

12-21-2021

A Hybrid Machine Learning Framework for Predicting Students' Performance in Virtual Learning Environment

Edmund Evangelista
Zayed University

Follow this and additional works at: <https://zuscholars.zu.ac.ae/works>



Part of the [Computer Sciences Commons](#), and the [Education Commons](#)

Recommended Citation

Evangelista, Edmund, "A Hybrid Machine Learning Framework for Predicting Students' Performance in Virtual Learning Environment" (2021). *All Works*. 4749.
<https://zuscholars.zu.ac.ae/works/4749>

This Article is brought to you for free and open access by ZU Scholars. It has been accepted for inclusion in All Works by an authorized administrator of ZU Scholars. For more information, please contact scholars@zu.ac.ae.

A Hybrid Machine Learning Framework for Predicting Students' Performance in Virtual Learning Environment

<https://doi.org/10.3991/ijet.v16i24.26151>

Edmund D. Evangelista

Zayed University, Abu Dhabi Campus, United Arab Emirates
edmund.evangelista@zu.ac.ae

Abstract—Virtual Learning Environments (VLE), such as Moodle and Blackboard, store vast data to help identify students' performance and engagement. As a result, researchers have been focusing their efforts on assisting educational institutions in providing machine learning models to predict at-risk students and to improve their performance. However, it requires an efficient approach to construct a model that can ultimately provide accurate predictions. Consequently, this study proposed a hybrid machine learning framework to predict students' performance using eight classification algorithms and three ensemble methods (Bagging, Boosting, Voting) to determine the best-performing predictive model. In addition, this study used filter-based and wrapper-based feature selection techniques to select the best features of the dataset related to students' performance. The obtained results reveal that the ensemble methods recorded higher predictive accuracy when compared to single classifiers. Furthermore, the accuracy of the models improved due to the feature selection techniques utilized in this study.

Keywords—machine learning, Weka, predictive model, ensemble, student performance prediction, classification algorithm, virtual learning

1 Introduction

The COVID-19 pandemic has affected many educational institutions in terms of their implemented teaching and learning pedagogies. Several schools, colleges, and universities have discontinued face-to-face teachings [1]. Instead, most universities shifted to virtual and digital strategies [2] using their chosen Virtual Learning Environment (VLE) platforms such as Blackboard, Canvas, and Moodle. VLE is a platform that allows students and teachers to interact, present, and share resources/activities to complete an entire online course or serves as a supporting feature in traditional teaching courses [3] [4].

Many universities have fully digitalized their operations [5], enabling faculty and students to carry out teaching and learning experiences amid the pandemic. Due to the significant demand for VLEs brought about by the lockdown of schools, student performance prediction and analysis is an essential task [6] that serves as supportive tools

to educators and metacognitive triggers to learners [7]. Prediction using machine learning techniques has enormous potential to assist faculty in identifying students' poor performance by enabling an early warning system [8]. Machine learning (ML) aims at creating algorithms that can learn and generate statistical models for data analysis and prediction. The ML algorithms should learn on their own, based on the data provided, and make accurate predictions without being explicitly programmed for a given task [9]. As a result, faculty can devote more attention to such underperforming students to prepare them for summative assessments on time. This effort usually leads to early detection of at-risk students, enhanced academic achievement, identification of weak learners, and trimming down of failure rates [10].

Various researchers conducted studies on this domain, but it mainly focused on using a single model prediction [11] or most commonly known as classification algorithms such as Decision Tree C4.5, Random Forest, Support Vector Machine (SVM), and Naive Bayes [12]. This study aimed to provide a machine learning framework to predict students' performance using a hybrid of classification and ensemble methods. The ensemble method uses multiple classification algorithms strategically generated and integrated to get a better prediction performance than the performance obtained from a single algorithm [13-14]. It combines the best-selected techniques as the final prediction model [15]. Furthermore, it combines different machine learning techniques into a single predictive model to reduce variability (bagging), bias (boosting), or improve results (stacking) [61]. The experimental results of various studies show that ensemble methods gained a higher accuracy performance when compared to a single classification model [15-17]. Moreover, the study utilized wrapper-based and filter-based feature selection techniques to identify the best features of the dataset used to build the final predictive model. The main objective of the ML framework is to compare the prediction models made implemented using both classification and ensemble methods, then choose the best one with the highest predictive accuracy.

As a result of these considerations, the current study posed the following research questions:

1. Do feature selection techniques improve the accuracy of the predictive model?
2. What is the best machine learning classification algorithm to predict students' performance in VLEs?
3. Does the use of ensemble methods help the predictive models to achieve better performance?

The remainder of this paper is divided into four (4) sections—Section 2 presents background and related works. Section 3 presents the methodology, data collection process, description, dataset generation, and model evaluation. Finally, section 4 discusses the obtained results, and Section 5 covers the conclusions and planned future works.

2 Background and related works

Machine learning (ML) for education is a new field in which a predictive model is created based on training data to predict students' performance [18]. The main goal is

to identify students who will have difficulty learning and take preventive measures. It has become a pressing need among educational institutions for a variety of purposes, such as detecting at-risk students [19], ensuring student retention [20], online learning behavior analysis [21], and many others. In addition, the wide use of VLEs generates large amount of data about teaching and learning interactions which can be helpful to discover hidden knowledge related to students' performance [10].

ML techniques can assist in this direction by providing a framework that can analyze the data of each learner gathered from the interaction logs of VLEs [22]. Every student's online interaction, such as a click, a page visited, or a video viewed, is recorded in the log history [23]. Data miners gather data from the log history and work with analysts toward making predictions for pedagogical intervention [24], such as feeding the result into an integrated system with the task of showing the predicted final grade of a student [25].

