Int. J. Advance Soft Compu. Appl, Vol. 13, No. 1, March 2021
Print ISSN: 2710-1274, Online ISSN: 2074-8523

An Improved Prediction System of Students' Performance Using Classification model and Feature Selection Algorithm

Md. Ahsan Arif[1], Asma Jahan[2], Mausumi Islam Mau[3], Razia Tummarzia[4]

Asian University of Bangladesh, Bangladesh
[1]e-mail: mdahsanarif@aub.edu.bd
[2]-mail: jahan76884@gmail.com
[3]e-mail: maus96607@gmail.com
[4]e-mail: rmnilimacse96@gmail.com

Abstract

Prediction on students' academic performance is a subject or situation under consideration of interest for both students and educational institutions at large. This is a strong emerging issue to be widely studied to determine the factors affecting students' performance in the sense of data mining. Respondent students' internal and external factors are students' demographic state, socio-economic condition, parental educational status, extra-curricular activities, teaching quality, and students' learning behavior that aims to predict performance predominantly. We use WEKA (Waikato Environment for Knowledge Analysis) tools to accomplish our research. This paper proposes the feature selection algorithm and classification model to develop student performance prediction system. The performance of the proposed methods is compared with another feature selection algorithm that is based on classification model. This paper aims to provide machine learning classifier algorithms with selected attribute to get better accuracy and comparable different feature selection algorithms because of its feature selection algorithm which is used to select the most valuable features. In the feature selection algorithm, WrapperSubsetEval method with Random Forest classification is better than other feature selection method and other classification algorithm to predict students' performance.

Keywords: *Data Mining, Internal and external factors, Feature selection algorithms, Classification model*

1 Introduction

The number of educational institutions has rapidly been growing day by day. In this situation, quality of education is essential to integrate and disseminate research findings about institutional ranking. Without quality education, it is not possible to explore students' idea that, they will not be able to develop their country eventually; as because, without idea there is no creativity and without creativity there is no development. To ensure quality education, our contemporary institutions are highly decorated to improve students' learning system targeting their success rate in the final examinations. The following research will predict students' performance, which will help us to understand students' learning ability of a course and will also ensure a teacher's teaching quality to judge their student's outcome of learning. The motto of the research will also be helpful for their authority of the institution and their parents to understand a student's learning behavior. If teachers are not able to predict students' performance under their assigned course load then those students may not be able to achieve or fulfill their academic knowledge in future. As a result, their pre-requisite knowledge on particular course may not be sufficient. Eventually, students may drop-out or fail in their final examination including their core courses.

2 Related Work

According to Anjana Pradeep et al. (2015), the number of students who are dropping out every semester or year has been increasing and the affected factor is shortage of quality education in their experience. According to Ahmed O. Ameen et al. (2019), student performance reduced depends on various factors; such as, teaching quality, availability of teaching resources etc. According to Pamela Chaudhury et al. (2016), determining some data encompasses all the features that might affects the performance of the student internal grades; such as, demographic factors, socio-economic conditions, personal characteristic traits and school related features have a high impact on how the students would perform.

According to Febrianti Widyahastuti and Viany Utami Tjhin (2017), emphasize that student performance depends on various attributes; such as, student behaviour, demographic situation and student information, psychological and socio-economic context. According to, Preet Kamal et al. (2017), determined that student performance depends on learning speed of a student. Who were the slow learners in a high school class their performance may be decreased? Usamah et al. (2013) stated that learning assessment and co-curricular activities effects on student performance. According to Anjana Pradeep et al. (2015), state that to solve the dropout problem student performance prediction is important. Predicting student dropout or failure is a difficult task caused by having multi-factor problem. To resolve this problem DM (Data Mining) algorithms and approaches is helpful on predicting students' failure. According to Pamela Chaudhury et al. (2016), describe that student performance is predicted by using different types of ML (Machine learning) classifier. Unbalanced data can degrade to get accurate accuracy of classifiers. The technique of oversampling and under sampling is helpful to solve this problem. According to Havan Agrawal et al. (2015), states that to identity high-risk students as well as to identify the features which affect the performance of students' ML technique is used.

