Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Hard Voting Classifier Based Students Performance Prediction Using Machine Learning

Ashwini Virulkar[1], Dr. Ashish Sasankar[2]
[1] *Department of Electronics and Computer Science, RTMNU, Nagpur, India*
Email: isha84210@gmail.com
[2] *Indraprasth New Arts Commerce and Science College, Wardha, India*
Email: ashishdigital14@gmail.com

ARTICLE INFO

Received: 18 Dec 2024

Revised: 10 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

Ensemble learning is an imperative feature of machine learning. It helps in improving the prediction accuracy. The research paper uses the ensemble techniques predict the performance of students. Holistic development of students is essential at every stage of life. Predicting the performance of student to avoid dropout ratio becomes important to improve the outcome of student's performance in field of education. It includes the various factors that have an impact on students' performance. The dataset used is from Kaggle Higher education students' performance evaluation. The performance of individual classifier is evaluated that is Decision Tree classifier, Gradient Boosting model and Random Forest which is further combined. The Voting Classifier is applied to advance the achievement of the collaborative model which gives the accuracy of 86%.

Keywords: Education, Ensemble, Random Forest, Voting classifier, Decision Tree, Gradient Boosting

INTRODUCTION

Holistic development of student is need of this competitive world. Formal education plays important role in all round development. Along with informal education also plays an important role. The credibility for the higher education improves with skill development along with scientific study. Through the learning analytics (LA) pattern for predicting students' performance is appreciable with supervised leaning techniques
[1]. The need to predict the performance of learners is important to identify the drop out ratio at the earliest. The improvisation of accuracies for predicting grades is important since small discrepancies leads to improper results [2]. Machine learning methods helps to predict the performance of learners. The ensemble techniques help to advances the accuracy of the forecast. The research paper focuses on the enhancing the accuracy of the models with the help of ensemble learning provided by Voting classifier.

LITERATURE REVIEW

Prediction of student's performance is important to identify them at risk. The research paper focuses on predicting the performance using random forest, gradient Boosting and voting classifier. The resultant output performed the peak of 90% accuracy [3].

In ML, assessing and evaluating the student's performance is a major task. Amani Khalifa et al., in the research paper enhances the performance of model using optimization technique. GridSearch_CV, HalvingGridSearch_CV, and Optuna are used where Optuna performs the best in tuning hyperparameter [4].

Kumar et. al., uses the ensemble learning models in his study to improvise the performance of model at 99.50% accuracy. Firstly, the individual model's performance is evaluated with Naïve Bayes (NB), RF, Decision Tree (J48), Decision Table (DT), (MLP), JRip, and Logistic Regression (LR). Ensemble learning predicts models' performance in an effective way [5].

G. Nassreddine et. al., in his research work proposes the model to deal with data unbalancing. It proposes XGboost ensemble learning with ADASYN. The prime features are evaluated by SHAP technique. It showed the absentee of students negatively impact the performance where attendance is important [6].

Yan et. al., proposes a model to predict the performance of potential learners. It performed the prediction using stacking ensemble model. Initially Random Forest (RF), Support Vector Machine model (SVM) and Adaboost, boosting technique are combined with the logistic regression through stacking. Comparative analysis is done and shows that proposed model does better than other models [7].

Gu et, al., in his article ranked pre-processed feature by ANOVA and Mutual Information. Further ranking is integrated using fuzzy inferences. Backward elimination feature selection method eliminates irrelevant features. Then dataset is modelled using Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), and Multilayer Perceptron (MLP), finally metamodel is settled. RMSE and MAPE is used to evaluate the performance which gives 0.06% and 0.03% [8].

T. S. Devkishan in his study ensembles Multilayer Perceptron, Decision Tree, Logistic Regression, KNN, Random Forest and SVM, Naive Bayes. The proposed mode shows perfect score (all 1.00) [9]

PROPOSED METHODOLOGY

A. Dataset

The dataset used is secondary dataset from Kaggle. The dataset used is from the college students “Higher education students performance evaluation” [10]. The faculty of engineering and of educational sciences learners studying at undergraduate level in 2019 data is utilized. The dataset is categorized under personal information of student, Demographic information, family background, academic performance. Dataset consists of 32 attributes.

