends on various factors. To perform such predictions the machine learning and data mining algorithms are useful. This paper presents investigation of application of C5.0, J48, CART, Naïve Bayes (NB), K-Nearest Neighbour (KNN), Random Forest and Support Vector Machine for prediction of students' performance. Three datasets from school level, college level and e-learning platform with varying input parameters are considered for comparison between C5.0, NB, J48, Multilayer Perceptron (MLP), PART, Random Forest, BayesNet, and Artificial Neural Network (ANN). Paper presents comparative results of C5.0, J48, CART, NB, KNN, Random forest and SVM on changing tuning parameters. The performance of these techniques is tested on three different datasets. Results show that the performances of Random forest and C5.0 are better than J48, CART, NB, KNN, and SVM.

Index Terms: Educational data mining, Machine learning, Random forest, C5.0.

1. Introduction

It is essential for every educational organization to facilitate high quality education to their students. Performance of student in academic is major concern for every institute as it linked to job opportunities and reputation of institution. One of field related to dealing with processing and analyzing of all educational data is educational data mining (EDM). EDM develops methods to understand student and their environment of learning [1]. It also helps to predict patterns that can be helpful to improvement of student performance. Prediction of student's academic performance is a difficult task because it depends on various demographic, socio-economic and past-academic factors. In this paper, the attributes responsible for affecting the academic performance of the student and the students' grades for three different datasets are determined.

In literature, different data mining (DM) algorithms and machine learning (ML) algorithms have experimented for this problem. Machine learning algorithms as said 'learn' from given data, discover hidden patterns and provide predictions, which allow engineers, researchers and scientists to make a reliable decision. Machine learning is broadly divided as supervised, unsupervised, and reinforcement learning. In supervised learning is done using training data which is analyses and builds model to perform predictions for training set. Classes or target variable is labelled in this case. DM and ML techniques are widely applied in field of analytics and predictions. The research work in [2-8] make use of such algorithms which are Logistic Regression, J48, Decision Tree, Support Vector Machine (SVM), NB, Random Tree, ANN, K-Nearest Neighbour (KNN), MLP, and Random Forest. In few cases, other algorithms are also used such as association rules [9], and clustering [10]. NB Tree is used in [11] for predicting status of student, length of study and GPA. Techniques such as REP Tree, PART, Decision Table, Decision Stump, and JRip [12-13] are used for student performance prediction. Results show that the algorithms that perform well for predicting grades are Random Forest, J48, CART, NB, KNN and SVM. To the best of our knowledge, the experimentations are not conducted on C5.0, which is an advanced version of C4.5 (also called as J48) for predicting grades of students. Hence, these seven algorithms are used for grade prediction.

Attribute selection is critical task in every DM and ML algorithm. Performance of algorithm depends on type of data it consists. Adding and removing of certain attributes can also change the performance of algorithm. For educational researches, data can be demographic, academic, and behavioural data. Most of cases demographic and academic data is used for student performance prediction.

- Factors such as age, gender, annual income, parents' occupation, parents' education, are included [2, 4].
- Academic data such as subject marks, previous examination marks are included in [2, 14],
- Previous semester marks, participation in activities are included in [3, 12],
- GPA, subject marks and assignment marks are included in [11, 15]
- Exam scores, absences and attendance is used in [2, 7].
- Consideration of behavioural data is done in dataset of Kalboard 360 available on Kaggle used in [5, 7].
- Comment based data is used in [16] for every lesson taught and various attributes are retrieved using text classification on those comments and [17] comprehends questionnaire based input variables.

In the literature, many machine learning algorithms tested for the problem. In most of cases in related work, the input parameters that are related target are not identified, such as in [2-3, 5]. Also, it is found out that the algorithms those are applied in previous work [7-8, 28, 40] are processed without fine tuning which doesn't pushes limits to know how far an algorithm can accurately predict results.

The research paper presents application machine learning on three different datasets. Performance of algorithm is analyzed by changing values of parameters. Results of decision tree are compared with other algorithms in literature.

The major research objectives of this paper are,

1. To find attributes having more influence on target variable using correlation. By using correlation coefficient it is expected that the attributes that are closely related to grades i.e. those who impact on academic performance most are to be identified.
2. To apply C5.0, J48, NB, Random forest, KNN, SVM, and CART. Using comparison of results of these seven algorithms we determine which of the algorithms can accurately predict the results when fine tuning is applied.
3. To study effect of tuning parameters on accuracy of classifiers.

Evaluation of these results are done using various measures such as precision, recall, True Positive Rate (TPR) and False Positive Rate (FPR).

Sections below are divided into 6 parts. Literature review on student performance prediction is described in section II. Section III contains problem formulation for current work and previous work. Methodologies used are mentioned in section IV. Results are discussed in section V and section VI explains conclusion.

2. Related Work

Research for predicting students' academic performance has been done for various kinds of datasets and using numerous methods. Datasets can be of type e-learning, university data, college data and the variety of methods applied are statistical, data mining techniques and machine learning algorithms.

A review of the various algorithms used and their accuracies obtained to solve student performance prediction presented below. In [3] algorithms such as Logistic Regression, SVM, Decision Tree, NB, NN, and KNN are applied. Experimentation is conducted by considering all input variables and attributes filtered through feature selection. Performance of KNN is found out to be better when all attributes are considered. SVM and Logistic Regression perform well for dataset with feature selection process. Conversely in the work mentioned in [18] has that Logistic Regression performs poor along with NB, but KNN has better accuracy results. The dataset in [3] and [18] differs in terms of number of records as well as [3] has previous academic data and [18] has course grades along with psychometric factors. In [19] using e-learning data and behavioural survey of student using such e-learning platform Logistic Regression model is applied to predict the failure of student in a particular course. Dataset considered is presumably small on which accuracy obtained is 73.7%. Logistic Regression, NN and Random Forest are used in [20]. All of these techniques provide low level of correctness in results, such weakness is overcome by inclusion of uncertain classes. Similar to [19], work in [21] follows same algorithm with almost same amount of records having accuracies of 78.6 % and 78.8%. In addition with algorithms used in [3], the work proposed in [22] has BayesNet and SMO where prediction of failure of a student in particular course is performed. Experimentations are conducted by taking into consideration dataset with and without filtering, discretization and rebalancing. Without filters performance of algorithms is lower than expected. With applying filters, most of algorithms has enhancement in their performance. Decision Trees, Random Tree and Random Forest have been giving best potential. In [23] the 3rd semester performance is carried out using Decision Tree and Random Tree. Results show that RT achieves 94.4% accuracy followed by J48 with 88.37%. Predictions are achieved high in [12], their work also convey that data mining techniques are not limited by size of datasets. [24] has multiclass classification performed by using algorithms such as RF, DT, SVM, NB, Boosting Trees and Bagging Trees. Above 2000 student records are considered for the prediction process. Their work focuses on obtaining results for degree level performance of students. The Random Forest achieves an accuracy of 96.17% which is best among the other algorithms implemented. In cases [2, 5, 24], Random Forest, J48, and NB tend to predict the results more accurately. For SVM, the results achieved are remarkable, as it has maximum F1-measure value after DT [4]. SVM and KNN both are to be found suitable for student academic predictions as mentioned in [6]. The algorithms CART and C5.0 are rarely used as per the best of our knowledge. Literature survey conveys that trees and Random Forests perform best for classification of grades of students.

The following literature discusses different kinds of outcomes predicted using ML algorithms. Most of the cases Grade Point Average (GPA) or cumulative GPA is predicted. The outcome can be either binary class such as pass or fail in particular subject or semester, successful or unsuccessful to complete graduation or degree otherwise target attribute can be multiclass such as in [8] end semester percentage is converted to five classes which are best, very good, good, pass and fail. The research work in [3, 4, 19, 21, 22] focus on acquiring predictions for a particular course. Classification performed is binary class, student will pass in the course or get failed is predicted. Whereas, [18] predict that whether student performs poor in academics or is a strong achiever, based on GPA values. Final GPA are determined in [15] by using demographic data, high school information, and family financial status. The research work in [25] predicts that whether a student will obtain his engineering degree or not using student academic data and background information and same with the case of [20]. Unlike these research papers the work mentioned in [26] predicts whether a student will obtain excellent grade, good grades, get passed or just get failed in a course with grades predicted to be from scale of 0 to 10 for final exam of a course. Similar to this, in [23] it is predicted either student will get 3rd semester performance as below average, average, above average or excellent. In [12] predicts the score of a course of students to be low, high or medium.

