# An approach for improved students' performance prediction using homogeneous and heterogeneous ensemble methods

**Edmund De Leon Evangelista[1], Benedict Descargar Sy[2]**
[1]Information Systems and Technology Management Department, Faculty of College of Technological Innovation, Zayed University (Abu Dhabi Campus), Zayed City, United Arab Emirates
[2]Information Technology Department, Faculty of School of Engineering, Architecture, and Information Technology Education, University of Saint Louis Tuguegarao, Tuguegarao City, Philippines

**Article Info**

***Article history:***

Received Nov 13, 2021
Revised May 18, 2022
Accepted Jun 14, 2022

***Keywords:***

Bagging and boosting
Ensemble methods
Machine learning
Student performance prediction
Weka experiment

**ABSTRACT**

Web-based learning technologies of educational institutions store a massive amount of interaction data which can be helpful to predict students' performance through the aid of machine learning algorithms. With this, various researchers focused on studying ensemble learning methods as it is known to improve the predictive accuracy of traditional classification algorithms. This study proposed an approach for enhancing the performance prediction of different single classification algorithms by using them as base classifiers of homogeneous ensembles (bagging and boosting) and heterogeneous ensembles (voting and stacking). The model utilized various single classifiers such as multilayer perceptron or neural networks (NN), random forest (RF), naïve Bayes (NB), J48, JRip, OneR, logistic regression (LR), k-nearest neighbor (KNN), and support vector machine (SVM) to determine the base classifiers of the ensembles. In addition, the study made use of the University of California Irvine (UCI) open-access student dataset to predict students' performance. The comparative analysis of the model's accuracy showed that the best-performing single classifier's accuracy increased further from 93.10% to 93.68% when used as a base classifier of a voting ensemble method. Moreover, results in this study showed that voting heterogeneous ensemble performed slightly better than bagging and boosting homogeneous ensemble methods.



***Corresponding Author:***

Edmund De Leon Evangelista
Department of Information Systems and Technology Management, College of Technological Innovation, Zayed University
Abu Dhabi Campus, United Arab Emirates
Email: undoevangelista@gmail.com, edmund.evangelista@zu.ac.ae

## 1. INTRODUCTION

Educational institutions utilize various web-based learning technologies to enhance their teaching and learning experience. It can take many forms, including massive open online courses (MOOCs), virtual learning environments (VLEs), and learning management systems (LMS) [1]. Students' have left a large amount of online learning data on these learning technologies that effectively predict students' academic performance and allow for pre-intervention of at-risk students [2]. Student performance is an essential part of the learning process, and predicting it is vital to identify students who are more likely to struggle academically in the future [3]. The majority of existing prediction models are built with machine learning techniques [4], which can discover variables that significantly influence students' performance, dropout, engagement, and interaction in online learning platforms [5].

Performance prediction requires an efficient approach in building a model that provides accurate predictions [6]. A traditional approach to achieve this is to employ several classification algorithms, compare the results, and select the single classifier with the highest predictive accuracy [7], [8]. However, single classifiers are commonly plagued by overfitting and local optimum issues, and they have remained an active research topic for performance enhancement [9]. This resulted in the emergence of ensemble learning, a learning method in which a collection of a finite number of single classifiers is trained for the same classification task to improve performance at the expense of computation [10]. The goal is to create a composite global model that can make more accurate and reliable decisions than the best model in the set [11], [12]. The advantage of ensembles is not that the best combination of classifiers outperforms the best classifier but that a combination of classifiers is less likely than a single classifier to misclassify unseen data samples [11].

Ensembles are classified into two types: i) homogeneous ensembles (HE) in which the base learning model is built using the same learning algorithm and ii) heterogeneous ensembles (HTE) in which different learning algorithms are combined to generate the base learning models [13]. Bagging and boosting algorithms are two well-known HE [14], while voting and stacking algorithms belong to HTE [15]. Much of the prior research on ensemble learning models have concentrated on HE [16]–[18] using a specific type of classifier or regressor [19]. However, this approach of relying alone on HE limits the capability of the model to achieve diversity and maximum predictive accuracy. No single classification algorithm is considered optimal for all cases, and only by combining various single classifiers can classification performance be improved [20], [21]. The main goal of ensembles is to improve generalization and diversity among the models to deal with dataset variance, and only HTE can achieve this better because it uses a diverse set of base classifiers [14], [22], [23]. Experimental results of various studies demonstrated the effectiveness of using HTE with significantly improved performance when compared to HE [15], [24]–[26].

Identifying the base classifiers to be used is a common challenge with HTE, also known as hybrid ensembles. Classifier performance varies across datasets, making it challenging to select the collection of classifiers that will best classify a given set of data [16]. In solving these issues, this study proposed an improved approach of predicting students' performance by using parallel prediction of bagging and boosting algorithms for HE and voting and stacking algorithms for HTE, respectively. The primary aim of the proposed work is to perform a comparative analysis of students' performance prediction using both types of ensembles and to use whichever ensemble type responds well on a given dataset.