B. Data Preparation

After importing the dataset, the data is pre-processed which is the second stage in developing a model. Pre-processing is performed to identify and remove null values and outliers. The supervised learning models is trained and tested in proportion of (80:20). Further cross validation is performed on hybrid model to train and test. The code is executed in python using google colab.

C. Evaluation of Training (Tnn) and Testing (Tsn)

Decision Tree classifier, Gradient Boosting model and Random Forest are trained and tested with proportion of (80:20). The dataset undergoes training (Tnn) and testing (Tsn) with cross validation method on fusion model. Recall, Accuracy, Precision, and F1 score are the performance measure utilized to predict the results.

D. Algorithm

```
Step 1: Read the dataset
Step 2: Apply preprocessing with splitting the data
Step 3: Train and test the model
Step 4: Model the dataset with the Decision Tree classifier, Gradient Boosting model
        and Random Fores
Step 5: Evaluate the model by ACCRc, Prc, Rec, F1 score
Step 6: Compare the values of Accuracy of different models and analyze it.
Step 7: Apply the voting classifier on the estimators of the Decision Tree classifier,
        Gradient Boosting model and Random Forest
Step 8: Evaluate the hybrid model by Accuracy, Precision, Recall, F1 score
Step 9: Stop
```

E. Working of the Model

Figure 1 gives the proposed framework. The data gathering is performed at the initial stage. Preprocessing is the next stage that is performed. Data splitting is performed and the dataset is modelled with Decision Tree classifier, Gradient Boosting model and Random Forest. Comparative analysis takes place with the performance metric including, Recall, Accuracy, Precision and F1 score. To improve the accuracy all the three models are ensembled. The estimators of all the three models are taken into consideration and voting classifier is applied. Hard voting classifier takes place. The hybrid model further is evaluated with Recall, Accuracy, Precision and F1 score.

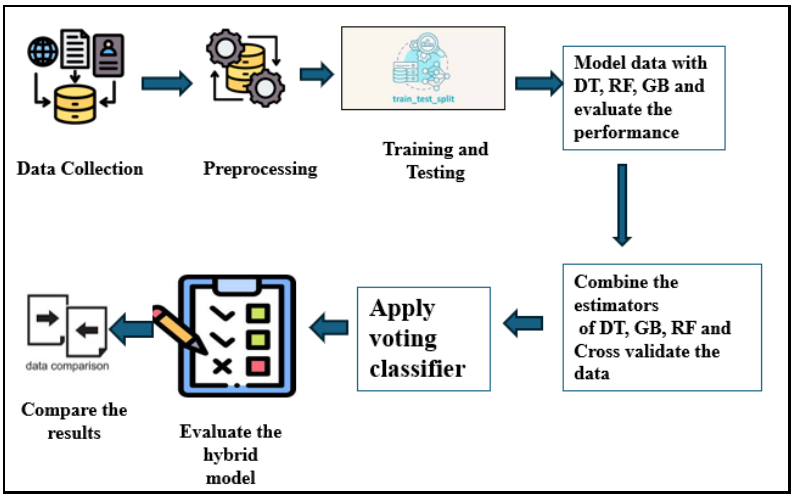
Data Collection
Preprocessing
train_test_split
Training and Testing
Model data with DT, RF, GB and evaluate the performance
Combine the estimators of DT, GB, RF and Cross validate the data
Apply voting classifier
Evaluate the hybrid model
data comparison
Compare the results

Figure 1: Working of the Model

EXPERIMENTAL METHODOLOGY

A. Classification techniques in Machine Learning

1. Decision Tree

Decision tree is a supervised learning model which helps to take decision by selecting the best choice from alternatives and mapping with the outcome.

Root node represents the initial stage in the entire dataset. Branches are the connectors that represents flow between the decisions. Internal node denotes the the stage where decisions are made with the help of data features whereas leaf nodes last stage where final predictions are made.