According to Y. K. Salal et al. (2019), emphasizes that researchers in educational data mining are used to CGPA and other internal performance (student attendance, class performance, co-curricular activities etc.) marks on predicting student performance. According to Sajadin Sembiring et al. (2011), proposed a model which can determine student conditions. They use kernel k-means clustering technique and Smooth Support Vector Machine (SSVM) classification to analyze the relationship between students' behavior and their success. Jabeen Sultana et al. (2019), proposed a model that works on classification technique to predict student performance. They collect data from a Saudi University database and also use WEKA and Rapid miner tools. Compared to other classification algorithms in this paper, MLP (Multilayer Perceptron) gave best precision and accuracy and it is more efficient than others algorithm. Mukesh Kumar and et al. (2017), discussed that student performance prediction system which is helpful to find out drop-out students by using DM (Data Mining) technique. They found that HSG, SSG and other related education are the main factors that could impact of a student's result, which is a cause of drop-out.

Ahmed Mueen et al. (2006), proposed a model to apply data mining techniques to predict and analyze students' academic performance based on their academic record and forum participation. Educational Data Mining (EDM) is an emerging tool for academic intervention. The educational institutions can use EDM for extensive analysis of students' characteristics. In this study, we have collected students' data from two undergraduate courses. Three different data mining classification algorithms (Naive Bayes, Neural Network, and Decision Tree) were used on the dataset. The prediction performance of the three classifiers is measured here and compared. In this research, teachers will be helpful to improve their students' academic performance.

3 Methodology

Educational Data Mining (EDM) describes the application of data mining, machine learning and statistics to information that is generated from educational settings. At the highest level, the field seeks to develop and improve methods for exploring this data, which often have multiple levels of meaningful hierarchy in order to discover new insights about how people learn in the context of such settings. In doing so, EDM has contributed to theories of learning investigated by the researchers in educational psychology and its learning-science. The Educational Data Mining (EDM) plays a vital role to develop learning system. Data Mining (DM)) can find unexplored information and developing methods is to explore unique patterns from a bulk amount of data to understand students in all aspects and their learning system as well (Hany M. et al, 2012). Data mining (DM) involves performance prediction, student modelling, domain modelling, analysis and visualization of student data, recommendation system, and grouping of students and so on (C. Romero et al., 2010). EDM applies techniques of statistics, machine learning, and data mining to analyze the data collected while teaching and learning. EDM reflects to use educational data repositories for understanding the learning techniques that helps educators to find knowledge that can be used to have better understanding of student's behavior, to improve teaching quality, improve student academic performance etc. Classification is one of the important research issues in machine learning; whereas the documents are classified with supervised knowledge and Prediction accuracy of the features selected from the feature selection algorithms that can also be evaluated through classification of algorithms. The main objective of this research work is to find the best classification algorithm and feature selection algorithm.

3.1 Participants

In this study among the participants are selected and based on their respective department. The required Data is collected from the recent students in a university. This dataset contains 33 attributes and 649 instances. In following both internal and external attributes are included because of academic performance is being affected by a multitude of heterogeneous factors (internal and external) that influenced student performance. Internal factors are demographic, socio-economic conditions, personal characteristic, marital status etc. The students, as to external factors, are name, age, gender, definition of learning facilities, study time, extra-curricular activities etc.

3.2 Using Tools

We used WEKA (Waikato Environment for Knowledge Analysis) software, version 3.8.6 and it is an open-source software. It provides tools for data preprocessing, implementation of several Machine Learning algorithms and visualization tools. Development of machine learning techniques is easier and it easily can be applied to real-world of data mining problems.

