Journal of Language and Linguistics in Society
ISSN: 2815-0961
Vol : 03, No.02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JLLS
DOI: https://doi.org/10.55529/jlls.32.19.33

An Effectual Model for Early Prediction of Academic Performance using Ensemble Classification

Chietra Jalota*

Associate Professor, Echelon Institute of Technology, Faridabad, India.

*Corresponding Email: *chitra19878@gmail.com*

Received: 02 December 2022 **Accepted:** 19 February 2023 **Published:** 22 March 2023

***Abstract:** In the past few years, researchers are focused towards educational data mining (EDM) to improve the quality of education. Student's academic performance prediction is a vital issue for improving the value of education. Research study conducted in the literature review mainly captivated the academic performance prediction at higher education. Though the academic performance at secondary level is infrequent, the same could be a scale for a student's performance at subsequent levels of education. Poor grades at lower levels also impact student's future performance. In this paper, an effectual model is built with the help of significant factors that affect a student's academic performance at secondary level using single and ensemble techniques of classificification For this, both single and ensemble classification techniques are used in this paper To do the same, three single classifiers (classification algorithm) i.e., MLP, Random Forest and PART along with three well recognized ensemble algorithms Bagging (BAG), LogitBoost (LB) and Voting (VT) are applied on the datasets. For better performance of aforementioned classifiers, blended versions (single + ensemble-based classifiers) of classification models are also built. Assessment metrics i.e., accuracy, precision, recall and F-measure used to evaluate the performance of our proposed model. Evaluation results shows that Logitboost with Random Forest outperformed with 99.8% accuracy. It is clearly visible from results that the proposed model is useful for academic performance prediction to improve learning outcomes in future.*

***Keywords:** Learning Management System, Educational Data Mining, Clustering, Classification, Prediction, Relationship Mining.*

1. INTRODUCTION

Educational data mining (EDM) is an interesting field of research to discover hidden patterns of educational data for various academic purposes [1,2]. By considering appropriate as well as significant factors for academic performance prediction, weak students would be identified at the initial stage so that timely action can be taken by academicians [3]. For the development of a nation, education plays a vital role. Prediction of academic performance is a major concern for all parents, teachers, departments and government. To inclined the students' performance

Journal of Language and Linguistics in Society
ISSN: 2815-0961
Vol : 03, No.02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JLLS
DOI: https://doi.org/10.55529/jlls.32.19.33

at secondary level, various factors i.e., psychological, schooling, family background demographic, social background, psychological factors must be studied and examined carefully [4]. As the value of these factors are not same for all students, therefore the academic performance of students is diverse. So, to predict student's academic performance, above-mentioned factors must be taken into consideration. Science and Maths are two benchmarks to predict the performance at higher level. Therefore, the performance at lower secondary level is very decisive as these two subjects are base for the students while setting their academic targets. Academic performance can be improved by recognizing feeble areas regarding personal traits and academics. To achieve this, the prediction system must be designed in such a way so that students' academic progress could be estimated at an early stage. This type of prediction is useful in the following areas:

i) Improvement in performance in the weak areas by students.
ii) It helps in the choice-based assortment of subjects.
iii) It provides a base while deciding career goals.
iv) Student's grade prediction on the basis of their previous performance.

Prediction of academic performance also restrains the failure and dropout ratios of students. Moreover, academic growth and professional achievement is also affected by subject selection at lower secondary level.

To figure out the required patterns or information on the basis of historical analysis of data, approaches of data mining are very useful [5]. To build a predictive model, numerous classification and regression approaches are used. Classification technique uses pre-classified data to illustrate the unclassified data [6]. Several algorithms such as logistic regression (LogitBoost), probability (Naïve Bayes, Bayes Network), decision trees (ID3, REP Tree, C4.5 etc.), backtracking (MLP) are used to perform classification. There are some limitations with these classifiers while scheming model's accuracy because these algorithms are single classifiers. Therefore, it is significant to improve the performance of single classifiers by ensembling of various classifiers [4]. To predict the academic performance of students on the basis of their academic and personal traits, a blended model of classification is developed by combining single and ensemble classifiers [5].

Grades prediction in educational data mining (EDM) uses academic data and other relevant features and can be used in several ways for improvement in the education quality. The major contributions of this paper include:

(a) Identification of Personal Traits

(b) Edification of a classification model to analyse the student\s personal as well as their academic detail for prediction of their academic improvement with higher accuracy and precision.

The rest of the paper is organized as follows: Section 2 presents a literature review that highlights the findings of prior research studies to recognize most common important features affecting students' academic performance and the data mining techniques used most frequently for academic performance prediction. Research Methodology is explained in Section 3 and Section 4 is about results and discussions. Limitations of this study is given in Section 5 and Section 6 is about conclusion and future work.

2. Literature Review

Journal of Language and Linguistics in Society
ISSN: 2815-0961
Vol : 03, No.02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JLLS
DOI: https://doi.org/10.55529/jlls.32.19.33

In this section, all previous papers related to academic performance prediction are discussed. In addition to this, research papers entailing an importance of factors/features selection while predicting academic performance are also taken into consideration. Below given is the summary of such research papers:
In the present time, Educational Data Mining (EDM) is rising because of tremendous growth of educational resources (internet, online tools) to deliver education [8].