The improved approach utilized base classifiers that perform well on the dataset and it eliminated weak classifiers to achieve better prediction performance. It involved three significant activities in constructing the final predictive model: i) various single classifiers such as multilayer perceptron (MLP), random forest (RF), naïve Bayes (NB), J48, JRip, OneR, logistic regression (LR), k-nearest neighbor (KNN), and support vector machine (SVM) would be trained to determine the best-performing and relatively top-performing classifiers on the given dataset. Weak classifiers were eliminated in the framework; ii) the base classifier of bagging and boosting HE would be the best-performing classifier among the single classifiers identified in the previous step. In contrast, the set of base classifiers for voting and stacking HTE would be the relatively top-performing classifiers; iii) these ensemble methods (bagging, boosting, voting, stacking) and their chosen base classifier/s were trained parallelly. Finally, the final predictive model for the future dataset was selected based on the ensemble method gaining the highest accuracy with the lowest misclassified instances.

This approach provided an equal opportunity for both types of ensembles to be trained and to determine when is the best time to use one over the other based on a given dataset. The open-source Weka Experimenter machine learning tool simplified the parallel training of various base classifiers and ensemble methods. The experimenter feature of Weka allowed performing parallel or "batches" of experiments on different algorithms, datasets, and parameters, thereby allowing to collect statistical comparisons of performance and to build models at a much lesser time. The remainder of this paper is divided into four sections. Section 2 presents background and related works. Section 3 presents the materials and methods. Finally, section 4 discusses the obtained results, and section 5 covers the conclusion.

## 2. BACKGROUND AND RELATED WORKS

This section briefly introduces an overview of ensemble learning methods, related works, and prediction approaches. The main goal of an ensemble classifier is to take advantage of the benefits of multiple classifiers and combine their outputs so that the predictive accuracy of the model improves [23]. The individual classifiers in an ensemble system are referred to as base classifiers. Ensembles are also known as a combination of experts, committees, multi-classifier systems, expert fusion, selection, or thinning [27]. It integrates a set of models used to solve different tasks to create an enhanced composite global model that produces more accurate and reliable estimates than a single model. It combines different machine learning techniques into a single predictive model to reduce variability (bagging), bias (boosting), or improve results (stacking and voting) [28].

Building HTE, which are made up of models generated by different base classifiers, rather than HE, which are made up of models developed by the same base classifier is one way to improve ensemble performance vastly [22]. When an ensemble is built with homogeneous models, neither high accuracy nor diversity will cause the ensemble to outperform the traditional individual classifiers [27]. Only with HTE that diversity can be achieved while maintaining the high accuracy of the models. Diversity occurs when various base classifiers are introduced into the ensemble model. It refers to the disagreement of these base classifiers over a set of data instances covering a wide range of the feature space because some base classifiers may make errors in different situations; strategically combining them can reduce total error [24]. There is no point in using an ensemble if there is no diversity because the output of the ensemble classifier will be identical to the output of each of the base classifiers.

The primary concern is finding the best ensemble combination (set of base classifiers) and approach to achieve diversity [14]. Many researchers proposed several measures to identify the “best” base classifiers among a set of alternatives, including accuracy, training time, classification time, and storage [20]. Luong *et al.* [29] proposed an ensemble system design that combines homogeneous and heterogeneous modules in a single framework using random projections and different learning algorithms. Experiments on well-known datasets show that the proposed ensemble system outperforms several well-known benchmark algorithms significantly.

Ostvar and Moghadam [23] introduced a heterogeneous dynamic ensemble classifier (HDEC), which used multiple classification algorithms and tested it on 12 standard datasets from the University of California Irvine (UCI) repository. They compared the performance of their proposed method to three cutting-edge ensemble approaches namely bagging, boosting, and stack generalization. The obtained results show that their proposed method is superior in terms of accuracy and geometric mean values.

For classifying spam emails, Wang [30] created a heterogeneous ensemble and a framework for constructing various ensembles. When compared to individual classifiers and other ensemble models, the results show that the heterogeneous ensemble can increase diversity and performance. Similarly, Haque *et al.* [14] suggested a genetic algorithm-based search method for finding the optimum combination from a pool of base classifiers to form a heterogeneous ensemble. Results of the proposed method show that genetic algorithm is a superior and reliable approach to heterogeneous ensemble construction.

In the same way, Shashank and Mahapatra [31] introduced a workflow for boosting prediction accuracy using a weighted heterogeneous ensemble of multiple machine learning classification algorithms to solve the classical problem of supervised rock-facies classification from well logs. Rather than attempting to find the perfect classifier for rock facies classification, they chose a weighted ensemble of three base-level classifiers: extremely randomized trees, support vector machines, and gradient boosting algorithms to obtain a more generalized predictor.

On the contrary, Espinoza *et al.* [32] assessed the performance of approaches based on homogeneous ensemble learning in detecting bogus online information. The application of several ensemble learning-based approaches to a collection of fabricated restaurant reviews developed by the researchers demonstrated that ensemble learning-based methods intercepted deceptive information better than conventional machine learning algorithms. In the same way, Gamie *et al.* [33] proposed a homogeneous ensemble model with boosting and other traditional classification algorithms on the OULAD education dataset to determine at-risk students. Their study also utilized brute force analysis to detect the best combination of feature subgroups. Results show that the ensemble model increased the classification accuracy by more than 85% compared to traditional classifiers.