3.3 Data Pre-Processing

Before applying the machine learning algorithm, it is essential to carry out some data pre-processing tasks; such as, data cleaning, integration, transformation and discretization. Data pre-processing task includes finding of relevant data, irrelevant data or missing data. Erroneous data or ambiguous data may be corrected or removed; whereas missing data must be supplied. Data pre-processing also includes removal of noise or outliers and collecting necessary information to model or account for noise removal. Raw data contain a mixture of relevant and irrelevant feature. Irrelevant features reduce the accuracy of machine learning algorithms and prediction of model cannot properly be predicted. The feature selection algorithms can help us to get the right feature.

Transformation: Transformation is the process of converting the data into a common format for processing. Some data may be encoded into more usable format. We have converted our data file into CSV format of WEKA (here CSV means Comma Separated Value).

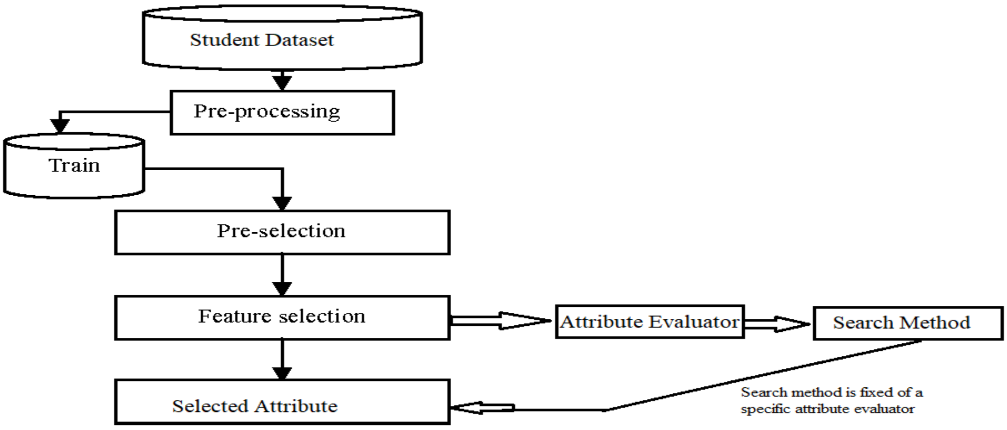


Figure 1: Data processing Steps in WEKA

3.4 Using Approach

We use feature selection algorithm and predictive model to predict student performance. The following figure 2 is the proposed approach here.

3.4.1 Feature Selection Algorithm

Feature selection is the first step to process a dataset which is greedy and enforces an important role to build a prediction model. Feature selection is one of the most periodic and notable technique in data pre-processing. It plays a vital role in improving the quality of prediction models. Feature selection eliminated unrelated data from the educational archive. It reduces the complexity and overfitting of a model. And also reduces the input variable number and automatically or manually select which attribute is the most important or suited to improve the accuracy of a predictive model. In machine learning feature selection, it is known as variable selection, attribute selection or variable subset selection. Feature selection algorithm is of three types. Such as;

- Wrapper Method
- Filter Method
- Embedded method

Wrapper method: Wrapper method creates different subset of input feature to get best performing metrics of a model. It is the best algorithm to find the best feature for best performance. Wrapper method includes recursive feature elimination, sequential feature selection algorithm and genetic algorithms.

Filter method: Filter method finds out the intrinsic properties of the feature. It evaluates the relationship between each input variable and the target variable; and the input variables will be used in the predictive model. Filter method includes information gain, chi-square test, fisher score, correlation coefficient and variance threshold technique.

Embedded Method: Embedded methods are a catch-all group of techniques which perform feature selection as part of the model construction process. It includes L1 (LASSO) regularization and decision tree algorithm.

Many feature selection techniques are supported in WEKA. The Feature selection is divided into two parts. Such as; Attribute Evaluator and Search Method and each section has multiple techniques.

Attribute Evaluators:

- CfsSubsetEval.
- ChiSquaredAttributeEval.
- ClassifierSubsetEval.
- ConsistencySubsetEval.
- GainRatioAttributeEval.
- InfoGainAttributeEval.
- OneRAttributeEval.
- PrincipalComponents.
- ReliefFAttributeEval.
- SVMAttributeEval.
- SymmetricalUncertAttributeEval.
- WrapperSubsetEval.