To recognize appropriate knowledge of educational datasets, various techniques of educational data mining (EDM) are used. Fundamentally, educational data (raw) is used as an input. After that learning analytics, data mining and machine learning techniques are applied on that raw dataset to get useful and meaningful information. This information is a base for comparative analysis of manual and predictable data. [9].

Information mining procedure was used by authors to determine factors affecting student's academic performance. Results of this study showed that cumulative evaluation points and internal assessments are two most substantial attributes/qualities used to predict academic performance of students. Apart from the above-mentioned attributes, some other significant attributes like social attributes, extracurricular activities, inner as well as personal appraisal were also identified for academic performance prediction [10].

In [11], to handle a variety of data, a group of data mining methods and list of data mining tasks were discussed in this study. To expand the understanding about data by an organization, various techniques of data mining were used for making modifications in the defined operations and procedures to make a rise in their profitability.

To measure the steady growth of academic performance of students in each semester, Cumulative Grade Point Average (CGPA) is a protruding factor [12]. Results of this study have also shown some significant factors like class tests, assignment completion, study hours etc. for academic performance prediction.

Problem of imbalanced classification on the Pima Indian Diabetes dataset using weka was discussed. A variety of filters i.e., Spread Subsample filter is used for majority class's under-sampling while Resample filter is used for random over-sampling of minority class were used for the balancing of class distribution in the pre-processing stage of the above-mentioned dataset, under weka. It was clearly mentioned in the result that by using sampling-based techniques all accuracy parameters such as Recall, F1-Score, Precision, and ROC area of minority class were improved [13].

For a classification problem, selection of significant features is very important. In this paper, two approaches of filter selection i.e. wrapper-based feature selection and correlation feature selection (CFS). It is clearly visible from results that highest accuracy with the correlation feature selection algorithms is achieved by using SMO and J48 whereas highest accuracy with the wrapper subset feature selection algorithms to predict various grades i.e., low, medium and high grade for the students is measured by using Naïve Bayes [14].

Five machine learning algorithms i.e. Support vector Machine, J48, Random Forest, Multilayer perceptron and Naïve Bayes along with statistical techniques for the enlightenment of academic performance and learning habits of students for the prediction of their performance level. It can be seen from the results that multilayer perceptron shows best performance as compared to other classifiers. [15]

Various methods of educational data mining (EDM) which can handle a variety of data, their application areas and list of data mining tasks were discussed in this study. Several objectives of data mining tools and techniques which are helpful for many organizations to expand their understanding about data i.e., knowledge by making modifications in the defined operations and procedures were also explained in this study [16].

There is an interconnection of students' extracurricular activities and his/her academic attainment. A strong predictor of academic performance prediction is student' presence in a class. [17]
In [18], authors have focused on demographic as well as academic attributes i..e, external and internal assessments marks and cumulative grade point average (CGPA) to predict the academic performance of students.
In [19], authors have used pre-university marks to predict academic performance of fourth year students of an undergraduate program. Their main emphasis was to predict academic performance on the basis of grades scored in internal and external assessments. Attributes like socio-economic and family attributes are not considered in this study while predicting academic performance.

In [20], a predictive investigation model using regression and ANOVA was built to infer the interaction of learners as well as to measure the level of satisfaction by university students in an online course. A questionnaire related to socioeconomic and demographics factors was given to students for data collection.

In [21], supervised machine learning techniques were applied for academic performance prediction of students in examination. Various factors like demographics and social interest were considered to presume the student's grade in their final examination.

For meaningful interpretation of massive educational data and to evaluate the usefulness of education, academic analysts mainly use data mining and machine learning approaches. [22].

In [23], an innovative machine learning model was built for prediction of academic performance of students in a degree program on the basis of their previous academic records.

In [24], a comparative analysis model using numerous techniques of machine learning on the basis of their time complexity was built. In this study, author has discussed the challenges and current limitations of machine learning as well as put emphasis on the composite methods as these are significant to improve the accuracy of predictive models using single classifiers.

Journal of Language and Linguistics in Society
ISSN: 2815-0961
Vol : 03, No.02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JLLS
DOI: https://doi.org/10.55529/jlls.32.19.33

To improve the working of composite models by fusion of base classifiers, Bagging, Boosting and stacking are special type of ensemble methods. For classification and prediction, bagging is used. In this paper, SMOTTEEN with an ensemble classifier is applied on an imbalanced dataset to improve accuracy [25].

To predict the student performance in final examinations on the basis of their demographic and social attributes, two prediction models using K-Nearest Neighbour (KNN) and support vector machine (SVM) were built [26].

Machine learning and feature selection techniques were used to select significant attributes and properties which have influential role on students' learning and performance results [27].

Machine learning-based ensemble methods which are also known as composite methods are a vital approach to strengthening single classifiers. By using this method, the power of multiple models is used to achieve improved prediction accuracy as compared to any of the individual models could achieve independently [28].

3. RESEARCH METHODOLOGY

Weka is used to perform data mining errands used in proposed research methodology for this paper A detailed explanation of research methodology is given in figure.1.