2. Random Forest

Decision Tree is single tree whereas random forest is the combination of multiple trees. It is an ensemble tree. It provides more accurate predictions. It predicts the numerical values. With the help of bagging or boosting it combines the results and make better prediction.

3. Gradient Boosting

One of the boosting algorithms include gradient boosting in which each new model trained minimizes the loss job. It trains the new weedy model by computing the gradient loss function as compared with predictions at every iteration. All the predictions are further combined which forms ensemble until the stopping criteria is met.

4. Voting Classifier

It is the type of ensemble technique that achieves experience through training on combinations of models and predict an outcome based on the model having highest resultant output. This becomes possible by averaging the output of each classifier. It is single model that learns through the combinations of models. It can be said as hybrid model.

B. Performance Metrics

The research paper focusses on evaluating the performance with the following performance metrics of Classification:

The performance metrics used for evaluation are Accuracy (ACCRr), Precision (Prc), Recall (Rec), and F1-Score(F1-S).

Thes metrics uses the parameters like True_Positive, True_Negative, False_Positive, False_Negative. Assume True_Positive = A, True_Negative = B, False_Positive = C, False_Negative = D.

They are calculated as follows

Accuracy $=\frac{A+B}{A+B+C+D}$ ---------------------------------------(1)

Precision (Prc) = $\frac{(A)}{(A+C)}$-------------------------------------(2)

Recall (Rec) = $\frac{(A)}{(A+D)}$ --------------------------------------(3)

F1-Score=2 $*\frac{Prc*Rec}{Prc+Rec}$ --------------------------------------(4)

RESULT ANALYSIS

The evaluation of the model applied on the dataset is used for comparative analysis. Performance metrics of classification is evaluated for DT, RF, GB and using voting classifier. Table 2 depicts the evaluation of classifiers using Recall, Accuracy, Precision, and F1 score.

Table1: Evaluation of Model

Classifier	Decision Tree (D_T)	Random Forest (R_F)	Gradient Boosting (G_B)	Voting Classifier (V_C)
Accuracy	0.76	0.75	0.76	0.86
Precision_ 0	0.8	0.77	0.8	0.89
Precision_ 1	0.67	0.71	0.67	0.82
Recall _0	0.84	0.89	0.84	0.89
Recall _1	0.6	0.5	0.6	0.82
F1-Score_ 0	0.82	0.83	0.82	0.89
F1-Score_ 1	0.63	0.59	0.63	0.82