Search Methods:

- BestFirst.
- GreedyStepwise.

- Ranker.

We use attribute evaluator as CfsSubsetEval, ClassifierSubsetEval, InfoGainAttributeEval and WrapperSubsetEval. In WEKA some attribute evaluator is fixed to use of specific search methods.

3.4.2 Predictive Model

Predictive model is the procedure that means; to create process and approve a model that may be used to future consequences. Predictive model is five types. Such as; Forest model, Classification model, Outlier model, Time series model and Clustering model. We use classification model to predict student performance. It includes linear classifier (Logistics regression, Naïve Bayes classifier, Fisher's linear discriminant), Support Vector Machine, Quadratic classifier, Kernel estimation (K-nearest neighbor), Decision tree (Random Forest), and Neural Networks (Learning vector quantization) machine learning algorithm.

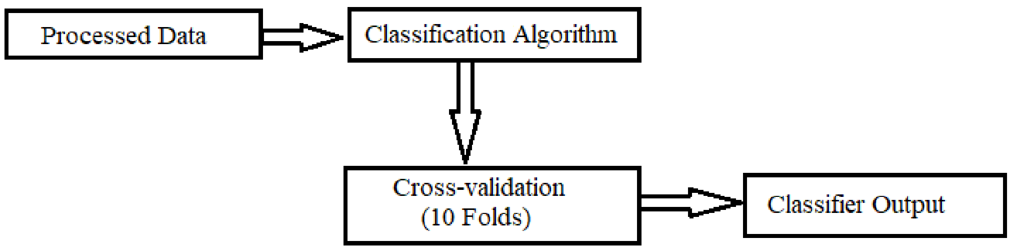


Figure 2: Working procedure of a classifier in WEKA

The performance of a predictive model depends on some evaluating factor those are classification accuracy, specificity, sensitivity/recall, precision, F1 score/F-measure, AUROC curve, kappa statistics, mean absolute error, root mean squared error, Relative absolute error, root relative squared error etc.

Accuracy: Accuracy is the correctly predicted data of a predictive model. If the accuracy of model is lower than that of the model it cannot properly predict.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{1}$$

Here, TP = True Positive, TN = True Negative, FP= False Positive, FN = False Negative

TP: If the real value is true and algorithm predict this value as a true value than it is called TP.

TN: If the real value id true but algorithm predict this value as a false value than it is called TN.

FP: If the real value is not true but algorithms predict this value as a true value than it is called FP.

FN: If the real value is true but algorithms predict this value as a false value than it is called FN.

Specificity: Specificity is the proportion of actual negatives, which got predicted as the negative (or true negative).

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}} \tag{2}$$

Sensitivity/Recall: Sensitivity/Recall describe that what proportion of actual positives was identified correctly?

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

Precision: Precision describe that what proportion of positive identifications was actually correct?

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

F1-score/F-Measure: F1 Score considers both precision and recall. It is the harmonic mean or average of the precision and recall.

$$F1\ Score = \frac{2 * (\text{Recall} * \text{Precision})}{\text{Recall} + \text{Precision}} \tag{5}$$

3.4.2.1 Logistic Regression

Logistic Regression (LR) transforms its output using the logistic sigmoid function to let return a probability value which can then be mapped to two or more discrete classes. In machine learning, we use sigmoid function to map predictions to probabilities.

$$\text{Sigmoid Function}, \text{S(z)} = \frac{1}{1 + e^{-z}} \tag{6}$$

In equation 6, here S(z) is sigmoid function and e is Euler's number.

In the below, Table 1 describes the evaluator performance of the selected attribute which are performing in Logistic Regression. For the different attribute evaluator TP, FP, Precision, Recall, F-measure and other performance evaluating metrics that is different.