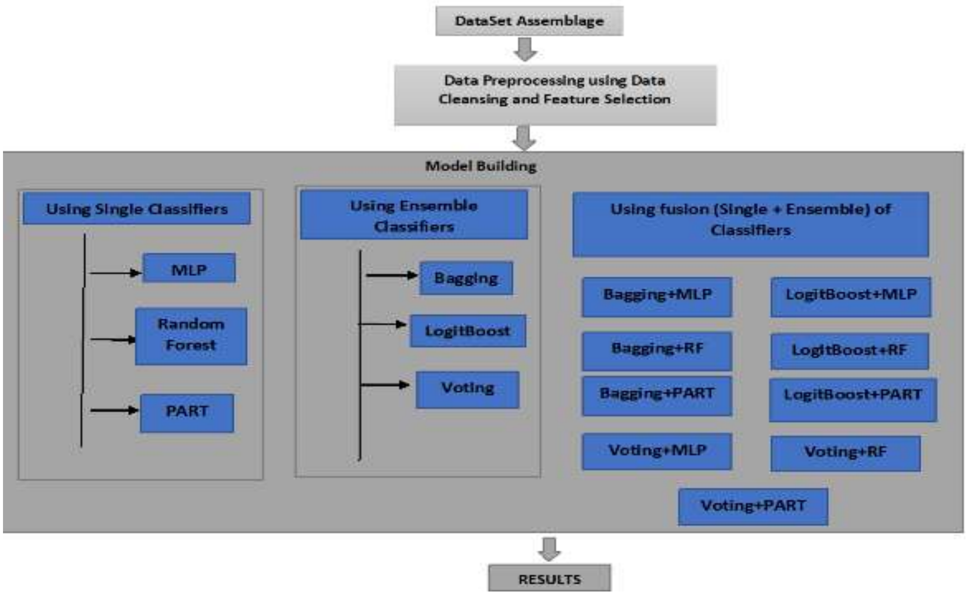


Figure 1. Proposed Methodology

We obtained an xAPI-Edu-Dataset from kaggle comprising of 480 instances It comprises 305 males and 175 females. The students in this dataset belongs to various origins such as Kuwait, Jordan, Palestine Iraq, Lebanon, Tunis, Saudi Arabia, Egypt, Syria, USA, Iran, Libya, Morocco, and Venezuela. This dataset is composed of two educational semesters in which 245 students belongs to first semester and 235 students belongs to second semester of the academic course.

Journal of Language and Linguistics in Society
ISSN: 2815-0961
Vol : 03, No.02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JLLS
DOI: https://doi.org/10.55529/jlls.32.19.33

Student's attributes of this dataset are visited resources, section, semester, topic, parent satisfaction from school etc. Students are categorized into three numerical intervals on the basis of their total grade i.e.; low-level (L) (0-69), mid-level (M) (70-89) and high-level (H) (90-100). A detailed description of student dataset is given in Table 1.

Following steps are an integral part of data pre-processing:

1. **Data cleansing:** It is a primary step of data pre-processing. It is used to clean dataset after removing all inappropriate attributes from the dataset. The dataset used in this study consists of 480 instances and there was no missing value.

2. **Feature/Attribute Selection:** To reduce dimensionality in the feature space of the given dataset, feature selection is used to avoid the model overfitting. By using feature selection, a subdivision of original features will be chosen to eliminate the duplicate and superseded features. In this paper, we have applied correlation-based analysis using ranker best first search method to appraise dynamic features to build models with good performance. Out of 16 features, 7 were selected on the basis of correlation with the final outcome for better results. The primary objective of the projected model is to predict the student's academic performance under the category M, L and H.

Table 1. ATTRIBUTE OF STUDENT'S DATASET (xAPI)

Attributes	**Dataset's probable value**
Demographic	Nationaity, Gender
Academic	Educational Stage, Grade level, Section
Behavioural	Hands Raised, LMS Accessing Rate, Overall Satisfaction of parents

In this paper, research methodology is primarily preceded by ensemble methods which includes bagging, boosting (logitboost) and voting. Classification as well as prediction is also possible by exploiting these ensemble methods. To attain higher prediction accuracy, models can build up using ensemble classifiers. Tuples are sorted randomly into different bags for model building in bagging ensemble method which is known as booostrap aggregation. In bagging, all models are built parallel. An average of all models is calculated for overall decision. By doing so, variance of the model will be lowered. It is clearly visible from results of that bagging attains the maximum efficiency comparative to the other methods. Bootstrap works better than Bagging. In bootstrap, weights are assigned to the erroneously tuples identified by previous classifiers. To build the final model, this weighted average is taken as a definitive input. Due to high efficiency of boosting algorithms, weak and low-performing models can be converted into robust models. AdaBoost, MultiBoost and GradientBoost are some common examples of boosting classifiers. Ensembling of voting, bagging and boosting with base classifiers is a recent trend of research. A prediction model on the basis of bagging and boosting was projected by Livieris et al. For automated classification of news articles in terms of fake/real content, bagging, boosting, and voting classifiers are used. To predict the software defects, a two-layer ensemble approach was proposed by Yang et al.