The studies highlighted in this section proved the dominance of ensemble models when compared to traditional single classifiers. Furthermore, among the two types of ensembles learning methods, various studies confirmed that HTE performed better in improving the predictive accuracy of models when compared to HE. However, this study aimed to utilize both HE and HTE to propose an improved approach for ensemble learning by providing a comparative analysis of their performance. The bagging and boosting HE used the best performing base classifier while voting and stacking HTE utilized the relatively top-performing base classifiers on a given dataset. This improved approach maximized the proper use of ensembles and let the framework decide which ensemble learning method best works on a given dataset.

## 3. MATERIALS AND METHODS

### 3.1. Data collection and pre-processing

This study used an open-access student performance dataset from the UCI machine learning repository to test the proposed model. If interested, the dataset may also be accessed here [34]. It included

649 student records and 32 attributes from two Portuguese secondary schools, containing student grades, social, demographic, and school-related information. Two datasets were available from this repository; the first dataset contained student performance data from a Mathematics lesson, while the second included a Portuguese lesson. This study made use of the second dataset only.

The dataset had undergone some pre-processing for data cleaning purposes, particularly the final grade attribute (target class). Since this is a classification task, the original numeric values (1-20) of this target class were transformed into nominal 'P' (>=10) and 'F' (<10) values. In addition, machine learning works well with numerical data; hence the other features containing nominal values ('Yes' or 'No') need to be converted into binary values (1 and 0).

### 3.2. Feature selection techniques

A feature selection technique, also known as evaluators, determined the best predictors or attributes of the training dataset highly correlated to the target class [35]. This study utilized info gain and chi-squared feature selection techniques to determine the best features associated with the target class (final grade). Info gain attribute eval measured the information gain, also called entropy, concerning the target class to assess the worth of an attribute. The value of an entry ranged from 0 (no information) to 1 (maximum information). Those attributes that contributed more information have a higher information gain value and got chosen. In contrast, those that did not contribute much information have a lower score and got removed.

On the other hand, chi-squared attribute eval computed the value of the chi-squared statistic relating to the target class to evaluate the value of a feature. It assessed the extent to which attributes contributed to the formation of clusters. As a result, the features were ranked on how much they contribute to the target class.

### 3.3. Model implementation

The main goal of this study was to provide an improved approach in predicting students' performance using homogeneous (HE) and heterogeneous (HTE) ensemble learning methods, as shown in Figure 1. The open-source WEKA machine learning tool was utilized to perform feature selection techniques and to build the proposed model. WEKA is a free and open-source data mining software that provides tools for tasks such as regression, classification, association rules mining, clustering, and visualization [36].

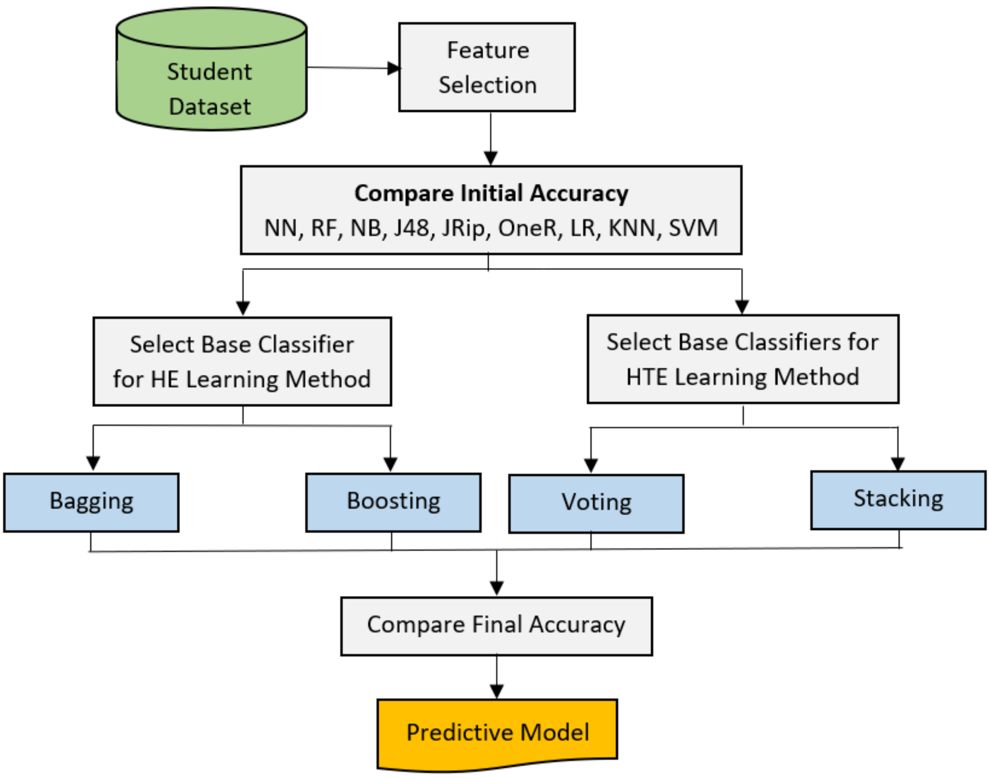


Figure 1. The proposed homogeneous and heterogeneous ensemble framework

The model started by implementing info gain attribute eval and chi-squared attribute eval feature selection techniques on the given dataset to determine the highly correlated attributes to the target class. Then, the model utilized various single classifiers such as MLP or neural networks (NN), RF, NB, J48, JRip, OneR, LR, KNN, and SVM to determine the base classifiers of the ensembles. Note that employing various classification algorithms in the framework aimed to achieve diversity while maintaining improved predictive accuracy.