Impact of different attributes on the performance of student performance is reviewed and presented. Most of the research works have included correlation methods to find out the influencing factors. Models those are trained on students of less age provided good results in [18]. The type of registration to University and income of student's family are found to be correlated achievement of student. The four major attributes that highly affect the performance of student found in [25] are First Year University GPA, CC BP transfer credit hours, first fall credits GPA and CC BP transfer GPA. Here all the attributes related to academic and student's background were considered. The work in [21] where e-learning data is taken for process, the significant predictors are found out to be date of first login to LMS, mode of study, previous academic performance record, and weighted average marks. For the results in [23], where 3rd semester performance is predicted, it is revealed that 2nd semester results, leadership and drive qualities correlate a lot with output variable. For [27], the ability to understand and handle basic subjects influence a lot on final result of degree. In [28], experimentations are conducted using dataset from UCI machine learning repository. On comparison with target variable it is found out that weekday alcohol consumption, romantic relationship and parents' education do affect student's performance. GPA, Participation rule, Test average, Lab test average, Assignment submit attendance, Final grade are considered as best attributes in [29] for undergraduate student data. In [30] where techniques such as Cluster Analysis and Association Rule Mining were used found a pattern that frequent occurrence of seven courses {MTH 111, STA 122, MTH 122, MTH 121, CSC 111, BIO 111, CSC 121} in failed students' data. These courses are found crucial for academic performance of a student. Survey presented in [31] for student performance concludes that CGPA and internal marks are important attributes. Conversely, evaluations done in [32] mention that grades do not necessarily affect outcome achievement and direct assessment has positive impact on student performance.

Table 1. Literature review

References	Objectives	Techniques	Types of attributes in dataset	Results
[2]	To use different machine learning algorithms for studying student performance.	J48, Naïve Bayes (NB), Random Forest (RF)	Demographic (age, gender) and academic parameters (project marks, midterm marks, quiz marks)	RF provides 100% accuracy in predicting appropriate student remarks
[3]	To find best prediction method to identify students at-risk.	Logistic Regression(LR), Support Vector Machine (SVM), Decision Tree (DT), Multilayer Perceptron (MLP), NB, K-Nearest Neighbour (KNN), Ensemble	Course related data (home works, quizzes).	Ensemble methods provides simple model than other models and gives best accuracy (84.6%)
[4]	To compare effectiveness of existing EDM techniques to early identify the students those are likely to fail.	NB, DT, Neural Network (NN), SVM	Demographic, e-learning (usage of educational tool, quiz, messages), and enrolment data (enrollment year)	Techniques analyzed in the study fulfill the research objective. SVM provides best results.
[5]	To improve the quality of education universities by predicting academic performances of students.	NB, DT, SVM, MLP	Three semester behavioural data	Pattern is found in the student data that current result is dependent on previous result. DT provides better accuracy than other methods.

[6]	To predict students' grades To find which of technique is suitable for regression	SVM, KNN	Demographic data (family income, parents' education), academic data (First year grade, Second year grade)	Both SVM and KNN are found to be suitable for regression
[7]	To study impact of behavioural features on academic performance of students.	DT, Artificial Neural Network (ANN), NB	Demographic data, behavioural data (visited resources, raised hands)	Accuracy to predict student class increases when behavioural features are included
[8]	To find which feature has highest impact on target class To find which technique outperforms most	RF, PART, J48, BayesNet	Socio-economic, demographic and academic information	RF provided 99% accuracy. Internal assessment impacts on final semester percentage
[11]	To build NBTree model for student performance prediction	NBTree	Personal data, academic data, admission data, education data	Gender attribute is found to be influencing on university level active student and credits for graduates. Faculty level active student have GPA as influencing factor where graduates find test score influencing.
[12]	To analyze student's data using data mining approaches with perspective to answer various HEIs questions	REPTree, J48, M5P	Personal data, academic data, registration data	REPTree is less sensitive for data having missing values hence provides better results
[13]	To build a framework for intelligent recommender system to predict student first year performance and recommend appropriate suggestions	PART, OneR, Decision Table, JRip, REPTree, J48, Random Tree (RT), Decision Stump, RF, MLP	Demographic data, education data	RT outperforms by giving 99.90% accuracy in 10-fold cross validation, and 99.82% in hold-out method
[14]	To predict performance of student in an engineering dynamic course by developing mathematical model and identify most appropriate model.	Multiple Linear Regression (MLR), MLP, Radial Basis Function (RBF), SVM	Course related data, GPA	SVM provides highest percentage of accurate predictions.
[15]	To determine student characteristics that are associated with their academic success	Microsoft Decision Tree Classifier	Registration data, academic data, financial status	For 1st model the type of registration has impacted the performance of student whereas for 2nd model monthly income of family impacts most.
[16]	To collect the comments that show student learning ability To find the factors that influence students' learning and build a student performance prediction model	DT, RF	Comment based data	RF gives best results. When extraction rules are applied the F-measure value increases for both algorithms.
[18]	To investigate the accuracies provided by models that predict students at risk in failing in first year college.	NB, Unpruned DT, LR, SVM, KNN, NN	Course related data, behavior related data	When induction is applied only KNN has increment in accuracy up to 100% vice versa for other methods. Self-efficacy, achievement motivation are best predictors when correlated with GPA.
[20]	To detect early detect students those are likely to fail To design algorithms for such predictions To analyze the results	LR, ANN, DT, RF	Past performance data and environmental factors	Accuracy is improved compared to their previous work as additional class known as Uncertain Class is included. RF performs best.
[24]	To early classify students into segments	RF, DT, SVM, NB, Bagged trees, Boosted trees	Socio-demographic data, socio-economic data, high-school background data, enrollment data	Enrollment average grade and average grade in 1st semester highest attribute importance. RF gives highest results of 96.1% accuracy.
[25]	To develop boosted Logistic Regression model for prediction graduation in engineering for student those are transferred from community college To compare statistics of this model with actual graduation rates To provide report on student academic variables that can be helpful to increase chances of successfully graduating in engineering college	Boosted LR	Academic, demographic and student background data	On an average of 35% difference in graduation rate statistics between observed and predicted are seen. Highest influencing attribute is found out to be first year GPA having over 39.5% variance inflation factor.

[27]	To predict the performance of bachelor degree engineering students based on matriculation and diploma level entries	NN	Subject grades, CGPA of semester 3	Results convey that when actual CGPA8 is low, the predicted value is higher than actual vice versa for higher CGPA8 actual value. The MSE obtained by student having entry with Matriculation is less than that of Diploma students
[28]	To determine the factors that influence the grades of student. To perform prediction of grades of students using various data mining algorithms	NB, MLP, J48	Demographic data, socio-economic data, academic data	J48 performs best with accuracy 73.92% Influencing factors are health, romantic relationship, parents' education and alcohol consumption.
[29]	To apply data mining techniques to predict and analyze students' academic performance using academic and forum data	NB, NN, DT	Demographic data, forum related data, academic data	NB performs best with 86% accuracy Best attributes determined are GPA, Test avg, Assignment submit, participation rate, attendance rate.
[32]	To identify relationship between courses provided to students To understand effects of these relations with learning and academic performance	Hierarchical Clustering, Association Rule Mining	Academic data, CGPA	RW students have highest average Relative Frequency (RF) followed by SPO students and second class students have low average RF.

3. Problem Defination

Classification maps data into predefined groups or classes. It is also referred to as a supervised learning technique because the classes are determined before examining data. It is used to predict the class for new datasets based on trained datasets. Classes are defined based on attribute values. Classification makes required data easy to find and retrieve. It includes two-phase which are building a model that includes training data based on which model is prepared and the using classification model, in which testing data is used to estimate the accuracy of classification rules. These rules can be applied to new data tuples if the accuracy is acceptable.

Predicting the performance of the student is a crucial task as it depends on various factors. Academic performance can be predicted in terms of grade. In the current study, students are classified into one of their classes i.e. grades predicted for them. Using such a supervised learning process the grades of students are predicted for three different datasets. The independent variables are fed as input to the classifiers for prediction of outcome i.e. dependent variable. Independent variables are demographic, socioeconomic, academic and behavioural factors and the outcomes are multiclass grades in the current context.

Dataset 1, which is acquired from the UCI Machine Learning Repository [33]. It is school students' records consisting of numerous demographic attributes such as age, gender, parents' education, parents' job, family income along with academic data such as first-year and second-year grades. These attribute act input for a classifier that predicts the appropriate class for a student which are A, B, C, D and F. Similarly, dataset 2 is also collected from UCI Machine Learning Repository [34] which a college student data. It has demographic data similar to dataset 1, academic data such as 10th percentage, 12th percentage and internal assessment. Output class (End semester percentage) has remarks such as Excellent, Very good, Good, Pass and Fail. Lastly, dataset 3 is based on e-learning information collected from Kaggle [35]. It contains demographic, past academic as well as behavioural data (parents' answer survey, visited resources, raised hands, discussion). For this dataset, the students are classified into one of three classes, High, Low or Medium.

For all of above research, DM and ML techniques are applied for predictions. Algorithms used are Naïve Bayes, Multilayer Perceptron, J48 for dataset in [28], Random Forest, PART, J48 and BayesNet for [8] and Decision Tree, ANN, and NB in [7].