Journal of Language and Linguistics in Society
ISSN: 2815-0961
Vol : 03, No.02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JLLS
DOI: https://doi.org/10.55529/jlls.32.19.33

3.1 Model Building Using Base Classifiers

Numerous classification algorithms are being used for classification and prediction like decision trees (REP Tree, C4.5, CART, and J48, etc.), probability (Naïve Bayes and Bayes Networks, etc.), backtracking (ANN like MLP, etc.), logistic regression (Logistic Boost, etc.). When used them independently for classification/prediction, they are referred to as a single base classifier. It is shown in the literature review that, the Multi-Layer Perceptron (MLP), Random Forest (RF) and J48 are the most efficient and frequently used algorithms as compared to other algorithms for performance prediction. It has been observed that RF took less training time for each data instance than the MLP. The MLP, RF, and PART are best for large and small size data sets and these classifiers have also shown better accuracy.

3.2 Model Building using Ensmeble Methods

To build a perfect predictive model using machine learning techniques, ensemble methods are used. Ensemble methods construct multiple models and unite them to get better results. Results produced using ensemble methods are generally more precise. Ensemble methods are also known as a significant and efficient development in the area of data mining and machine learning. In this paper, we use three main ensembling techniques i.e., Bagging (BAG), Logit Boost (LB), and voting (VT). Two main categories of ensemble methods are homogenous and heterogeneous. In homogenous ensemble methods, a single algorithm is applied on various training datasets to build numerous classifiers e.g., bagging and boosting. On the other hand, in heterogeneous ensemble methods; dissimilar algorithms are used to control training datasets to build different models e.g., voting and stacking.

3.3 Model Building using Hybrid Ensmeble Methods

In this step, hybridization of models takes place in which hybrid ensemble-based models are build, train and test with base classifiers. This blending of ensemble models with base classifiers improves the generality and prediction ability of ensemble models. This hybridization technique always builds precise and effectual models of machine learning. Given below is a list of nine hybrid ensemble models:

i) Multilayer Perceptron with Bagging (BAG + MLP)
ii) Random Forest with Bagging (BAG + RF)
iii) PART with Bagging (BAG + PART)
iv) Multilayer Perceptron with LogitBoost (LB + MLP)
v) Random Forest with LogitBoost (LB+RF)
vi) PART with LogitBoost (LB+PART)
vii) Multilayer Perceptron with Voting (VT + MLP)
viii) Random Forest with Voting (VT + RF)
ix) PART with Voting (VT + PART).

4. RESULTS AND DISCUSSIONS

In this paper, WEKA is used `to build various models to evaluate students' performance. A series of different experiments were conducted to evaluate the student's performance. After experiments, a comparative analysis was done through single classifiers, ensemble-based classifiers and by fusion of single and ensemble classifiers. The time complexity of each

Journal of Language and Linguistics in Society
ISSN: 2815-0961
Vol : 03, No.02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JLLS
DOI: https://doi.org/10.55529/jlls.32.19.33

algorithm plays an important role to check the efficiency of an algorithm and the same can be denoted in terms of Big O notation. [24]. To determine the progresses in the performance of different models, a comparative analysis has been performed which determines the competent model in predicting student academic performance.

4.1 Results of Single and Ensemble Classifiers

After the data pre-processing stage, three base classifiers (MLP, RF and PART) applied to the given dataset. Among these three base classifiers, it is clearly visible from figure 2 that Random Forest (RF) achieved the highest accuracy (75.3), precision (75.2), recall (75.3) and F-measure (75.3). There are two parts of figure; performance of classifiers is described through bar graph in the first part while the same is done through tabular form in the second part.

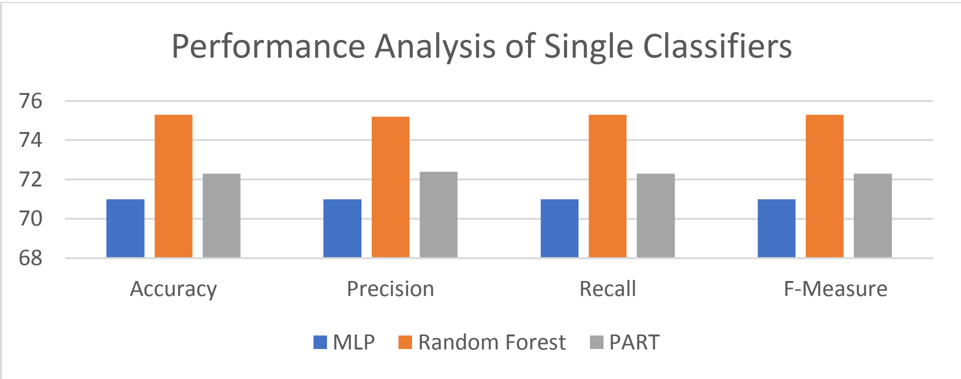


Figure 2. Single Based Classifiers

Name of the Classifier	Accuracy	Precision	Recall	F-Measure
MLP	71	71	71	71
Random Forest	75.3	75.2	75.3	75.3
PART	72.3	72.4	72.3	72.3

After the results of applying single based classifiers on given dataset, three different ensemble classifiers i.e., bagging, Logitboost, and Voting are also applied on the same dataset. Among these three ensemble classifiers, bagging outperformed the other classifiers, achieving higher accuracy i.e., 85.6 as shown in Figure 3. This figure is having two parts; Performance parameters i.e.; Accuracy, precision, recall, and F-score is shown through bar chart in the first part while the second part specifies the performance of classifiers through the same measures but in tabular form. Ensembled classifiers have shown better performance for accuracy, precision, recall and F-score.