Table 1: An average accuracy measure of Logistic Regression with attribute evaluators

Evaluator	TP	FP	Precision	Recall	F-measure
CfsSubsetEval	0.693	0.399	0.685	0.693	0.688
ClassifierSubsetEval	0.761	0.317	0.756	0.761	0.757
InfoGainAttribute	0.701	0.341	0.697	0.701	0.697
WrapperSubsetEval	0.764	0.293	0.762	0.764	0.763

3.4.2.2 Naïve Bayes

One of the main positions of the Naïve Bayes (NB) algorithm is that each characteristic is individualistic, gives good output for the problem considered (Pushpa et al. 2017). It is a probabilistic classifier which means that it can predict on the basis of the probability of an object. Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem. The formula for Bayes theorem is given as below:

$$\mathrm{P(A|B)} = \frac{\mathrm{P(B|A)P(A)}}{\mathrm{P(B)}} \tag{7}$$

In equation 7, here A and B means events, P(A|B) means probability of A given B is true, P(B|A) means probability of B given A is true and P(A), P(B) means the independent probabilities of A and B.

In the below, Table 2 shows the evaluator performance of the selected attribute which are performing in Naïve Bayes classifier. For the different attribute evaluators TP, FP, Precision, Recall, F-measure and other performance evaluating metrics is different.

Table 2: An average accuracy measure of Naïve Bayes with attribute evaluators

Evaluator	TP	FP	Precision	Recall	F-measure
CfsSubsetEval	0.701	0.403	0.691	0.701	0.693
ClassifierSubsetEval	0.735	0.335	0.731	0.735	0.733
InfoGainAttribute	0.698	0.360	0.695	0.698	0.689
WrapperSubsetEval	0.761	0.290	0.760	0.761	0.761

3.4.2.3 Stochastic Gradient Descent

Stochastic gradient descent (SGD) is a constant method for optimizing an objective function with appropriate precision properties. This algorithm is the most efficient and easily implemented to be compared with other machine learning models (Ofori, Gitonga, 2020). In SGD it uses only a single sample.

In the below, Table 3 describes the evaluator performance of the selected attribute which are performing in SDG. For the different attribute evaluator TP, FP, Precision, Recall, F-measure and other performance evaluating metrics is different.

Table 3: An average accuracy measure of SGD with attribute evaluators

Evaluator	TP	FP	Precision	Recall	F-measure
CfsSubsetEval	0.695	0.412	0.684	0.695	0.686
ClassifierSubsetEval	0.758	0.321	0.753	0.758	0.754
InfoGainAttribute	0.698	0.352	0.694	0.698	0.691
WrapperSubsetEval	0.760	0.306	0.756	0.760	0.757

3.4.2.4 K-Nearest Neighbor

K-Nearest Neighbor (KNN) took less time to identify the students' performance as slow learner, average learner, good learner and excellent learner (Mayilvagan and Kalpanadevi, 2014). It can easily be implemented and robust to the noisy training data. It can be used for classification and regression predictive problems also.

In the below, Table 4 describes the evaluator performance of the selected attribute, which are performing in KNN. For the different attribute evaluator TP, FP, Precision, Recall, F-measure and other performance evaluating metrics is different.

Table 4: An average accuracy measure of KNN with attribute evaluators

Evaluator	TP	FP	Precision	Recall	F-measure
CfsSubsetEval	0.659	0.479	0.640	0.659	0.642
ClassifierSubsetEval	0.655	0.446	0.645	0.655	0.648
InfoGainAttribute	0.622	0.416	0.618	0. 622	0.619
WrapperSubsetEval	0.684	0.410	0.676	0.684	0.679

3.4.2.5 Decision Tree

Decision Tree (DT) techniques give up awareness into the decision making processed (Quadri and Kalyankar, 2010). The data in the decision tree algorithm does not need to be normalized. There are five commonly used algorithms for the decision tree are; ID3, CART, CHAID, C4.5 algorithm and J48. Decision trees are built on using algorithms ID3, C4.5 and J48 is an execution version of C4.5 in WEKA software (Patel, Prajapati, 2018). We use J48 algorithm here. This algorithm is dependent on the Entropy and the information gains concepts.

$$Entropy, E(S) = \sum_{i=1}^{c} -(p_i \log_2 p_i) \tag{8}$$

In equation 8, we use entropy to measure disorder or uncertainty of the model. Here, Pi is simply the frequentist probability of an element or class ‘i’ in our data.