The single classifiers were trained in parallel using Weka experimenter. Then, the comparative analysis of their accuracy was used to determine the best and relatively top-performing algorithms, respectively, as shown in a sample setup in Figure 2. The parallel training of models was performed based on 10-folds cross-validation and default parameters available in the configuration of Weka experimenter.

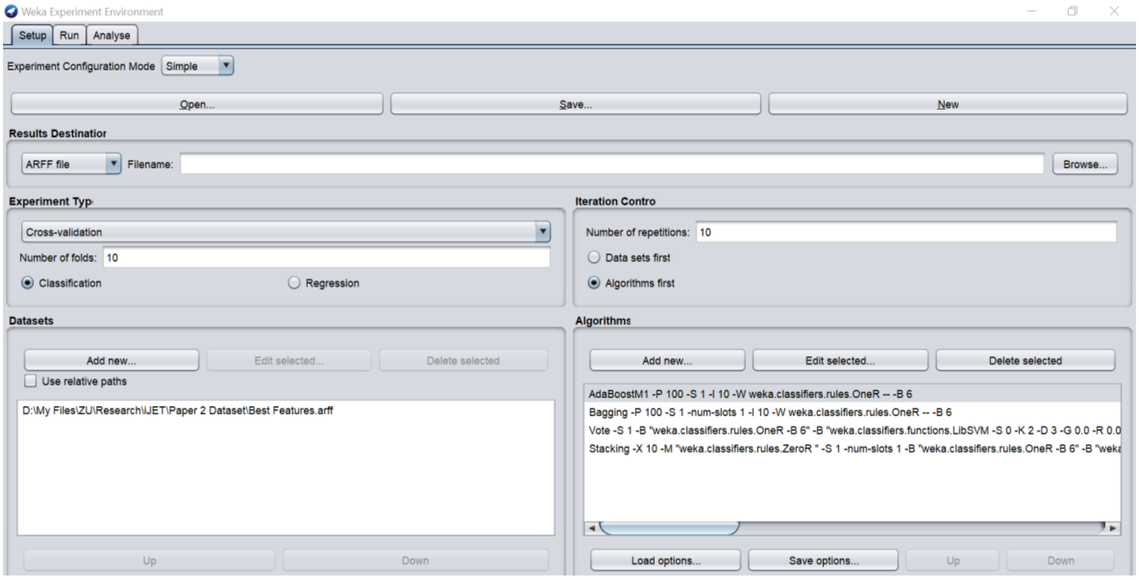


Figure 2. Weka experimenter setup for comparative analysis of ensemble methods

Weak classifiers having the lowest accuracy and lowest F-score were eliminated in the framework. bagging and boosting HE used the best algorithm as its base classifier. In contrast, voting and stacking HTE used the top-performing algorithms (with the highest accuracy and F-score) as its base classifiers. Figure 3 shows the setup on adding multiple base classifiers in Weka for voting and stacking HTE, respectively.

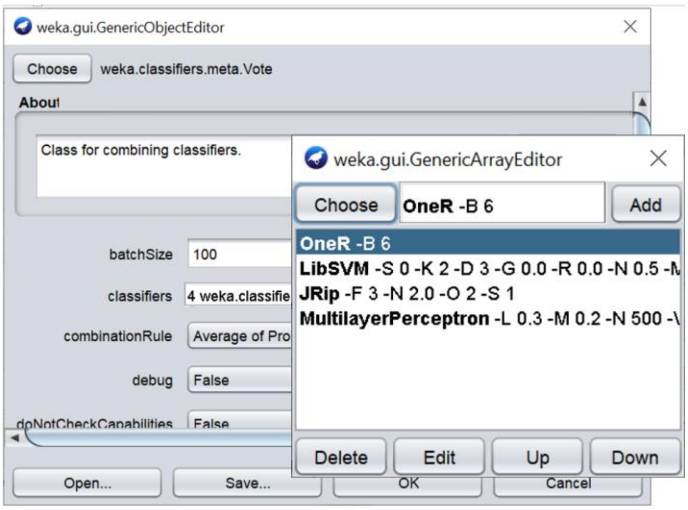


Figure 3. Weka experimenter setup for adding of multiple base classifiers

In bagging, each base classifier was trained with a bootstrapped replica of the training dataset [23]. Then, the final decision was generated by applying a majority voting on each base classifier's decision. On the other hand, boosting builds the base classifiers iteratively, each compensating for the shortcomings of its predecessors. Meanwhile, voting creates two or more sub-models, then each sub-model makes predictions, which are then combined in some way to get the mean or mode of the predictions [6]. In the same way, stacking combines predictions from multiple models to create a new model, which is then used to make predictions on the test dataset [19]. Thus, it aims to improve a classifier's predictive power. The four (4) ensemble methods and their base classifiers would undergo parallel training in the final step. The final predictive model would then use the ensemble method that generates the highest predictive accuracy with the lowest misclassified instances.