4. Methodology

In research work mentioned in [8, 28] datasets are collected, on which classifier models are applied and results are evaluated. Whereas, in [4-7] certain feature selection techniques are applied before the data classification along with fine tuning is done. Based on these methodologies the proposed methodology is constructed and described as follows:

A. Data Collection

Data is collected from UCI ML Repository and Kaggle. Two of student academic datasets are from UCI Machine Learning. The first dataset contains 395 records of students with 33 attributes [33].The second UCI Machine Learning dataset contains 133 student records with 22 features [34]. The third dataset is collected from Kaggle [35] which consists of 16 features and 480 student records.

B. Methodology

Fig. 1 presents proposed methodology for student performance prediction. The methodology is applied to all three datasets.

Step number two is conversion of values. Input parameters are demographic and academic data. Output is class consisting respective grade categories. The dataset is imported and converted to numeric format in order to apply Pearson correlation between all input attributes to target attribute [36]. The attributes having correlation less than 0.05 are eliminated. Attributes with high correlation are used for further processing. Step number 4 is application of classification technique. The J48 and C5.0 are applied in selected attributes and predictions are performed. Various evaluation measures such as precision, recall, FPR (False Positive Rate), TPR (True Positive Rate) are used for analysis.

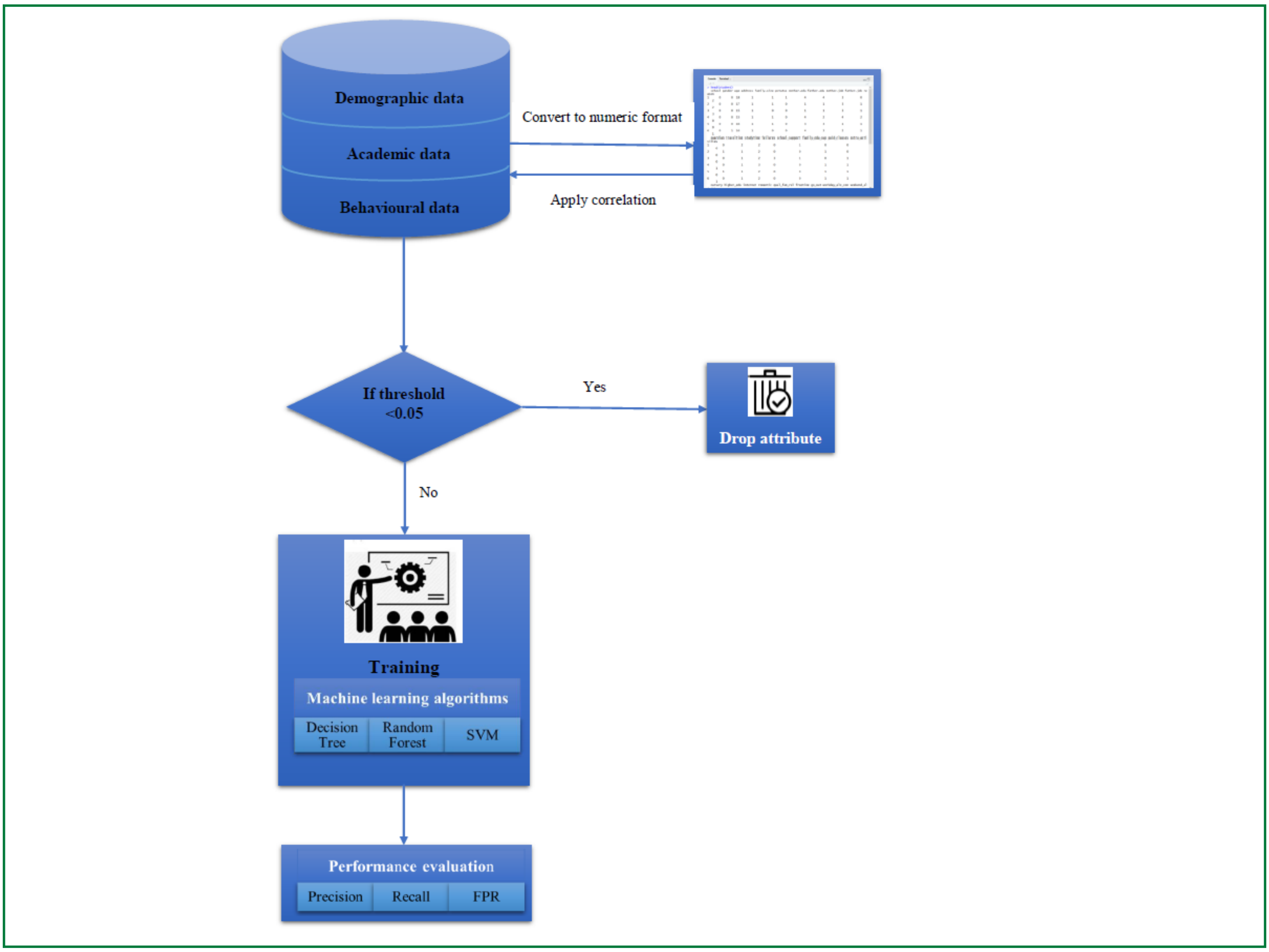

Fig. 1. Proposed methodology for student performance prediction

Fig. 2 presents model for prediction system for which the attributes with high correlation with target variable are chosen as input. Decision tree algorithm is applied on those selected attributes as well as tuning parameters are altered which gives result to classification of students marks in their appropriate grades.

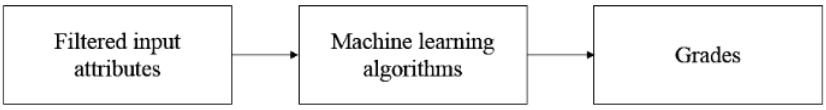

Fig. 2. Model for student performance prediction

The strategy applied for previous work [7] includes data collection, data preprocessing, data discretization, data cleaning followed by applying feature selection techniques such as feature ranking. These filtered students records are then classified to appropriate classes. Various evaluation measures are applied to results acquired by techniques used in [7]. Whereas in [8], a dataset is directly provided as input to classifiers in WEKA and various feature selection methods are applied such as gain-ratio attribute, information- gain attribute evaluation. Results produced by supervised learning methods are evaluated using accuracy, Mean Absolute Error (MAE), Root Mean Square Error (RMSE) etc. The research work conducted in [28] doesn't have preprocessed data, the algorithms are straightaway applied on the whole dataset to obtain accurate results. The performance influencing attributes are selected using analysis performed on the graphs plotted between input variables and target attribute. It can be concluded that determining the correlation between all input attributes and target attribute and fine-tuning of algorithms are not performed in methodologies utilized in [7, 8, 28] which are included in current work.

C. Pearson Correlation

Correlation is measure of strength between two variables [36]. Its value lies between "-1" to "+1". '+' indicated positive relationship and '–' indicates negative relationship. *'0'* indicates no relationship between two variables. In current system, Pearson correlation is used to find relationship between target variable and other variables. Pearson correlation is calculated as follows [37].

$$r = \frac{N\sum xy - \sum x \sum y}{\sqrt{N\sum x^2 - (\sum x^2)}\sqrt{N\sum y^2 - (\sum y^2)}} \tag{1}$$

r = Pearson r correlation coefficient
N = number of observations
Σxy = sum of the products of paired scores

D. Machine Learning Algorithms

C5.0

C5.0 is improvement of C4.5 which in terms of boosting process. The J48 is java implementation of C4.5. R language provides tuning parameters for C5.0 which are trials and Confidence Factor (CF). Value of CF lies between 0 to 1. Number of trials represents number of boosting iterations to be performed. Value 1 indicates only one model is used. CF controls the post-pruning of decision tree. Decrease in value of CF leads to decrease in amount of post-pruning [38].

J48

J48 is java implementation of Quinlan's C4.5 algorithm. It uses entropy to calculate information gain [4]. To select appropriate features for constructing a tree, information gain is calculated for each feature and the attribute with highest value is chosen to be root node. Recursively parent nodes are created and process is stopped when a node has all of its instances belonging to same class. Advantage of J48 is that it can be used for both categorical and continuous variables and can handle missing data [1]. Tuning parameters used are reduced pruned tree, unpruned tree and binary splits.

CART

Classification and Regression Trees (CART) are introduced by Breiman [36]. They can handle both continuous and categorical data. CART uses Gini index for attribute selection. It used cost complexity for pruning of tree. CART can handle missing data. It creates binary tree i.e. it is made up of at most two nodes. CART does not rely on distribution of data. Outliers have less impact on CART [39]. The tuning parameter used is feature selection method which is either information gain or gini index.