In the below, Table 5 describes the evaluator performance of the selected attribute which are performing in J48. For the different attribute evaluator TP, FP, Precision, Recall, F-measure and other performance evaluating metrics is different.

Table 5: An average accuracy measure of Decision Tree with attribute evaluators

Evaluator	TP	FP	Precision	Recall	F-measure
CfsSubsetEval	0.673	0.449	0.658	0.673	0.661
ClassifierSubsetEval	0.747	0.325	0.743	0.747	0.744
InfoGainAttribute	0.606	0.439	0.599	0.606	0.601
WrapperSubsetEval	0.700	0.391	0.692	0.700	0.694

3.4.2.6 Random Forest

Random forest (RF) makes multiple decision trees and combines to gain a more accuracy rates and fixed prediction (Unal, 2020). This algorithm takes less training time than other classifier algorithms and represents high accuracy.

In the below, Table 6 describes the evaluator performance of the selected attribute which are performing in Random Forest. For the different attribute evaluator TP, FP, Precision, Recall, F-measure and other performance evaluating metric is different.

Table 6: An average accuracy measure of Random Forest with attribute evaluators

Evaluator	TP	FP	Precision	Recall	F-measure
CfsSubsetEval	0.712	0.467	0.705	0.712	0.679
ClassifierSubsetEval	0.753	0.360	0.747	0.753	0.742
InfoGainAttribute	0.716	0.353	0.719	0.716	0.703
WrapperSubsetEval	0.775	0.355	0.775	0.775	0.760

3.4.2.7 Support Vector Machine

Support vector machine (SVM) algorithm presents one of the most robust prediction methods based on the statistical learning framework. Hamalainen et al. (2006), chose Support Vector Machine as their prediction technique because it suited well in small datasets. A support vector machine constructs a hyperplane or set of hyperplanes in a high or infinite-dimensional space, which can be used for classification, regression. SVM algorithms use a set of mathematical functions that are explained as the kernel. The RBF kernel function has been used in SVM to classify the dataset.

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{9}$$

In equation 9, here σ is the variance and our hyperparameter.

In the below, Table 7 describes the evaluator performance of the selected attribute which are performing in Support Vector Machine. For the different attribute evaluator TP, FP, Precision, Recall, F-measure and other performance evaluating metrics is different.

Table 7: An average accuracy measure of SVM with attribute evaluators

Evaluator	TP	FP	Precision	Recall	F-measure
CfsSubsetEval	0.700	0.404	0.689	0.700	0.691
ClassifierSubsetEval	0.750	0.339	0.744	0.750	0.744
InfoGainAttribute	0.696	0.353	0.692	0.696	0.690
WrapperSubsetEval	0.775	0.314	0.751	0.755	0.752

4 Results, Analysis and Discussions

This study, after using feature selection algorithm, we understand which attribute is not good and which good one(s) for our predictive model. In the below, Table 8, Table 9 and Table 10 is the comparison study of which algorithm and which feature selection method is the best for student performance prediction model.

Here, Table 8 describes that Random Forest algorithm with WrapperSubsetEval method is good for student performance prediction model.

Table 8: Comparison among different algorithm accuracy with WrapperSubsetEval

Evaluator: WrapperSubsetEval	
Algorithm	Correctly Classified Instance /Accuracy
Logistic Regression	76.4253%
Naïve Bayes	76.1171 %

Stochastic Gradient Descent	75.963 %
K-Nearest Neighbor	68.4129 %
Decision Tree	69.9538 %
Random Forest	77.5039 %
Support Vector Machine	75.5008 %

In the below, Table 9 describes that Random Forest algorithm with CfsSubsetEval method is good for student performance prediction model.