Journal of Language and Linguistics in Society
ISSN: 2815-0961
Vol : 03, No.02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JLLS
DOI: https://doi.org/10.55529/jlls.32.19.33

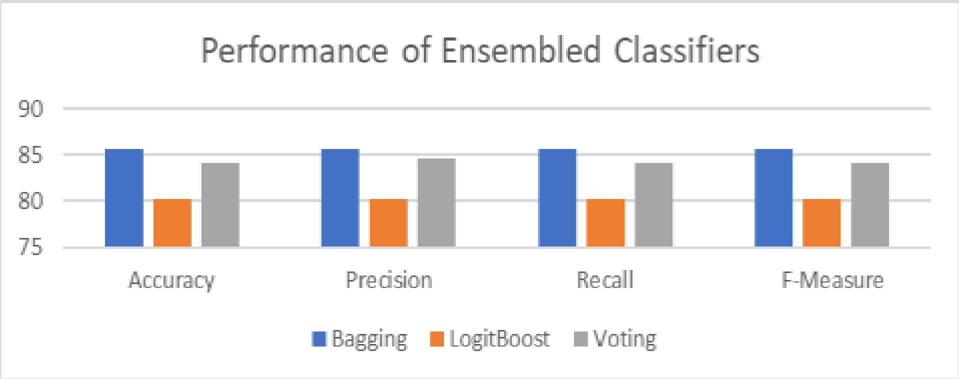


Figure 3. Ensemble Classifiers

Name of ensemble classifier	**Accuracy**	**Precision**	**Recall**	**F-Measure**
Bagging	85.6	85.7	85.6	85.6
LogitBoost	80.2	80.2	80.2	80.2
Voting	84.1	84.6	84.2	84.2

4.2 Results of Ensemble Models

The purpose behind this phase was to produce the hybrid models of single classifiers with ensemble classifiers. In this phase, nine hybrid models i.e., fusion of BAGGING (BAG) with MLP, RANDOM FOREST and with PART, fusion of LogitBoost (LB) with MLP, RANDOM FOREST and with PART as well as fusion of Voting (VT) with MLP, RANDOM FOREST and with PART. Results of these fusion models have been shown in figures 4-6.

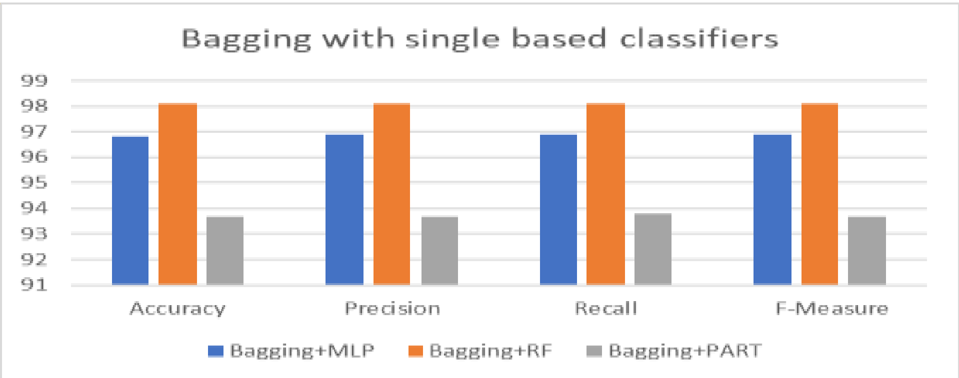
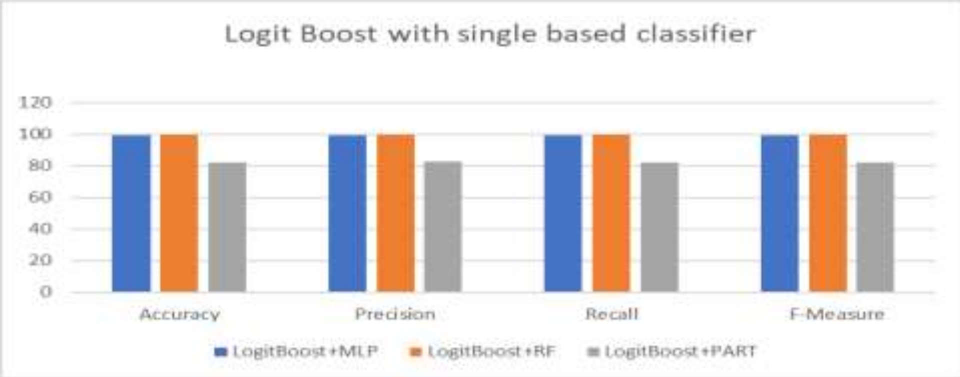