Naïve Bayes

Naïve Bayes is based on Bayes theorem. It assumes that features are independent of each other. Bayes theorem is stated as follows:

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)} \tag{2}$$

P (x) - is the prior probability of predictor.
P (c) - is the prior probability of class.
P (x|c) - is the likelihood which is the probability of predictor given class.
P (c|x) - is the posterior probability of class (c, target) given predictor (x, attribute)

Using prior probability of given class, posterior probability is predicted. If the probability of certain class is found out to be high then that instance of record belongs to respective class with high probability. Advantage of NB is it requires short time for training. It requires large amount of dataset to produce more accurate results. Tuning parameter used is number of Laplace.

K Nearest Neighbour

KNN calculates distance between all instances and searches for K instances which are having minimum distance. The majority of class present in k instances is applied to the test instance. It is simplest algorithm of all other machine learning algorithms [40]. KNN is easy to understand and implement. It works well when numbers of classes are more. Although KNN has ease of understanding it is called as lazy learner and has high memory cost [40].

Random Forest

Random forest is made up of many decision trees. Each tree learns from randomly selected sample of data. Samples are drawn with help of bootstrapping process. The predictions are made by taking an average of predictions of each tree. Random forest does not selects all of the features available in dataset. In the end votes for each target are calculated. The class with highest vote is assigned to the test instance.

Support Vector Machine

SVM plots each item in dimensional space. For two input attributes it will be two dimensional spaces. A hyper plane is placed such that for binary class, its separates instances into two classes. The distance between nearest instances from both classes should not be far from hyper plane. SVM gives better performance for binary classification problems. Kernel used in SVM can be linear, radial basis, sigmoid or polynomial.

E. Performance evaluation metrics

Confusion Matrix

A confusion matrix has actual values versus predicted values. Assuming confusion matrix for binary class which is either 'Yes' or 'No' representation will be as follows:

		Predicted	
		Yes	**No**
Actual	**Yes**	TP	FN
	No	FP	TN

TP (True Positive) – When predicted class and actual class both are Yes.
FN (False Negative) – When class is predicted to be No and has actual value as Yes.
FP (False Positive) – When predicted class is Yes and has actual value as No.
TN (True Negative) – When both predicted and actual values are No

Classification accuracy

Accuracy is calculated by taking sum of correctly classified instances divided by total number of instances.

$$Accuracy = \frac{Correctly\ classified\ instances}{Total\ instances} \tag{3}$$

Precision

Precision is True Positives divided by sum of True Positives and False Positives shown as follows [41].

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

It describes correctness of a classifier. It defines proportion of accurately categorizing instances. Precision is useful when FP is high.

Recall

It is also called as sensitivity, and shows probability of detection of relevant instances in dataset. It is calculated as follows:

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

True Positive Rate (TPR)

It is same as recall having same formula for calculation. It helpful when there is high cost associated.

False Positive Rate (FPR)

It is also called as Fall-out and identifies proportion of incorrectly classified instances. It is probability that instances will be predicted wrongly. It has following formula:

$$FPR = \frac{FP}{FP + TN} \tag{6}$$

5. Experimental Details, Results and Discussion

This section presents experimental details on varying datasets, obtained results and discussion. To test the performance of the proposed methodology, three datasets are used.

A. Results of Pearson correlation

Dataset 1 has 395 records along with 33 attributes [33]. These data were collected from two Portuguese schools to predict the target variable which is final marks of Math subject G3. The G3 marks were converted into 5 grades. 12 attributes were selected for further processing after applying Pearson Correlation between input attributes and target output. Selected attributes are shown in Table 2 along with their correlation coefficient values against target output 'grade'. It is clear that G1 and G2 show highest influence on target variable i.e. grade. Other high influencing factors are mother's education, father's education, and interest for higher education.

Table 2. Attributes with their Pearson correlation coefficient values for Dataset 1

Attribute	Pearson correlation coefficient value
Gender	0.11
Address	0.10
Mother's Education	0.22
Father's education	0.15
Mother's job	0.05
Study time	0.10
Higher education	0.15
Internet access	0.09
Family relationship	0.05
G1 score	0.86
G2 score	0.87

Dataset 2 is from UCI Machine Learning Repository has 131 student records along with 22 attributes [34]. Dataset was made from three different Indian colleges. 13 attributes were selected after applying Pearson Correlation with respect to target output. These selected attributes are shown in Table 3 along with correlation coefficient value. The attributes with the highest correlation with target attribute i.e. end semester percentage are 10th percentage, 12th percentage, internal assessment, and attendance percentage.

Table 3. Selected attributes and their Pearson correlation coefficient value for Dataset 2

Attribute name	Pearson correlation coefficient value
10^{th} percentage	0.66
12^{th} percentage	0.63
Internal assessment percentage	0.54
Where student lives	0.11
Admission category	0.40
Income	0.14
Father's qualification	0.33
Mother's qualification	0.29
Father's occupation	0.05
Number of friends	0.14
Study hours	0.20
Medium	0.32
Attendance percentage	0.44

The third dataset is from Kalboard 360 having 480 records of students and 16 features [36]. This dataset was used by authors Amrieh, E. A., Hamtini, T., & Aljarah, I. (2015, November) [7] in which 13 attributes are choosen from Kalboard 360 dataset and additional an attribute named as teacher id.

Table 4 shows the features selected after Pearson correlation is applied along with their correlation coefficient values. Number of hands raised, visited resources have highest correlation to target class compared to other attributes. Attributes having high correlation are announcement view, relation with parents, number of discussions done, parent's school satisfaction and parents' answering survey.

Table 4. Attributes and their correlation values for Dataset 3

Attribute name	Correlation value
Nationality	0.1855
Place of birth	0.1831
Educational stages	0.0839
Grade levels	0.0672
Topic	0.1653
Relation	0.4011
Raised hands	0.6462
Visited resources	0.6770
Announcement view	0.5273
Discussion	0.3081
Parent answering survey	0.4354
Parent school satisfaction	0.3759

B. Results for C5.0

Using filtered attributes decision tree model was built using library C50.

For C5.0 there are control parameter such as number of trials and confidence factor (CF). Fig. 3-5 show effect of CF on accuracy of decision tree for three datasets. As the CF increases the accuracy increases rapidly and stays constant after certain values.

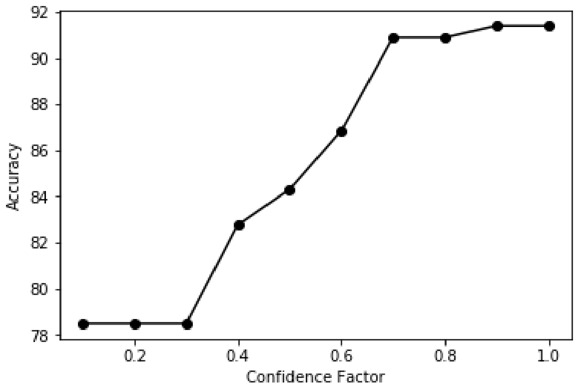

Fig. 3. Behaviour of accuracy of C5.0 with respect to Confidence Factor for Dataset 1

For dataset 1 accuracy stays constant up to value of CF = 0.3. There is sudden rise in accuracy till CF = 0.7 and rises gradually till end. The difference in accuracy observed at the initial point and endpoint is around 13.5%

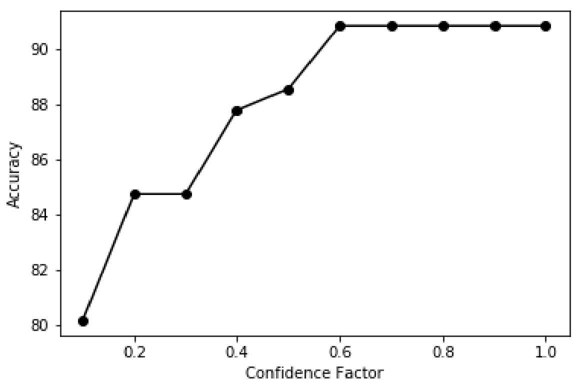

Fig. 4. Behaviour of accuracy of C5.0 with respect to Confidence Factorfor Dataset 2

Here, the nature of the graph containing accuracies, is like a stair step. The accuracy stays unchanged from CF = 0.6 till the end and the accuracy elevation is around 10%.

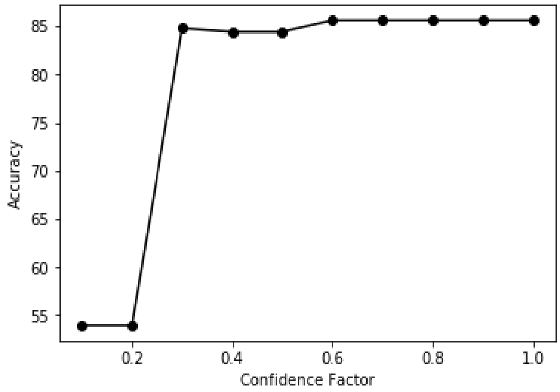

Fig. 5. Behaviour of accuracy of C5.0 with respect to Confidence Factor for Dataset 3

A sudden increase in accuracy is observed in Fig. 3 for values from CF = 0.2 to 0.3. Later, there is small amount of gradual rise after CF =0.3 till CF =1. Majority difference in the accuracies for initial and final values of CF are observed in dataset 3 having 30%.