Table 9: Comparison among different algorithm accuracy with CfsSubsetEval

Evaluator: CfsSubsetEval	
Algorithm	Correctly Classified Instance /Accuracy
Logistic Regression	69.3374 %
Naïve Bayes	70.1079 %
Stochastic Gradient Descent	69.4915 %
K-Nearest Neighbor	65.9476 %
Decision Tree	67.3344 %
Random Forest	71.1864 %
Support Vector Machine	69.9538 %

In the below, Table 10 describes that Logistic Regression algorithm with ClassifierSubsetEval method is good for student performance prediction model.

Table 10: Comparison among different algorithm accuracy with ClassifierSubsetEval

Evaluator: ClassifierSubsetEval	
Algorithm	Correctly Classified Instance /Accuracy
Logistic Regression	76.1171%
Naïve Bayes	73.4977 %
Stochastic Gradient Descent	75.8089 %
K-Nearest Neighbor	65.4854 %
Decision Tree	74.7304 %
Random Forest	75.3467 %
Support Vector Machine	75.0385 %

In the below, table 11 describe that Random Forest algorithm with InfoGainAttributeEval method is good for student performance prediction model.

Table 11: Comparison among different algorithm accuracy with InfoGainAttributeEval

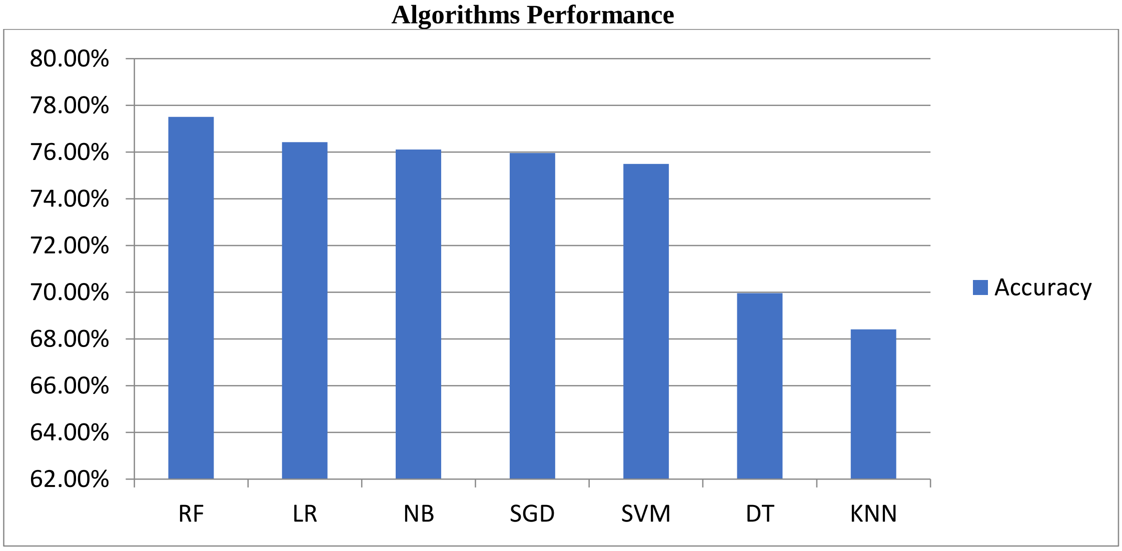
Evaluator: InfoGainAttributeEval	
Algorithm	Correctly Classified Instance /Accuracy
Logistic Regression	70.1079 %
Naïve Bayes	70.1079 %
Stochastic Gradient Descent	69.7997 %
K-Nearest Neighbor	62.2496 %
Decision Tree	60.5547 %
Random Forest	71.6487 %
Support Vector Machine	69.6456 %

5 Conclusion

We use attribute evaluator as CfsSubsetEval, ClassifierSubsetEval, InfoGainAttributeEval and WrapperSubsetEval. These attribute evaluators represent filter method and wrapper method as a feature selection algorithm. WrapperSubsetEval is the attribute evaluator of wrapper method. And this evaluator performance is best for our model. In future, we will try to increase our algorithm accuracy.