## 4. RESULTS AND DISCUSSION

This study used an open-access student performance dataset from the UCI machine learning repository. The target class was final grade which contained a nominal value of P (pass) or F (fail). All experiments were carried out using the Weka machine learning tool.

### 4.1. Top features selected

Figure 4 reveals the ranking of the top attributes selected based on merit by the feature selection techniques used in this study. As shown in Figure 4(a), only the attributes G2, G1, failures, higher, and study time contributed more information gain than all of the other attributes of the dataset. Likewise, Figure 4(b) selected the same attributes, confirming that they are the most compelling features of the target class. Both evaluators agreed that out of 32 features of the original dataset, only the attributes G2, G1, failures, higher, and study time have a significant degree of relationship with the target class.

| average | merit | average | rank | attribute |
|---|---|---|---|---|
| 0.387 | +- 0.007 | 1 | +- 0 | 20 G2 |
| 0.324 | +- 0.007 | 2 | +- 0 | 19 G1 |
| 0.089 | +- 0.005 | 3 | +- 0 | 3 failures |
| 0.053 | +- 0.006 | 4 | +- 0 | 9 higher=no |
| 0.018 | +- 0.004 | 5 | +- 0 | 2 studytime |
| 0 | +- 0 | 8.9 | +- 1.58 | 4 schoolsup=no |
| 0 | +- 0 | 9.5 | +- 1.12 | 6 paid=yes |
| 0 | +- 0 | 9.8 | +- 2.99 | 8 nursery=no |
| 0 | +- 0 | 9.9 | +- 1.81 | 7 activities=yes |
| 0 | +- 0 | 10.7 | +- 1.55 | 5 famsup=yes |

(a)

| average | merit | average | rank | attribute |
|---|---|---|---|---|
| 367.193 | +- 7.923 | 1 | +- 0 | 20 G2 |
| 301.677 | +- 7.882 | 2 | +- 0 | 19 G1 |
| 92.157 | +- 4.848 | 3 | +- 0 | 3 failures |
| 56.251 | +- 6.834 | 4 | +- 0 | 9 higher=no |
| 15.048 | +- 2.541 | 5.3 | +- 0.46 | 2 studytime |
| 0 | +- 0 | 8.9 | +- 1.58 | 4 schoolsup=no |
| 0 | +- 0 | 9.5 | +- 1.12 | 6 paid=yes |
| 0 | +- 0 | 9.8 | +- 2.99 | 8 nursery=no |
| 0 | +- 0 | 9.9 | +- 1.81 | 7 activities=yes |
| 0 | +- 0 | 10.7 | +- 1.55 | 5 famsup=yes |

(b)

Figure 4. Ranking of top attributes based on merit using (a) info gain attribute evaluator and (b) chi-squared attribute evaluator

Similarly, Table 1 shows the attribute description and type of data, along with possible values of the processed dataset. The proposed model used the top-ranked features identified by the evaluators. To save some space but still be able to visualize the original dataset, you may opt to view the complete 32 features here [35].

Table 1. Description of dataset attributes

| Feature | Description | Type | Values |
|---|---|---|---|
| G1 | First period grade | Numeric | 0-20 |
| G2 | Second period grade | Numeric | 0-20 |
| Failures | Number of past class failures | Numeric | 0-3 |
| Higher | Wants to take higher education | Numeric | 0-1 |
| Study time | Home to school travel time | Numeric | 1-4 |
| FG | Final grade | Nominal | P and F |

### 4.2. Performance accuracy of the single classifiers

Table 2 shows the performance of the trained models using single classifiers and their accuracy when used as a base classifier of bagging and boosting HE. The setup was based on 10-folds cross-validation and default parameters of the algorithms in Weka. Cross-validation is a statistical method for assessing and comparing learning algorithms that divide data into two parts: learning or training a model and validating the model [36].

Based on the experiments shown in Table 2, the OneR classifier gained the highest accuracy of 93.10% among the other single classifiers. Moreover, it is the only algorithm in this experiment that gained a consistent highest accuracy of 93.62% and 93.08% when used as base classifier of bagging and boosting ensemble methods respectively. Therefore, it would serve as the base classifier for bagging and boosting HE of this study.

Table 2. Performance of the single classifiers and its HE accuracy

| Algorithm | Accuracy | HE Accuracy | |
|---|---|---|---|
| | | Bagging | Boosting |
| Multilayer perceptron (NN) | 92.63% | 92.45% | 92.43% |
| Random forest (RF) | 91.85% | 92.00% | 91.43% |
| Naïve Bayes (NB) | 90.89% | 90.88% | 92.22% |
| J48 | 92.36% | 92.72% | 91.57% |
| JRip | 93.00% | 92.86% | 91.75% |
| OneR | 93.10% | 93.62% | 93.08% |
| Support vector machine (SVM) | 93.02% | 92.76% | 91.52% |
| K-nearest neighbor (KNN) | 90.06% | 89.60% | 89.41% |
| Logistic regression (LR) | 92.48% | 92.54% | 92.48% |

Furthermore, this study used the top four performing classifiers (OneR, SVM, JRip, and NN) as the base classifiers of voting and stacking HTE. In addition, this study also utilized the top, middle, and bottom performing classifiers (OneR, LR, and KNN) for further comparative analysis to determine if it would impact the performance of the ensemble model if the classifiers would come from varying accuracies. Accordingly, this combination was referred to as voting2 and stacking2 HTE in this study.