It is clear from these figures that for these three datasets the accuracy is at its peak when CF reaches approximately at value 0.6.

Fig. 6-8 show how accuracy improves by incrementing boosting trials. Behaviour for trials is similar to that of CF, here as number of boosting trials are incremented the accuracies raise eventually.

For dataset 1 as shown in Fig. 6 when CF is set 0.9 initially accuracy lies approximately 91% when the number of trials is one, as trials are increased so does the accuracy drastic increase in accuracy of C5.0 is found from trials = 4 to trials = 6. The highest point of accuracy achieved is at trial = 9 with approximately 98.6%.

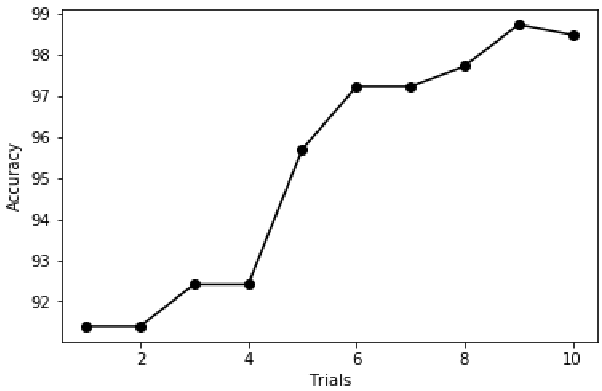

Fig. 6. Behaviour of accuracy of C5.0 with respect to trials when CF = 0.9 for Dataset 1

Concerning Fig. 6 CF for dataset 2 is fixed at 0.6 value and the number of trials from 1 to 10 is tested. Accuracy at trials = 0 is around 90.5% which elevates suddenly at 3rd to 4th trial by 5% as displayed in fig. 9. The highest accuracy is reached when the number of trials is 9 and percentage accuracy lies between 99 and 100. Accuracy drops at trial number 10. The graphs in Fig. 6 and Fig. 7 almost follow the same pattern both datasets have rising accuracy points at a similar number of trials (trial = 3 and trial =4 resp.) and drop at trial number 10.

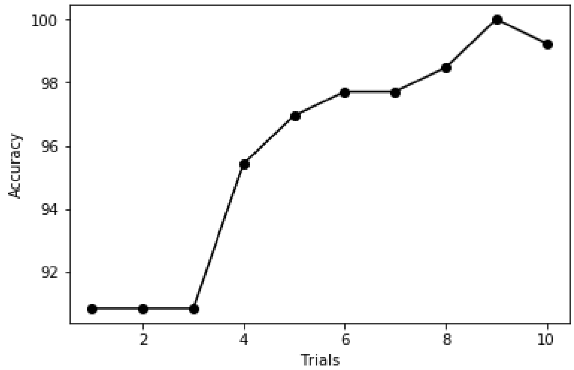

Fig. 7. Behaviour of accuracy of C5.0 with respect to trials when CF = 0.6 for Dataset 2

Fig. 8 shows accuracy behaviour for dataset 3 when CF is set as 0.6 and number of trials from 0 to 10. Similar to graphs for dataset 1 and 2, accuracy doesn't increase till reaches 3 trials. Irregular increase and decrease in the accuracies are observed from 3rd trial till 7th trial which is not in cases of datasets 1 and 2. Similar to Fig. 6 and Fig. 7 there is a gradual increase in accuracy from 7th trial onward and reaches up to 93% in this case.

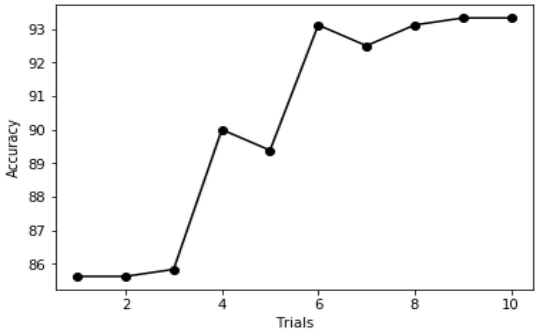

Fig. 8. Behaviour of accuracy of C5.0 with respect to trials when CF = 0.6 for Dataset 3

Table 5 shows comparison of C5.0 for dataset 1 with algorithms in previous work 3 shows a sample output decision tree with good results for student academic dataset.

Table 5. Performance comparison between NB, MLP, J48, proposed J48, and C5.0 for Dataset 1

Algorithm	**Time taken to build model (in sec)**	**Correctly classified instances (in %)**	**Incorrectly classified instances (in %)**
NB[8]	0.03	68.60	31.39
MLP[8]	29.06	51.13	48.86
J48[8]	0.17	73.92	26.07
C5.0	7.74	98.48	1.52

On comparison of these results with algorithms used in research paper [28] the results are shown below in Table 4 which shows comparison between Na ïve Bayes, Multilayer Perceptron, J48 and C5.0. It is clear that C5.0 gives 98.48% accuracy and highest among the algorithms used in previous work [8].

On comparison of these results with algorithms used in research paper [8] the results are shown in Table 6 shows comparison between Random Forest, PART, J48, BayesNet and C5.0. Best performance is achieved by C5.0 followed by Random Forest which acquires 99% accuracy and BayesNet performs poor among these algorithms.

Table 6. Performance comparison between Random Forest, PART, J48, Bayes Net and C5.0 for Dataset 2

Algorithm	**Number of instances**	**Correctly classified instances**	**Accuracy (in %)**
Random forest [28]	300	297	99
PART [28]	300	223	74.33
J48 [28]	300	219	73
BayesNet [28]	300	196	65.33
C5.0	131	131	100

Overall highest accuracy attained in experimentation of [7] is 73.8% by Artificial Neural Network. Using 480 records and applying Pearson correlation to eliminate features decision tree is applied with its control parameters. Comparing those results here accuracy achieved is 93.33% for Confidence Factor 0.6 and 9 boosting trials. Table 7 shows that C5.0 gives best performance

Table 7. Performance of Decision tree, ANN, NB and C5.0 for dataset 3

Algorithm	**Accuracy (in %)**
Decision tree [7]	61.30
ANN [7]	73.80
NB [7]	72.50
C5.0	93.33

C. Results for J48

Results of J48 for three datasets are presented in Table 8. The graphs 10-12 shows that when the number of instances are increased, the accuracy drops. After a certain limit all of dataset is classified to the class which has maximum number of instances. Table 7 shows accuracies achieved using changing tuning parameters of J48 for three datasets. Use of unpruned tree and binary splits enhances the accuracy of J48 and pruning process reduces accuracy.

Table 8. Performance of J48 for various tuning parameters on dataset 1, 2 and 3

J48 tuning Parameter	**Accuracy (in %)**					
	Dataset 1		**Dataset 2**		**Dataset 3**	
	True	False	True	False	True	False
Reduced prune tree	79.75	86.08	71.76	82.44	68.54	74.38
Unpruned tree	94.68	86.08	90.08	82.44	85.62	74.38
Binary splits	94.18	86.08	89.31	82.44	84.58	74.38

For dataset 1 in as shown in Fig.9 observations convey that till minimum instances per leaf are increased up to 50, the decrease in accuracy occurs at slow rate and drops suddenly after that up to 33% and decrement continues slowly after that. Lowes accuracy of 40% is consistent from instances = 70 till the end.

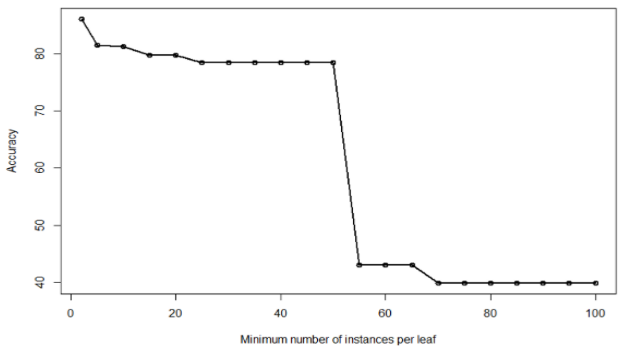

Fig.9. Nature of accuracy of J48 on changing minimum number of instances per leaf for dataset 1 [34]

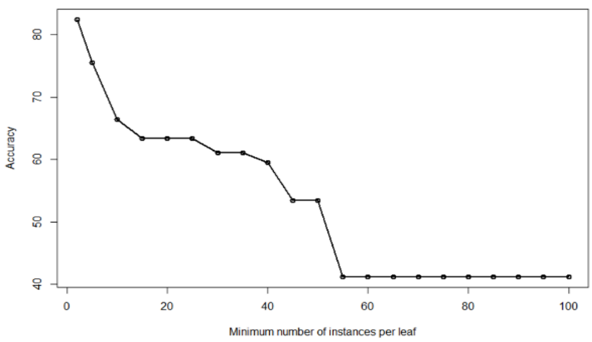

Fig. 10.Nature of accuracy of J48 on changing minimum number of instances per leaf for dataset 2 [35]

In Fig. 10 the accuracy drops fast till the number of instances is 10 and thereafter it falls off in medium rate till instances are set 50 and drops to its lowest accuracy when instances are 60. After that, the accuracy gets frozen at 40% till the end.