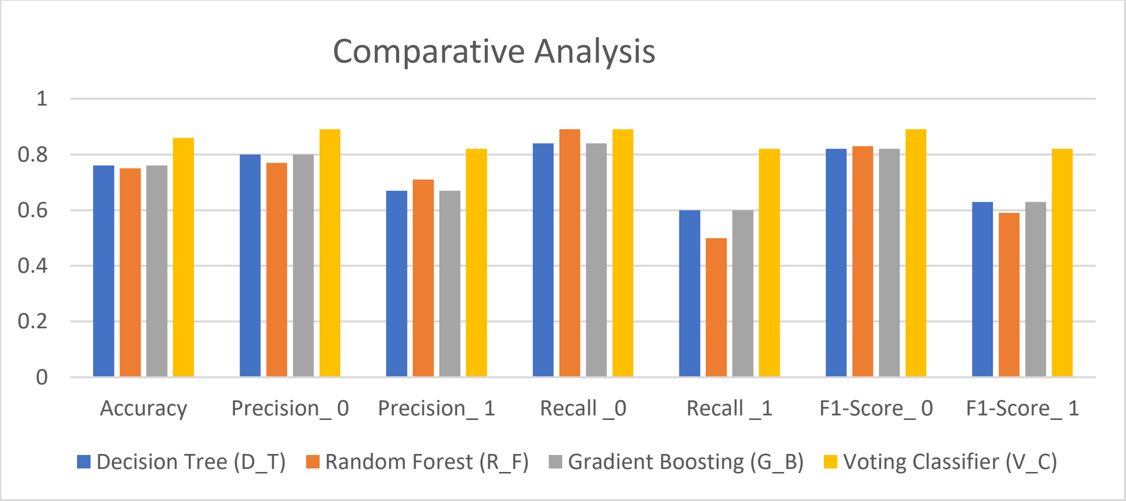

Figure2: Comparative Analysis of Model

Assessment of the table

Table1 shows the evaluation of the classifiers. The models like Decision Tree classifier, Gradient Boosting model and Random Forest show the accuracy with 76%, 75% and 76%. The accuracy of DT, RF and Gb is acceptable but it is not considered an optimal. The estimators of DT, RF and Gb are combined and voting classifier is applied which gives accuracy of 86%. Voting classifier proves to be best classifying model. It has high and balanced precision, recall and F1 score. The recall, precision, F1-Score is calculated for both classes i. e class 0 and class1.The ensemble model with voting classifier gives more efficient results.

CONCLUSION

From the findings the voting classifier is the best performing model with 86% accuracy. It outperforms all the models. It illustrates strong generality by merging the strengths of multiple classifiers. Ensemble techniques provide good predictions. It gives balances class wise performance creating suitable in both cases of false positives and false negatives. For deployment or final reporting, the Voting Classifier should be selected as the optimal model due to its superior accuracy, robustness, and consistent performance across classes.

REFRENCES

[1] Daud, Ali & Aljohani, Naif & Abbasi, Rabeeh & Lytras, Miltiadis & Abbas, Farhat & Alowibdi, Jalal. (2017). Predicting Student Performance using Advanced Learning Analytics. 10.1145/3041021.3054164.

[2] Rehman, Muhammad & Iftikhar, Asim & Muhammad, Saghir & Ahmed, Rizwan. (2025). Student Academic Performance Prediction using Ensemble Learning Methods. Journal of ICT, Design, Engineering, and Technological Science. 9. 10.33150/JITDETS-9.1.2.

[3] M. G. K. Reddy and N. Sharma, "Enhancing Student Performance Predictions Using Ensemble Models and Hyperparameter Tuning," 2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI),

[4] Amani Khalifa, Fatma Ben Said, Yessine Hadj Kacem, Optuna-Optimized Meta-Learner and Ensemble Learning Models for Student Performance Prediction, Procedia Computer Science, Volume 270, 2025, Pages 1826-1835, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2025.09.303.Gwalior, India, 2025, pp. 1-6, doi: 10.1109/IATMSI64286.2025.10985712.

[5] Kumar, M., Bhardwaj, V., Thakral, D., Rashid, A., Ben Othman, M.T. (2024). Ensemble learning based model for student's academic performance prediction using algorithms. Ingénierie des Systèmes d'Information, Vol. 29, No. 5, pp. 1925-1935. https://doi.org/10.18280/isi.290524

[6] G. Nassreddine, L. Saleh, M. A. Majzoub and A. E. Arid, "SHAP Explainability: An Ensemble Learning Approach for Student Performance Prediction," 2025 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), Bali, Indonesia, 2025, pp. 432-438, doi: 10.1109/IAICT65714.2025.11100707.

[7] Yan, Lijuan & Liu, Yanshen. (2020). An Ensemble Prediction Model for Potential Student Recommendation Using Machine Learning. Symmetry. 12. 728. 10.3390/sym12050728.

[8] Gu, J. Predicting student academic achievement using stacked ensemble learning with deep neural networks and fuzzy-based feature selection. Sci Rep 15, 37195 (2025). https://doi.org/10.1038/s41598-025-20779

[9] T. S. Devkishan, S. K. Singh and A. K. Bharti, "Ensemble Learning for Student Performance Assessment: Identifying and Analyzing Significant Affecting Factors in Higher Education," 2024 7th International Conference on Contemporary Computing and Informatics (IC3I), Greater Noida, India, 2024, pp. 271-277, doi: 10.1109/IC3I61595.2024.10829365.

[10] Higher Education Students Performance Evaluation dataset: https://archive.ics.uci.edu/dataset/856/higher+education+students+performance+evaluation, accessed on October 2025