Figure 4. Bagging with single classifiers

Name of fusion Classifier	Accuracy	Precision	Recall	F-Measure
Bagging+MLP	96.8	96.9	96.9	96.9
Bagging+RF	98.1	98.1	98.1	98.1
Bagging+PART	93.7	93.7	93.8	93.7

Journal of Language and Linguistics in Society
ISSN: 2815-0961
Vol : 03, No.02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JLLS
DOI: https://doi.org/10.55529/jlls.32.19.33



	Accuracy	Precision	Recall	F-Measure
LogitBoost+MLP	99.5	99.6	99.6	99.6
LogitBoost+RF	99.8	99.8	99.9	99.9
LogitBoost+PART	82.4	82.5	82.4	82.3

Figure 5. Boosting (Logit Boost) with single based classifiers

	Accuracy	Precision	Recall	F-Measure
Voting+MLP	97.7	97.7	97.7	97.7
Voting+RF	99.8	99.9	99.8	99.8
Voting+PART	92.2	92.4	92.3	92.3

Figure 6. Voting with single based classifiers

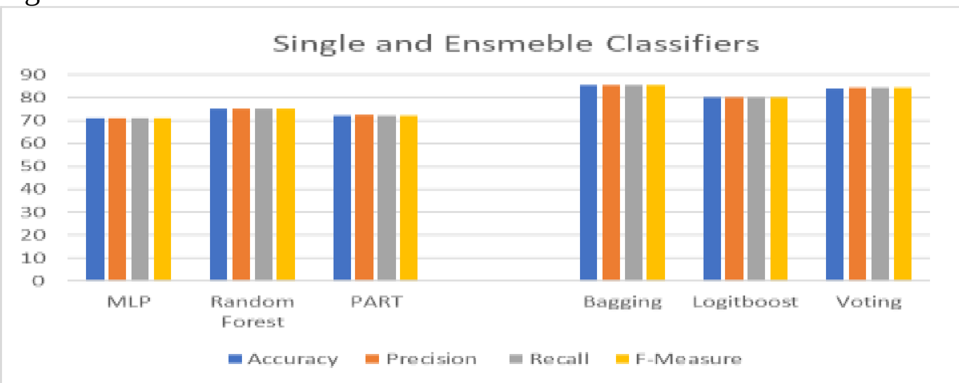
Results of fusion-based models shows that bagging (BAG) with Random Forest (RF) is having the highest accuracy (i.e., 98.1%). Other parameters of performance i.e., precision, F-score and recall have also shown better results for this model. It is shown in figure 5 that there is highest accuracy while fusing LogitBoost with Random Forest (RF). This model has also achieved good performance for precision, recall, and F-score. The results related to the fusion of voting (VT) with various single classifiers showed that VT + MLP achieved greater accuracy (i.e., 97.7%), as shown in Figure 6. This model also showed better performance in terms of precision, recall, and F-score. Every figure comprises two parts; performance of classifiers in terms of

Journal of Language and Linguistics in Society
ISSN: 2815-0961
Vol : 03, No.02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JLLS
DOI: https://doi.org/10.55529/jlls.32.19.33

accuracy, precision, recall, and F-score is shown in first part with the help of bar chart. The second part indicates the performance of classifiers in tabular form through the same measures.

4.3 Comparative Analysis of Single, Ensemble and Fusion Models

To examine the performance of different classifiers used in this paper, a relative as well as comparative analysis was made and the same is shown in figure. 7. It is clearly visible from the results shown in fig. 7 that there is an outstanding performance of ensemble-based models with respect to accuracy, recall, precision, and F-score as compared to single-based classifiers. With the help of bar charts and numeric terms, performance of above-mentioned models is shown in figure.7



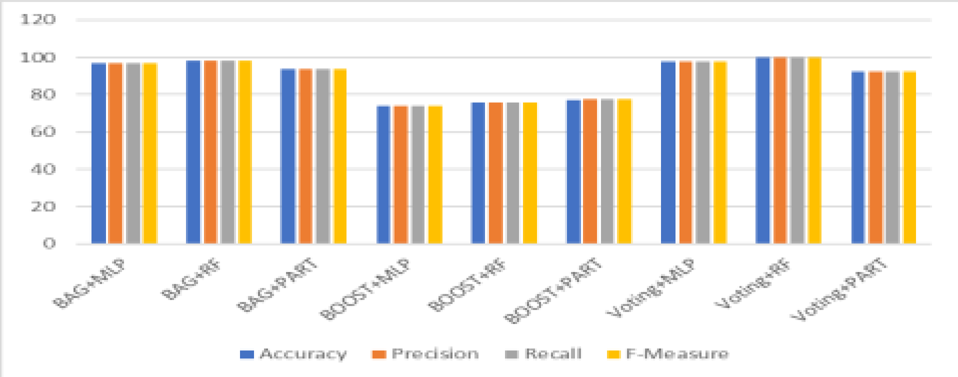
Single Based Classifiers				Ensemble Based Classifiers			
Accuracy	**71**	**75.3**	**72.3**	Accuracy	**85.6**	**80.2**	**84.1**
Precision	**71**	**75.2**	**72.4**	Precision	**85.7**	**80.3**	**84.2**
Recall	**71**	**75.3**	**72.3**	Recall	**85.6**	**80.2**	**84.2**
F-Measure	**71**	**75.3**	**72.3**	F-Measure	**85.6**	**80.2**	**84.3**

Figure 7. Comparison of single and ensemble-based classifiers

The aim of this study was to recognize the best fusion model having high performance by making a comparative analysis of all fusion-based models. Results shown in fig. 7 shows that there is a significant improvement in precision and accuracy while using fusion-based models. This has been proven from the results that precision and accuracy of the student prediction model can be improved by fusion based models.