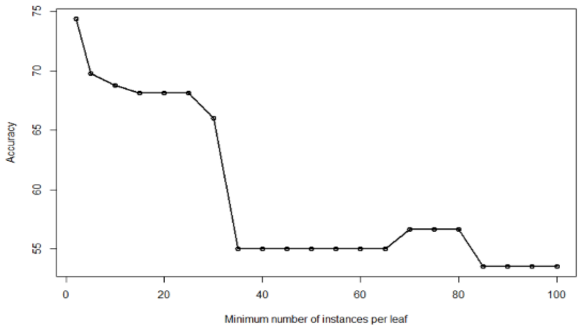

Fig. 11.Nature of accuracy of J48 on changing minimum number of instances per leaf for dataset 3 [36]

From Fig. 11, sudden accuracy drop point occurs from 20 to 30 instances having a difference of 25%. Consistency of 55% accuracy is seen from 50 to 65 instances thereafter, an abnormal increase and decrease occur and the lowest accuracy of 50% stays constant from 85 instances onward. Graphical behaviour of the third dataset is different than previous datasets.

D. Results for CART

Following table presents results for CART algorithm by application of two feature selection techniques. It is seen that using information gain the CART algorithm gives poor performance compared to Gini index for dataset 1 whereas for dataset 2 and 3 information gain gives better performance for CART.

Table 9. CART performance for different feature selection methods for dataset 1, 2 and 3

Splitting index	**Accuracy (in %)**		
	Dataset 1	**Dataset 2**	**Dataset 3**
Information gain	78.48	74.81	80.00
Gini index	79.48	70.99	79.49

E. Results for Naïve Bayes

Naïve Bayes has Laplace as tuning factor. On increasing the number of Laplace the accuracy decreases as shown in Fig. 13-15. Table 10 presents results for Naïve Bayes when Laplace is not used and maximum accuracy attained using Laplace parameter.

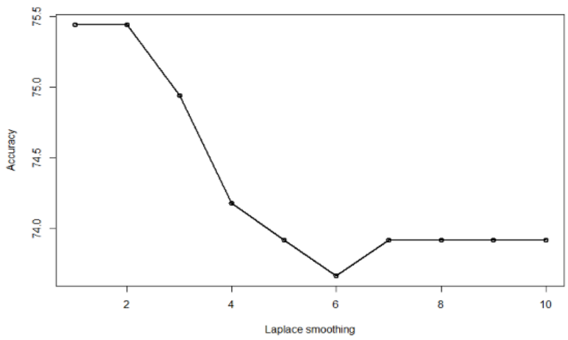

Fig. 12. Nature of performance of Naïve Bayes for different values of Laplace for dataset 1.

Above graph shows that the accuracy of Naïve Bayes drops from 75.4% till 73.9% when Laplace value consecutively increased till 10 for dataset 1. The performance drop at lowest at Laplace = 6, then gets increased and stabilizes when value equals 7.

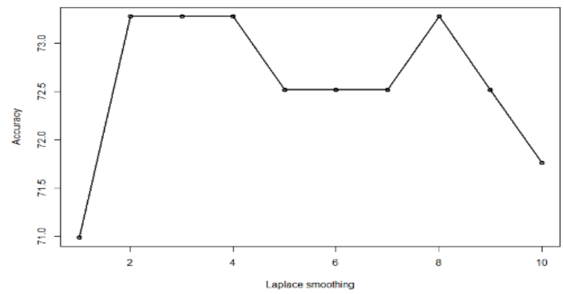

Fig.13. Nature of performance of Naïve Bayes for different values of Laplace for dataset 2.

The graph in Fig. 13 behaves abnormally in nature as there is no predictability about increase or decrease in the accuracies for dataset 2. Accuracy rises at sudden at initial value then has stability when Laplace values are from 2 to 4. Accuracy drops at Laplace = 5 stay same till Laplace =7. Again the increase and decrease are seen at the end. Compared to the graph for dataset 1, Laplace has irregular effects on accuracy for dataset 2.

Fig. 14 shows the effects of Laplace on dataset 3, which much simpler to understand than that of dataset 1 and 2. The accuracy at Laplace = 1 is 82% and gradually drops to 70% with no sudden rise in the process, as observed in Fig. 12 and Fig. 13

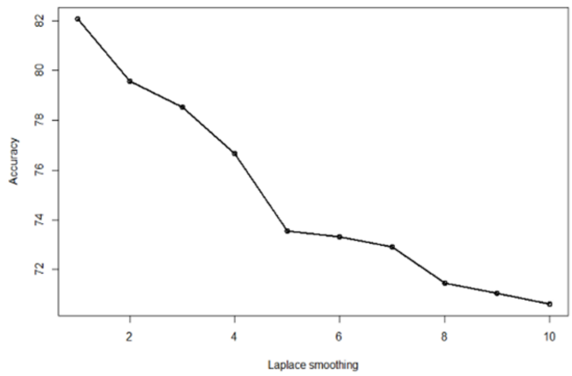

Fig. 14. Nature of performance of Naïve Bayes for different values of Laplace for dataset 3.

Table 10. Performance of Naïve Bayes for dataset 1, 2 and 3

Tuning parameter of Naïve Bayes	Accuracy (in %)		
	Dataset 1	Dataset 2	Dataset 3
Accuracy without Laplace	75.95	72.52	85.21
Accuracy with Laplace	75.44	73.28	82.08

F. Results for KNN

Table 11 presents accuracies of KNN for different values of K. Maximum accuracy is achieved when value of K equals 3. The performance of KNN decreases as value of K is increases.

Table 11. Performance of KNN for different values of K

K	Accuracy (in %)		
	Dataset 1	Dataset 2	Dataset 3
3	82.28	77.86	81.25
5	77.97	67.18	77.92
7	76.20	63.36	73.75

G. Results for Random Forest

Table 12 shows performance of Random forest when numbers of trees are changed. As the number of tree increases so does the accuracy. For dataset 1, the maximum accuracy obtained when the number of trees is 250. Dataset 2 achieves 100% accuracy when the number of trees is 20. Finally, for dataset 3, 100 trees can obtain maximum accuracy of 98.12%.

Table 12. Performance of Random Forest for dataset 1, 2 and 3

Number of trees	Accuracy (in %)		
	Dataset 1	Dataset 2	Dataset 3
1	74.94	83.97	82.92
2	80.25	82.44	84.58
5	95.70	93.13	92.92
10	98.23	98.47	95.83
20	99.24	100	97.29
50	99.24	100	97.92
100	99.49	100	98.12
250	99.75	100	98.12
500	99.75	99.24	98.12

H. Results for SVM

SVM has kernel as tuning parameter. Kernel can be linear, radial basis, sigmoid and polynomial. Table 13 shows results for different kernels for SVM. For each dataset, SVM with linear kernel gives best performance amongst other kernels.

Table 13. Performance of SVM for dataset 1, 2 and 3

Kernel	Accuracy (in %)		
	Dataset 1	Dataset 2	Dataset 3
Linear	81.77	82.26	77.08
Radial basis	71.65	63.36	72.29
Sigmoid	45.32	58.03	67.92
Polynomial	32.91	41.22	43.96

Table 14 shows comparison between C5.0, J48, CART, Naïve Bayes, KNN, Random Forest and SVM. These algorithms are applied on three datasets containing varying attributes. For all three datasets Random forest performs best performance followed by C5.0 and J48. CART performs low as compared to others for all three datasets. Naïve Bayes performs poor for dataset 1 and dataset 2 and SVM has low performance for dataset 3.

Table 14. Performance comparison between J48, CART, Naïve Bayes, KNN, Random forest and SVM for three datasets

Algorithms	Accuracy (in %)		
	Dataset 1	Dataset 2	Dataset 3
C5.0	**98.48**	**100**	**93.33**
J48	94.68	90.08	85.62
CART	79.49	74.81	80.00
Naïve Bayes	75.95	75.52	85.21
KNN	82.28	77.86	81.25
Random Forest	**99.75**	**100**	**98.12**
SVM	81.77	86.26	77.08

Performance evaluation for Random Forest and C5.0 are shown below:

Table 15. Confusion matrix of Random Forest for Dataset 1

Grade		Predicted				
		A	B	C	D	Fail
Actual	A	40	0	0	0	0
	B	0	60	0	0	0
	C	0	0	62	0	0
	D	0	0	0	103	1
	Fail	0	0	0	0	129

Table 15 shows confusion matrix for Random Forest for 5 classes. For grades A to D students are correctly classified to their actual classes. Only one student from fail class is predicted to class D.