### 4.3. Performance accuracy of HE and HTE learning methods

Figure 5 compared the accuracies of the various ensemble learning methods implemented in this study based on 10-folds cross-validation and default parameters of the ensemble methods in Weka. It can be seen that voting HTE achieved the highest accuracy, which confirms the common observation that heterogeneous ensembles perform better than homogeneous ensembles. However, it can also be observed that boosting HE is just slightly lower than voting HTE but way higher than the stacking HTE. This observation agrees with the idea proposed in this study that when dealing with ensembles, both HE and HTE should be given equal chances to be trained to determine which type of ensemble performs well in any given dataset. Moreover, it can be observed that voting and boosting HTE using top-performing base classifiers perform better than voting v2 and stacking v2 HTE that uses top, middle, and weak performing base classifiers. It means that when dealing with HTE, top-performing base classifiers perform better when used as the base classifiers of ensemble methods.

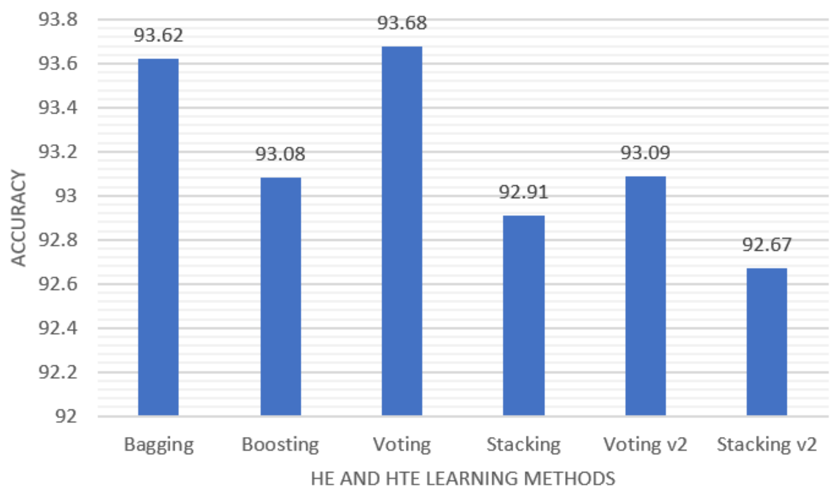


Figure 5. Performance accuracy of HE and HTE learning methods

### 4.4. Performance accuracy of single classifiers and ensemble methods

Figure 6 demonstrates the performance accuracy of single classifiers and ensemble methods. It shows that bagging and voting ensemble methods outperformed the other single classifiers. It is worth mentioning that the OneR algorithm gained the highest accuracy of 93.10% among all single classifiers used in this study. Similarly, the bagging ensemble has used the same algorithm as its base classifier and further increased the accuracy to 93.62%. It confirms the observation of most studies that ensembles indeed improve the accuracy of single classifiers when used as its base classifier.

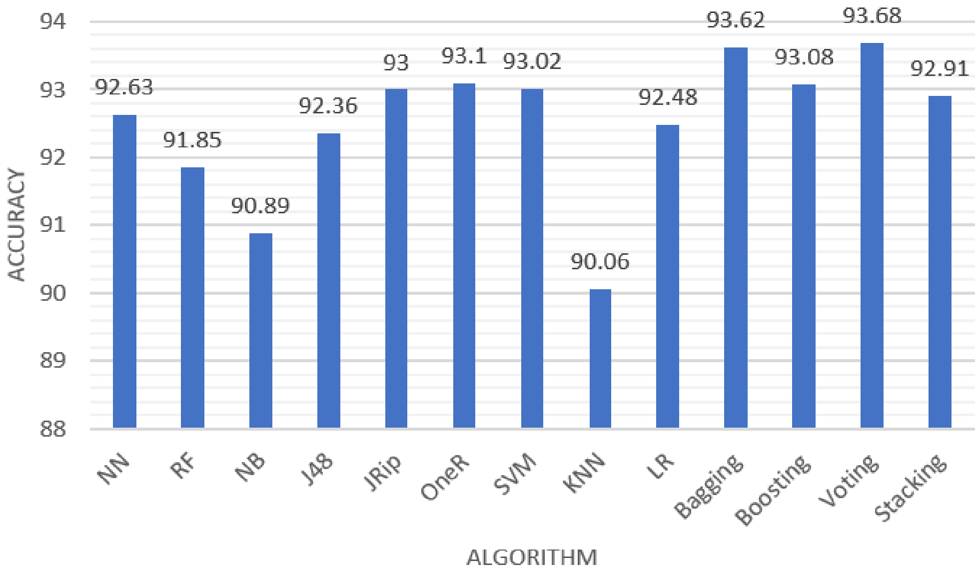


Figure 6. Performance accuracy of single classifiers and ensemble methods

## 5. CONCLUSION

This study aimed to compare students' performance prediction using homogeneous and heterogeneous ensembles and to adopt whichever ensemble type performs well on a given dataset. It provided an approach for utilizing single classifiers to improve its accuracy by using them as base classifiers of bagging, boosting, voting, and stacking ensemble learning methods. The comparative analysis of the model's accuracy showed that the best-performing single classifier's accuracy increased further from 93.10% to 93.68% when used as a base classifier of a voting ensemble method. Moreover, results in this study showed that voting heterogeneous ensemble performed slightly better than bagging and boosting homogeneous ensemble methods. It simply means that when dealing with ensemble methods, no specific type of ensemble is better than the other ensembles. Therefore, there is a need to train and select the ensemble method that performs well on a given dataset. However, for both types of ensemble methods to achieve improved performance accuracy, one must select the appropriate base classifiers from a set of single classifiers. The enhanced approach proposed in this study proved that top-performing single classifiers are good candidates as base classifiers of an ensemble method. Future work may include testing the framework with other datasets to explore different ways to improve its performance. In addition, the framework needs to perform optimization techniques to find out the algorithm parameters and configuration that will maximize the potential that any ensemble methods may achieve.