With the help of bar chart and numeric terms, performance of various fusion models is shown in figure 8. Results of figure 8 clearly indicates that Voting and Random Forest (RF) achieved highest accuracy with respect to all measures i.e., accuracy, precision, recall and F-Measure. This model Voting+Random Forest (RF) achieved 99.8 accuracy and 99.9% precision, 99.8% recall and 99.8% F-score which is highest among all other fusion models. Thus, to make a significant improvement in student prediction model fusion-based models are always preferable

Journal of Language and Linguistics in Society
ISSN: 2815-0961
Vol : 03, No.02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JLLS
DOI: https://doi.org/10.55529/jlls.32.19.33



	Accuracy	Precision	Recall	F-Measure
BAG+MLP	96.8	96.9	96.9	96.9
BAG+RF	98.1	98.1	98.1	98.1
BAG+PART	93.7	93.7	93.8	93.7
BOOST+MLP	73.9	73.7	74	73.8
BOOST+RF	75.6	75.6	75.6	75.6
BOOST+PART	77.2	77.5	77.3	77.3
Voting+MLP	97.7	97.7	97.7	97.7
Voting+RF	99.8	99.9	99.8	99.8
Voting+PART	92.2	92.4	92.3	92.3

The F-Measure of Voting+RF has produced 99.8% accuracy which is a significant result. To classify illustrations for students' performance evaluation, fusion-based models provide better as well as precise outcomes.

Limitations

In this paper, we only consider the physical learning environment to predict student's performance. We should consider both physical and virtual educational environment to encompass for evaluation of students' performance prediction.

One more drawback of this paper is that the factors considered for this study are related only with a definite/specific educational level.

5. CONCLUSION AND FUTURE WORK

Several data mining techniques have been instigated on academic data as a standard procedure for the prediction of academic performance of students. An early prediction is valuable for those who are at-risk as well as facing difficulty to get good grades in the class. It is significant intermittently to support such students by predicting their performance regularly. To make predictions about such student's, an integrated ensemble model comprises features like students' demographic, family, social and academic attributes is developed in this paper. To assess a student at an initial stage, this model is highly beneficial. Out of various models i.e., single, ensemble, and fusion-based ensemble classifiers developed in this paper, model build up with LogitBoost and Random Forest (RF) proves the best model to predict students'

Journal of Language and Linguistics in Society
ISSN: 2815-0961
Vol : 03, No.02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JLLS
DOI: https://doi.org/10.55529/jlls.32.19.33

performance at the lower secondary level because of highest accuracy amongst all other models. In future, a system with meta-analysis could be developed as a decision support method for a large dataset on the basis of this model. Moreover, hybrid feature selection methods can also be considered in future to predict academic performance of students. For further enhancement of this model, gradient as well as extreme gradient boosting could be used.

6. REFERENCES

1. Kamran, S. Nawaz, I. Aslam, S. Zaheer, S. and Shaukat, U. (2016), "Student's performance in the context of data mining," in Proc. International Multi-Topic Conference (INMIC), pp.1-8.
2. Kamran, S. Nawaz, I. Aslam, S. Zaheer, S. and Shaukat, U. (2017), Student's Performance: A Data Mining Perspective, Lambert Academic Publishing, India.
3. Abid, A. Kallel, I. Blanco, I. and Benayed, M. (2018), "Selecting relevant educational attributes for predicting students' academic performance in Intelligent Systems Design and Applications," in Proc. Springer 17th International Conference on Intelligent Systems Design and Applications (ISDA 2017), Delhi, India, pp. 650–660.
4. Phua, E.J. and Batcha, N.K. (2020), "Comparative analysis of ensemble algorithms' prediction accuracies in education data mining". Journal of Critical Review, No.7, pp.37–40.
5. Abu, A. (2016), "Educational data mining & students' performance prediction", International Journal of Advanced Computer Science and Applications, No. 7, pp.212–220.
6. Romero, C. and Ventura, S. (2010) "Educational Data Mining: A Review of the State of the Art,". IEEE Trans. Syst. Man Cybern. Part. C (Appl.Rev.), Vol.40, No.6, pp.601–618.
7. Arun, D.K. Namratha, V. Ramyashree, B.V. Jain, Y.P. and Choudhury, A.R. (2021), "Student academic performance prediction using educational data mining,". In Proc of the 2021 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, pp. 1–9.
8. Alturki, S. and Alturki, N. (2021), "Using Educational Data Mining to Predict Students' Academic Performance for Applying Early Interventions," J. Inf. Technol. Educ. Innov. Pract. Vol. 20, pp.121–137.
9. Trautwein, U. Lüdtke, O. Marsh, H.W. Köller, O. and Baumert, J. (2006), "Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics," Journal of Educational Psychology, pp.788–806.
10. Li, F. Zhang, Y. Chen, M. and Gao, K. (2019), "Which Factors Have the Greatest Impact on Student's Performance," Journal of Physical Conference Series, pp.1288-1292.
11. Mehra C. and Agrawal R. (2020), "Comparative study of resampling techniques of imbalanced dataset", International Journal of Advance Science and Technology, Vol. 29, No. 3, pp.12699-12710.
12. Jalota, C. and Agrawal R. (2020), "Feature Selection Algorithms and Student Academic Performance: A Study", in Proc. Springer International conference on innovative computing and communication (ICICC-2020), 2020, pp. 317-328.