Table 16. Confusion matrix of Random Forest for Dataset 2

End semester percentage		Predicted			
		Best	Good	Pass	Very Good
Actual	Best	8	0	0	0
	Good	0	54	0	0
	Pass	0	0	27	0
	Very Good	0	0	0	42

For dataset 2, Random Forest provides predictions for 4 classes. These classes indicate end semester percentage grade. All the grades are correctly classified to their respective classes.

Table 17. Confusion matrix of Random Forest for Dataset 3

Class		Predicted		
		High	**Low**	**Medium**
Actual	**High**	140	1	1
	Low	0	127	0
	Medium	2	6	203

Table 17 has confusion matrix for dataset 3 only low class students are correctly classified to their actual class. One of high class student is misclassified as low and one more student is misclassified as medium class. Eight medium class students are missclassified out of which 2 belong to high class and 6 belong to low class.

Table 18. Performance evaluation of Random Forest

Random Forest	TP rate	FP rate	Precisi on	Recall	Class
Dataset 1	1.000	0.000	1.000	1.000	A (16-20)
	1.000	0.000	1.000	1.000	B (14-15)
	1.000	0.000	1.000	1.000	C (12-13)
	1.000	0.003	0.990	1.000	D (10-11)
	0.992	0.000	1.000	0.992	Fail
Dataset 2	1.000	0.000	1.000	1.000	Best
	1.000	0.000	1.000	1.000	Good
	1.000	0.000	1.000	1.000	Pass
	1.000	0.000	1.000	1.000	Very Good
Dataset 3	0.992	0.009	0.979	0.992	High
	1.000	0.020	0.947	1.000	Low
	0.957	0.000	1.000	0.957	Medium

Concerning table 18, observed that there are accurate classifications of students to their appropriate classes for datasets 1 and 2. Hence, the FR rates are zero for these datasets. Some of the misclassifications occur in dataset 3. Hence, precision for the high and low class is not 1 as well as FPR is not 0, recall is not 1 for medium and high class.

Table 19. Confusion matrix of C5.0 for Dataset 1

Grade		Predicted				
		A	**B**	**C**	**D**	**Fail**
Actual	**A**	39	0	0	0	0
	B	1	60	0	0	0
	C	0	0	59	1	0
	D	0	0	3	101	0
	Fail	0	0	0	1	130

Table 19. shows confusion matrix of C5.0 for dataset 1 where misclassifications occur for class B, C, D and Fail. One student of class B is misclassified as student of class A. One student of class C is predicted as student of class D. Three students of class D are incorrectly predicted as class C. One student of Fail class is misclassified to class D.

Table 20. Confusion matrix of C5.0 for Dataset 2

End semester percentage		Predicted			
		Best	**Good**	**Pass**	**Very Good**
Actual	**Best**	8	0	0	0
	Good	0	54	0	0
	Pass	0	0	27	0
	Very Good	0	0	0	42

The above confusion matrix shows that all students are correctly classified to their rightful classes giving 100% accuracy.

Table 21. Confusion matrix of C5.0 for Dataset 3

Class		**Predicted**		
		High	**Low**	**Medium**
Actual	High	131	0	11
	Low	1	123	3
	Medium	8	9	194

From above table it is clear that 11 students of High class are misclassified as Medium class students. For Low class, one is missclassified to high and 3 students to medium class. Medium class has 17 misclassifications, 8 belong to high class and 9 belong to low class.

Table 22. Performance evaluation of C5.0

C5.0	**TP rate**	**FP rate**	**Precis ion**	**Recall**	**Class**
Dataset 1	0.975	0.000	1.000	0.975	A (16-20)
	1.000	0.003	0.983	1.000	B (14-15)
	0.951	0.003	0.983	0.951	C (12-13)
	0.980	0.010	0.971	0.980	D (10-11)
	1.000	0.003	0.992	1.000	Fail
Dataset 2	1.000	0.000	1.000	1.000	Best
	1.000	0.000	1.000	1.000	Good
	1.000	0.000	1.000	1.000	Pass
	1.000	0.000	1.000	1.000	Very Good
Dataset 3	0.922	0.032	0.935	0.922	High
	0.968	0.011	0.931	0.968	Low
	0.919	0.062	0.932	0.919	Medium

For evaluation of results produced by algorithms, the measures Precision, Recall, TPR and FPR used in current work. FPR rate for dataset 1 for prediction in class A implies that the students those are observed to be failed are not misclassified to any other class hence value is zero and the precision for class A implies that all the positive classes are rightfully classified to their respective classes producing no false negatives hence according to the equation (1) value is calculated as 1. The performance of C5.0 for dataset 2 is high hence there are no occurrences of false positives and true negatives which makes precision, recall and TPR values 1 and no misplacement of records hence FPR is zero. For dataset 3, most of the misclassifications occur for Medium class students hence having the high FPR among other classes. None of the classes in dataset 3 is accurately places hence they are not having other evaluation value as 1.

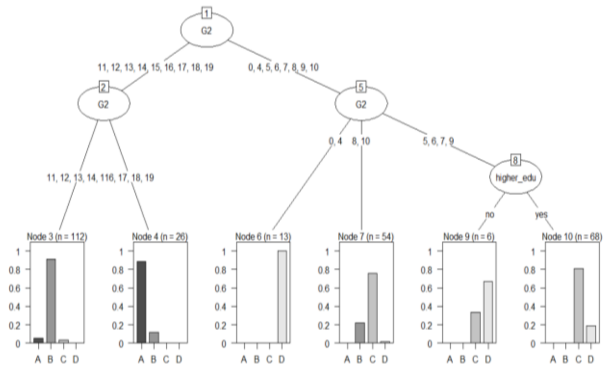

Fig. 15. A sample Decision Tree for dataset 1

The above figure shows a sample C5.0 Decision Tree for dataset 1. For building such a decision tree classifier, this algorithm calculates entropies of each attribute present in the training set. The formula for entropy is:

$$\mathrm{E(A)} = \sum_{i=1}^{v} \frac{p_i + n_i}{p + n} I(p_i, n_i) \tag{7}$$

Where,
p is number of records in positive class, P
n is number of records in negative class, N

After determining entropies, the information required is calculated as follows

$$\mathrm{I}(p,n) = -\frac{p}{p+n}\log_2\frac{p}{p+n} - -\frac{n}{p+n}\log_2\frac{n}{p+n} \tag{8}$$

Finally Information Gain (IG) is calculated having formula as follows:

$$\mathrm{gain}(A) = \mathrm{I}(p,n) - \mathrm{E}(A) \tag{9}$$

The variable having the highest IG is selected as a root node, which is G2 in this case. When attributes are available in continuous form, a particular threshold considered for splitting such as for G2, the values are divided from 0 to 10 and 11 to 20 forming two nodes. Again for the next level of tree G2 is split into two partitions. Each of the leaf nodes shows bar graph containing the distribution of instances for each of the target classes. For example, according to the following rule, a student having 18 marks in G2 will be classified into class A.

G2 in ------> {11, 12, 13, 14, 15, 16, 17, 18, 19}:
: G2 in ---> {11, 12, 13, 14, 15}: B (112/10)
: G2 in ---> {16, 17, 18, 19}: A (26/3)

Attribute importance determined by C5.0 model, in this case, has G2 with 75% and High school education with 25% importance. Compared to Table 2, which has correlated variables with target class, G2 has the highest place in both correlation and variable importance.

6. Conclusions

Prediction of students' performance is key factor to improve reputation of educational institutions. To overcome challenges in student performance prediction various classifiers such as C5.0, J48, CART, Naïve Bayes, SVM, KNN and Random forest are implemented for three datasets. Datasets are collected from school, college and e-learning platform.

Pearson correlation is applied on each dataset and features that are highly correlated to target output are chosen. The highly correlated attributes found are from academic and behavioural factors. For dataset 1, G1 and G2.For dataset 2, 10^{th} percentage 12^{th} percentage, internal assessment and attendance are impacting factors on student performance. For dataset 3, behavioural attributes namely number of raised hands and visited resources have high correlation.

On selected parameters the various classifiers are applied along with their tuning parameters such as trials and Confidence Factor in C5.0 ,minimum number of instances per leaf for J48, feature selection method in CART, Laplace for Naïve Bayes, value of k in KNN, number of trees to be formed in Random forest and type of kernel in SVM.