The outcome of this research shows that, Random Forest algorithm is best for the student performance prediction system and WrapperSubsetEval method is the best attribute selector.

Algorithms Performance



References

[1] Pushpa, SK., Manjunath, TN., Mrunal, TV., Singh, Amartya., Suhas, C. (2017). Class result prediction using machine learning**.** *International Conference On Smart Technologies For Smart Nation, IEEE,978-1-53860569.*

[2] Ofori, F., Maina, E., Gitonga R., (2020). Using Machine Learning Algorithms to Predict Students Performance and Improve Learning Outcome: A Literature Based Review. Journal of Information and Technology, 4(1), 33-55.

[3] Mayilvaganan, M., Kalpanadevi, D. (2014). Comparison of classification techniques for predicting the performance of student's academic environment. Communication and Network Technologies (ICCNT), 2014 *International Conference on, IEEE,pp.113–118.*

[4] Quadri, M.N., Kalyanka, N.V. (2010). Drop Out Feature of Student Data for Academic Performance Using Decision Tree Techniques. Global Journal of Computer Science and Technology, Vol. 10 Issue 2 (Ver 1.0).

[5] Patel, H., Patel, P. (2018). Study and Analysis of Decision Tree Based Classification Algorithms. *INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING, 6(10),2347-2693.*

[6] Ünal, F. (2020). Data Mining for Student Performance Prediction in Education. Support vector machine, from Wikipedia, the free encyclopedia.

[7] Ainen, W., Vinni, M. (2006) Comparison of machine learning methods for intelligent tutoring systems, Intelligent Tutoring Systems. Springer*, pp. 525–534.*

[8] Pradeep, A., Das, S., Kizhekkethottam, J. (2015), Students dropout factor prediction using EDM techniques. *International Conference on Soft-Computing and Networks Security (ICSNS)IEEE Xplore, 10.1109/7292372.*

[9] Chaudhury, P., Mishra, S., Kumar,H., Kishore, T. (2016), Enhancing the capabilities of Student Result Prediction System. Proceedings of the Second *International Conference on Information and Communication Technology for Competitive Strategies, 978-1-4503-3962.*

[10] Widyahastuti, F., Tjhin,V. (2017), Predicting Students Performance in Final Examination using Linear Regression and Multilayer Perceptron. IEEE, 978-1-5090-4688.

[11] Kamal, P., Ahuja, S.A (2017), Review on Prediction of Academic Performance of Student Sat-Risk Using Data Mining Techniques. This article is published with open access at www.chitkara.edu.in/publications,*Vol. 5, No.1 pp. 30–39.*

[12] Mat, U., Buniyamin, N., Arsad, P. (2013), An Overview of Using Academic Analytics to Predict and Improve Students' Achievement: A Proposed Proactive Intelligent Intervention Rosni Abu Kassim. *IEEE 5th Conference on Engineering Education (ICEED), 978-1-4799-2332.*

[13] Agrawal, H., Mavani, H. (2015), Student Performance Prediction using Machine Learning. *International Journal of Engineering Research & Technology (IJERT), 4(03), 2278-0181.*

[14] Salal, Y.K., Abdullaev, S.M., Kumar, M. (2019), Educational Data Mining: Student Performance Prediction in Academic. *International Journal of Engineering and Advanced Technology (IJEAT), 8(4), 2249-8958.*
[15] Sultana, J., Rani, U., Farquad, M. (2019), An Efficient Deep Learning \Method to Predict Student's Performance. https://www.researchgate.net/publication
[16] Kumar, M., Singh, A.J., Handa, D. (2017), Literature Survey on Student's Performance Prediction in Education using Data Mining Techniques. I.J. Education and Management Engineering, 6, 40-49.