Journal of Language and Linguistics in Society
ISSN: 2815-0961
Vol : 03, No.02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JLLS
DOI: https://doi.org/10.55529/jlls.32.19.33

13. Jalota C. and Agrawal R. (2019), "Analysis of data mining using classification", in Proc. IEEE International conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCON-2019), pp. 247-253.
14. Mehra, C. and Agrawal, R. (2021)," Educational data mining approaches, challenges and goals: A Review", International Journal of Information, Communication and Computation Technology, Vol8, No.2, pp.442-447.
15. Shaukat, K. Luo, S. Varadharajan, V. Hameed, I.A. Chen, S. Liu, D. and Li, J. (2020), "Performance Comparison and Current Challenges of Using Machine Learning Techniques in Cybersecurity,". Energies 2020, Vol 13, No.10.
16. Mehra C. & Agrawal R, " Educational data mining approaches, challenges and goals: A Review", International Journal of Information, Communication and Computation Technology, 2021(2): pp.442-447.
17. Alturki, S.; Alturki, N. "Using Educational Data Mining to Predict Students' Academic Performance for Applying Early Interventions". J. Inf. Technol. Educ. Innov. Pract. 2021(20), pp.121–137.
18. Trautwein, U.; Lüdtke, O.; Marsh, H.W.; Köller, O.; Baumert, J, "Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics,". J. Educ. Psychol. 2006, pp.788–806.
19. Tomasevic, N.; Gvozdenovic, N.; Vranes, S, "An overview and comparison of supervised data mining techniques for student exam performance prediction,". Comput. Educ. 2019, pp.143-149.
20. Md Zubair Rahman, A.M.J, "Model of Tuned J48 Classification and Analysis of Performance Prediction in Educational Data Mining". International Journal of Applied Engineering Research,2018(20), pp.14717-14727.
21. Zollanvari, A.; Kizilirmak, R.C.; Kho, Y.H.; Hernandez-Torrano, D, "Predicting Students' GPA and Developing Intervention Strategies Based on Self-Regulatory Learning Behaviors,". IEEE Access, 2017(5), pp.23792–23802.
22. Aucejo, E.M.; French, J.; Ugalde Araya, M.P.; Zafar, B,. "The impact of COVID-19 on student experiences and expectations: Evidence from a survey,". J. Public Econ. 2020, 191.
23. Tomasevic, N.; Gvozdenovic, N.; Vranes, S, "An overview and comparison of supervised data mining techniques for student exam performance prediction,". Comput. Educ. 2019, pp.143-153.
24. Shaukat, K.; Luo, S.; Varadharajan, V.; Hameed, I.A.; Chen, S.; Liu, D.; Li, J,. "Performance Comparison and Current Challenges of Using Machine Learning Techniques in Cybersecurity,". Energies 2020,13(10),
25. Xu, J.; Moon, K.H.; van der Schaar, M, "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs,". IEEE J. Sel. Top. Signal. Process. 2017(11), pp.742–753.
26. Shaukat, K.; Luo, S.; Varadharajan, V.; Hameed, I.A.; Chen, S.; Liu, D.; Li, J, "Performance Comparison and Current Challenges of Using Machine Learning Techniques in Cybersecurity,". Energies, 2020, 13.
27. Hassan, H.; Ahmad, N.B.; Anuar, S, "Improved students' performance prediction for multi-class imbalanced problems using hybrid and ensemble approach in educational data mining," J. Phys. Conf. Ser. 2020, pp.1529-1540.

Journal of Language and Linguistics in Society
ISSN: 2815-0961
Vol : 03, No.02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JLLS
DOI: https://doi.org/10.55529/jlls.32.19.33

28. Wood, L.; Kiperman, S.; Esch, R.C.; Leroux, A.J.; Truscott, S.D, “Predicting dropout using student- and school-level factors: An ecological perspective,”. Sch. Psychol. Q. 2017, 32, pp.35–49.
29. Nahar, K.; Shova, B.I.; Ria, T.; Rashid, H.B.; Islam, A.H.M.S, “Mining educational data to predict students’ performance,”. Educ. Inf. Technol. 2021, 26, pp.6051–6067.
30. Madni, H.A.; Anwar, Z.; Shah, M.A, “Data mining techniques and applications—A decade review,” In Proc of the International Conference on Automation and Computing (ICAC), Huddersfield, UK, September 2017; pp. 1–7.