Increase in value of trials and Confidence Factor have drastically enhanced performance of C5.0. Increase in number of instances in leaf node for J48 and use of Laplace in Naïve Bayes tends to give poor performance. For all datasets, use of unpruned tree in J48 gives better performance. When CART used information gain performance obtained is better for dataset 2 and 3, but performs poor for dataset 1. For Random forest when number of trees is increased accuracy increases. SVM with linear kernel gives better results for all three datasets and polynomial kernel performs poor for each case in dataset. It is observed that performance of Random forest and C5.0 performs best for all three datasets.

References

[1] http://educationaldatamining.org/
[2] Mahboob, T., Irfan, S., & Karamat, A., "A machine learning approach for student assessment in E-learning using Quinlan's C4. 5, Naive Bayes and Random Forest algorithms," *IEEE 19th International Multi-Topic Conference (INMIC)*, pp. 1-8, December 2016.
[3] Marbouti, F., Diefes-Dux, H. A., & Madhavan, K., "Models for early prediction of at-risk students in a course using standards-based grading", *Computers & Education*, vol 103,pp 1-15, 2016.
[4] Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J., "Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses", *Computers in Human Behavior*, vol 73, pp 247-256, 2017.

[5] Zhang, X., Xue, R., Liu, B., Lu, W., & Zhang, Y., "Grade Prediction of Student Academic Performance with Multiple Classification Models", *IEEE 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pp. 1086-1090, 2018.
[6] Al-Shehri, H., Al-Qarni, A., Al-Saati, L., Batoaq, A., Badukhen, H., Alrashed, S., ... & Olatunji, S. O., "Student performance prediction using support vector machine and k-nearest neighbor", *IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)* pp. 1-4, April 2017.
[7] Amrieh, E. A., Hamtini, T., & Aljarah, I., "Preprocessing and analyzing educational data set using X-API for improving student's performance", *IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, pp. 1-5, November 2015.
[8] Hussain, S., Dahan, N. A., Ba-Alwib, F. M., & Ribata, N., "Educational data mining and analysis of students' academic performance using WEKA", *Indonesian Journal of Electrical Engineering and Computer Science*, vol 9(2), pp. 447-459, 2018.
[9] Parack, S., Zahid, Z., & Merchant, F.," Application of data mining in educational databases for predicting academic trends and patterns", *IEEE International Conference on Technology Enhanced Education (ICTEE)*, pp. 1-4, January 2012.
[10] Alfiani, A. P., & Wulandari, F. A. (2015). Mapping student's performance based on data mining approach (a case study). *Agriculture and Agricultural Science Procedia*, *3*, 173-177. 2015
[11] Christian, T. M., & Ayub, M., "Exploration of classification using NBTree for predicting students' performance", *IEEE International Conference on Data and Software Engineering (ICODSE)*, pp. 1-6, November 2014.
[12] Natek, S., & Zwilling, M.,"Student data mining solution–knowledge management system related to higher education institutions" *Expert systems with applications*, vol. 41(14), pp. 6400-6407, 2014.
[13] Goga, M., Kuyoro, S., & Goga, N., "A recommender for improving the student academic performance", *Procedia-Social and Behavioral Sciences*, vol. 180, pp. 1481-1488, 2015.
[14] Huang, S., & Fang, N., "Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models" *Computers & Education*, vol. 61, pp. 133-145, 2013.
[15] Guruler, H., Istanbullu, A., & Karahasan, M., "A new student performance analysing system using knowledge discovery in higher educational databases", *Computers & Education*, vol. 55(1), pp. 247-254, 2010.
[16] Sorour, S. E., & Mine, T., "Building an interpretable model of predicting student performance using comment data mining", *IEEE 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, pp. 285-291, July 2016.
[17] Vandamme, J. P., Meskens, N., & Superby, J. F., "Predicting academic performance by data mining methods", *Education Economics*, vol. 15(4), pp. 405-419, 2007.
[18] Gray, G., McGuinness, C., & Owende, P., "An application of classification models to predict learner progression in tertiary education", *IEEE International Advance Computing Conference (IACC)*, pp. 549-554, February 2014.
[19] Macfadyen, L. P., & Dawson, S., "Mining LMS data to develop an "early warning system" for educators: A proof of concept", *Computers & education*, vol. 54(2), pp. 588-599, 2010.
[20] Hoffait, A. S., & Schyns, M., "Early detection of university students with potential difficulties", *Decision Support Systems*, vol. 101, pp. 1-11, 2017.
[21] Palmer, S., "Modelling engineering student academic performance using academic analytics", *International journal of engineering education*, vol. 29(1), pp. 132-138, 2013.
[22] Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S., "Web usage mining for predicting final marks of students that use Moodle courses", *Computer Applications in Engineering Education*, vol. 21(1), pp. 135-146, 2013.
[23] Mishra, T., Kumar, D., & Gupta, S., "Mining students' data for performance prediction", *Proceedings of international conference on advanced computing & communication technologies*, pp. 255-263, February 2014.
[24] Miguéis, V. L., Freitas, A., Garcia, P. J., & Silva, A., "Early segmentation of students according to their academic performance: A predictive modelling approach", *Decision Support Systems*, vol. 115, pp. 36-51, 2018.
[25] Laugerman, M., Rover, D. T., Shelley, M. C., & Mickelson, S. K., "Determining graduation rates in engineering for community college transfer students using data mining", *International Journal of Engineering Education*, vol. 31(6A), pp. 1448, 2015.
[26] Romero, C., & Ventura, S., "Data mining in education", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3(1), pp. 12-27, 2013.
[27] Arsad, P. M., & Buniyamin, N., "A neural network students' performance prediction model (NNSPPM)", *IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)*, pp. 1-5, November 2013.
[28] Roy, S., & Garg, A., "Predicting academic performance of student using classification techniques", *4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)*, pp. 568-572, October 2017.
[29] Mueen, A., Zafar, B., & Manzoor, U., "Modeling and predicting students' academic performance using data mining techniques", *International Journal of Modern Education and Computer Science*, vol. 8(11), pp. 36, 2016.
[30] Khan, I. H., "A Unified Framework for Systematic Evaluation of ABET Student Outcomes and Program Educational Objectives", 2019.
[31] Kumar, M., Singh, A. J., & Handa, D., "Literature survey on student's performance prediction in education using data mining techniques", *International Journal of Education and Management Engineering*, vol. 7(6), pp. 42-49, 2017.
[32] Inyang, U. G., Eyoh, I. J., Robinson, S. A., & Udo, E. N., "Visual Association Analytics Approach to Predictive Modelling of Students' Academic Performance", 2019.
[33] https://archive.ics.uci.edu/ml/datasets/Student+Academics+Performance#
[34] https://archive.ics.uci.edu/ml/datasets/student+performance
[35] https://www.kaggle.com/aljarah/xAPI-Edu-Data
[36] Badr, G., Algobail, A., Almutairi, H., & Almutery, M., "Predicting students' performance in university courses: a case study and tool in KSU mathematics department", *Procedia Computer Science*, vol. 82, pp. 80-89, 2016.
[37] https://www.statisticssolutions.com/correlation-pearson-kendall-spearman/
[38] Drazin, S., & Montag, M., "Decision tree analysis using weka", *Machine Learning-Project II, University of Miami*, pp. 1-3, 2012.

[39] Azmi, M. S. B. M., & Paris, I. H. B. M., "Academic performance prediction based on voting technique", *IEEE 3rd International Conference on Communication Software and Networks*, pp. 24-27, May 2011.
[40] Amra, I. A. A., & Maghari, A. Y., "Students performance prediction using KNN and Naïve Bayesian", *8th International Conference on Information Technology (ICIT)*, pp. 909-913, May 2017.
[41] Sokolova, M., & Lapalme, G., "A systematic analysis of performance measures for classification tasks", *Information processing & management*, vol. 45(4), pp. 427-437, 2009.

Authors' Profiles

Madhuri T. Sathe completed M. Tech. in Computer Science and Engineering from Rajarambapu Institute of Technology, Rajaramnagar in year the 2018. Her areas of interest are classification, machine learning, prediction and analytics.

Amol C. Adamuthe received Ph.D. from Mumbai University and Master of Technology in Computer Engineering from Dr. B. A. Technological University, Lonere, MS, India. He is currently an Assistant Professor at Rajarambapu Institute of Technology, Sakharale, Sangli, MS, India. His areas of interest are technology forecasting, optimization algorithms, soft computing, and cloud computing. He has published more than 45 papers at the international and national level. He is the recipient of the Institution of Engineers (India) Young Engineers Award for year 2018-19 in Computer Engineering discipline.

How to cite this paper: Madhuri T. Sathe, Amol C. Adamuthe, "Comparative Study of Supervised Algorithms for Prediction of Students' Performance", International Journal of Modern Education and Computer Science(IJMECS), Vol.13, No.1, pp. 1-21, 2021.DOI: 10.5815/ijmecs.2021.01.01