## REFERENCES

[1] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi, "Student engagement predictions in an e-learning system and their impact on student course assessment scores," *Computational Intelligence and Neuroscience*, vol. 2018, pp. 1–21, Oct. 2018, doi: 10.1155/2018/6347186.

[2] X. Li, X. Zhu, X. Zhu, Y. Ji, and X. Tang, "Student academic performance prediction using deep multi-source behavior sequential network," in *Advances in Knowledge Discovery and Data Mining*, Springer International Publishing, 2020, pp. 567–579.

[3] N. M. Aslam, I. U. Khan, L. H. Alamri, and R. S. Almuslim, "An improved early student's academic performance prediction using deep learning," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 12, Jun. 2021, doi: 10.3991/ijet.v16i12.20699.

[4] L. Yan and Y. Liu, "An ensemble prediction model for potential student recommendation using machine learning," *Symmetry*, vol. 12, no. 5, May 2020, doi: 10.3390/sym12050728.

[5] M. Adnan *et al.*, "Predicting at-risk students at different percentages of course length for early intervention using machine learning models," *IEEE Access*, vol. 9, pp. 7519–7539, 2021, doi: 10.1109/ACCESS.2021.3049446.
[6] E. D. Evangelista, "Development of machine learning models using study behavior predictors of students' academic performance through moodle logs," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 6S3, pp. 22–27, 2019.
[7] T. Anderson and R. Anderson, "Applications of machine learning to student grade prediction in quantitative business courses," *Global Journal of Business Pedagogy*, vol. 1, no. 3, pp. 13–22, 2017.
[8] X. Fan, C.-H. Lung, and S. A. Ajila, "Using hybrid and diversity-based adaptive ensemble method for binary classification," *International Journal of Intelligence Science*, vol. 08, no. 03, pp. 43–74, 2018, doi: 10.4236/ijis.2018.83003.
[9] A. O. M. Abuassba, D. Zhang, X. Luo, A. Shaheryar, and H. Ali, "Improving classification performance through an advanced ensemble based heterogeneous extreme learning machines," *Computational Intelligence and Neuroscience*, vol. 2017, pp. 1–11, 2017, doi: 10.1155/2017/3405463.
[10] L. Yang, "Classifiers selection for ensemble learning based on accuracy and diversity," *Procedia Engineering*, vol. 15, pp. 4266–4270, 2011, doi: 10.1016/j.proeng.2011.08.800.
[11] K.-W. Hsu, "A theoretical analysis of why hybrid ensembles work," *Computational Intelligence and Neuroscience*, vol. 2017, pp. 1–12, 2017, doi: 10.1155/2017/1930702.
[12] A. Satyanarayana and M. Nuckowski, "Data mining using ensemble classifiers for improved prediction of student academic performance," *CUNY Academic Works*, 2016.
[13] Y. Chali, S. A. Hasan, and M. Mojahid, "Complex question answering: homogeneous or heterogeneous, which ensemble is better?," in *Natural Language Processing and Information Systems*, Springer International Publishing, 2014, pp. 160–163.
[14] M. N. Haque, N. Noman, R. Berretta, and P. Moscato, "Heterogeneous ensemble combination search using genetic algorithm for class imbalanced data classification," *Plos One*, vol. 11, no. 1, e0146116, Jan. 2016, doi: 10.1371/journal.pone.0146116.
[15] S. Tewari and U. D. Dwivedi, "A comparative study of heterogeneous ensemble methods for the identification of geological lithofacies," *Journal of Petroleum Exploration and Production Technology*, vol. 10, no. 5, pp. 1849–1868, Jun. 2020, doi: 10.1007/s13202-020-00839-y.
[16] R. S. Akinbo and O. A. Daramola, "Ensemble machine learning algorithms for prediction and classification of medical images," in *Machine Learning - Algorithms, Models and Applications*, 2021.
[17] M. Ballings, D. Van den Poel, N. Hespeels, and R. Gryp, "Evaluating multiple classifiers for stock price direction prediction," *Expert Systems with Applications*, vol. 42, no. 20, pp. 7046–7056, Nov. 2015, doi: 10.1016/j.eswa.2015.05.013.
[18] Z. Wang, R. Srinivasan, and Y. Wang, "Homogeneous ensemble model for building energy prediction: a casestudy using ensemble regression tree," *ACEEE Summer Study on Energy Efficiency in Buildings*, pp. 1–12, 2016.
[19] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A comprehensive evaluation of ensemble learning for stock-market prediction," *Journal of Big Data*, vol. 7, no. 1, Dec. 2020, doi: 10.1186/s40537-020-00299-5.
[20] E. Alshdaifat, M. Al-hassan, and A. Aloqaily, "Effective heterogeneous ensemble classification: An alternative approach for selecting base classifiers," *ICT Express*, vol. 7, no. 3, pp. 342–349, Sep. 2021, doi: 10.1016/j.icte.2020.11.005.
[21] J. Han, M. Kamber, and J. Pei, "Mining frequent patterns, associations, and correlations," in *Data Mining*, Elsevier, 2012, pp. 243–278.
[22] J. N. van Rijn, G. Holmes, B. Pfahringer, and J. Vanschoren, "The online performance estimation framework: heterogeneous ensemble learning for data streams," *Machine Learning*, vol. 107, no. 1, pp. 149–176, Jan. 2018, doi: 10.1007/s10994-017-5686-9.
[23] N. Ostvar and A. M. E. Moghadam, "HDEC: a heterogeneous dynamic ensemble classifier for binary datasets," *Computational Intelligence and Neuroscience*, vol. 2020, pp. 1–11, Dec. 2020, doi: 10.1155/2020/8826914.
[24] S. Gu and Y. Jin, "Multi-train: A semi-supervised heterogeneous ensemble classifier," *Neurocomputing*, vol. 249, pp. 202–211, Aug. 2017, doi: 10.1016/j.neucom.2017.03.063.
[25] S. Lertampaiporn, T. Chinae, N. Chakarida, K. Boonserm, and R. Marasri, "Heterogeneous ensemble approach with discriminative features and modified-SMOTEbagging for pre-miRNA classification," *Nucleic Acids Research*, vol. 41, no. 1, pp. e21--e21, Jan. 2013, doi: 10.1093/nar/gks878.
[26] M. Sabzevari, G. Martínez-Muñoz, and A. Suárez, "Building heterogeneous ensembles by pooling homogeneous ensembles," *International Journal of Machine Learning and Cybernetics*, vol. 13, no. 2, pp. 551–558, Feb. 2022, doi: 10.1007/s13042-021-01442-1.
[27] N. M. Baba, M. Makhtar, S. A. Fadzli, and M. K. Awang, "Current issues in ensemble methods and its applications," in *Journal of Theoretical and Applied Information Technology*, 2015, vol. 81, no. 2, pp. 266–276.
[28] S. Kausar *et al.*, "Mining smart learning analytics data using ensemble classifiers," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 12, Jun. 2020, doi: 10.3991/ijet.v15i12.13455.
[29] A. V. Luong *et al.*, "A homogeneous-heterogeneous ensemble of classifiers," in *Communications in Computer and Information Science*, Springer International Publishing, 2020, pp. 251–259.
[30] W. Wang, "Heterogeneous Bayesian ensembles for classifying spam emails," in *The 2010 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2010, pp. 1–8, doi: 10.1109/IJCNN.2010.5596545.
[31] S. Shashank and M. P. Mahapatra, "Boosting rock facies prediction: weighted ensemble of machine learning classifiers," *Day 2 Tue, November 13, 2018*, Nov. 2018, doi: 10.2118/192930-MS.
[32] L. Gutierrez-Espinoza, F. Abri, A. S. Namin, K. S. Jones, and D. R. W. Sears, "Fake reviews detection through ensemble Learning," *arxiv.org/abs/2006.07912*, Jun. 2020.
[33] E. A. Gamie, S. A. El-Seoud, and M. A. Salama, "Comparative analysis for boosting classifiers in the context of higher education," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 10, Jun. 2020, doi: 10.3991/ijet.v15i10.13663.
[34] P. Cortez and A. M. G. Silva, "Using data mining to predict secondary school student performance," in *Proceedings of 5th Future Business Technology Conference (FUBUTEC 2008)*, pp. 5-12, 2008.
[35] E. Evangelista, "A hybrid machine learning framework for predicting students' performance in virtual learning environment," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 24, pp. 255–272, Dec. 2021, doi: 10.3991/ijet.v16i24.26151.
[36] M. Salihoun, "State of art of data mining and learning analytics tools in higher education," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 21, Nov. 2020, doi: 10.3991/ijet.v15i21.16435.

**BIOGRAPHIES OF AUTHORS**

**Edmund De Leon Evangelista** is an Assistant Professor at the College of Technological Innovation at Zayed University, Abu Dhabi Campus, United Arab Emirates. In 2019, he received his Ph.D. in Information Technology from Saint Paul University Philippines. His research interests include machine learning, data mining, and software engineering. He has over nineteen years of experience working both in the IT Industry and in IT Academia. He has also held positions as Team Leader, Software Engineer, and Web/Moodle Developer within the IT industry (Oman, Kuwait, and the Philippines). He can be contacted at email: undoevangelista@gmail.com, edmund.evangelista@zu.ac.ae.

**Benedict Descargar Sy** is an Assistant Instructor and Part-time Graduate School professor at the University of Saint Louis. In 2019, he obtained his Ph.D. in Information Technology from Saint Paul University Philippines. His research study interests include machine learning and data-mining. In addition, he has over 11 years of teaching experience and two years of working in the IT industry. He can be contacted at email: benedictsy@usl.edu.ph.