Many researchers conducted studies to predict students' performance in VLEs based on various ML frameworks. For example, Soni et al. [26] analyzed the pupil's performance from their last performances using classification algorithms and prepared a dataset of about 2000 students with 50 attributes. Results reveal that students' performance does not rely only on their marks but also on their extracurricular and personal habits. Ünal [27] used feature subset selection and classification operations to predict student performances using two publicly available datasets. The study used classification techniques such as decision trees, random forest, and Naive Bayes to compare their accuracy rates. Similarly, Adnan et al. [28] proposed a predictive model that analyzes the problems faced by at-risk students. They trained and tested their model using various machine learning (ML) and deep learning (DL) algorithms to characterize the learning behavior.

Moreover, Cui et al. [62] proposed an emotion recognition model that monitors each student's real-time emotional state during English learning. For example, when frustration or boredom is detected, machine learning will switch to contents that interest the student or are easier to learn, keeping the student engaged in learning. Finally, Saqib et al. [8] applied in their model a combination of logistic and regression algorithms on historical data to predict the final grade of students taking the same course in the next term. Experimental results show linear discrimination analysis as the most effective approach to correctly predicting students' performance outcomes in final exams.

Unlike the other studies that focused mainly on using a single classification algorithm, this study aimed to present a framework that will utilize a hybrid of classification and ensemble methods. Ensemble learning is an ML process to improve prediction performance by strategically combining the predictions from multiple learning algorithms [29-31]. Therefore, this study aimed to utilize single classifiers and ensemble methods to exhaust all options in determining the ML model that provides the highest accuracy in predicting students' performance based on the data extracted from the VLEs.

3 Materials and methods

The following subsections provide a brief overview of the research materials and methodology used in this study.

3.1 Dataset and target class

This study made use of an open-access dataset from the repository of Optimized Computing and Communications (OC2) used previously as lab's work on student prediction in eLearning environments using machine learning methods. The dataset contributors [32-36] were cited accordingly in this study to reuse their dataset. If interested, the dataset may also be accessed here [37-38]. Four (4) datasets are available on this public repository, but this study used the Student Performance Prediction - Multiclass Scenario dataset.

The dataset comes from a second-year undergraduate Science course at the University of Western Ontario containing grades of the 486 students in the different assignments, quizzes, and exams [32]. Table 1 shows the attribute description and type of data, along with possible values.

Table 1. Description of dataset attributes

Feature	Description	Type	Values
Id	Student Id	Nominal	Std000-485
Quiz01	Quiz1 Score	Numeric	0-10
Assign01	Assignment1 Score	Numeric	0-8
Midterm	Midterm Score	Numeric	0-20
Assign02	Assignment2 Score	Numeric	0-12
Assign03	Assignment3 Score	Numeric	0-25
Final Exam	Final Exam Score	Numeric	0-35
Final Grade	Final Grade Score	Numeric	0-100
Total	Final Grade	Nominal	G, F, W

In this dataset, all features have been scaled to 100 to improve the accuracy of the classifiers. The final grade is 100% instead of 110% due to an additional 10% applied into Assignment03's score as a curve or extra credit to assist students in their course grades. The total final grade of the students serves as the target class in this study. It is a multiclass variable that groups students into three categories, namely:

1. Good (G) – course grade of the student is between 70-100%
2. Fair (F) - course grade of the student is between 51-69%
3. Weak (W) - course grade of the student is between <50%

Depending on the type of student intervention configured to be detected by a machine learning framework, the students under the Weak category are the most vulnerable at risk of failing in a course. When building a predictive model, the range of final grades in each type is set based on the educational institution's policy.

3.2 Proposed framework and machine learning tool

The primary aim of this study was to provide a machine learning framework to predict students' performance using a hybrid of single classifiers and ensemble methods. Unlike traditional ML approaches, which train data using a single learning model, ensemble methods attempt to use a collection of models, then combine them to vote on their results [10]. Thus, the model gaining the most efficient accuracy will predict the future dataset gathered from the VLE. The framework of this study was inspired based upon Cross Industry Standard Process for Data Mining (CRISP-DM) [39]. It is the most used methodology for developing data mining and knowledge discovery projects [40]. However, this framework combines the strength of the CRISP-DM model in its data mining process while it introduces a hybrid approach of using classification algorithms and ensemble methods to optimize students' performance detection. Figure 1 illustrates the proposed framework consisting of four significant steps.

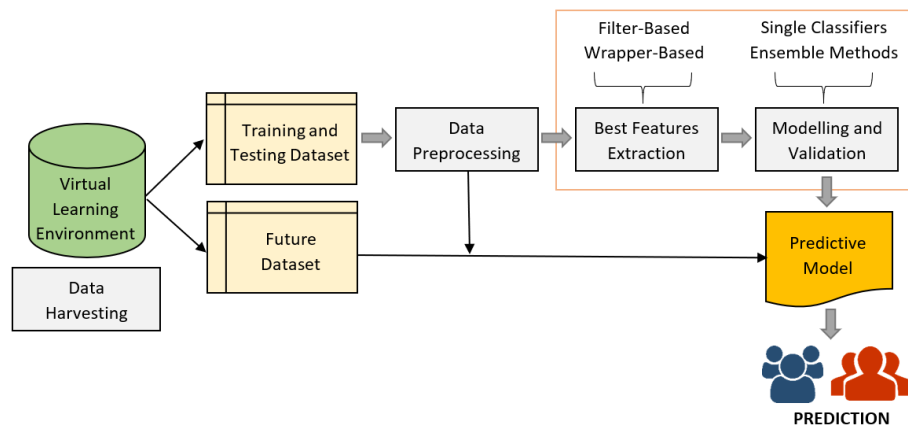


Fig. 1. The proposed framework

This study used the Weka machine learning tool to build a predictive model. Weka is a Java-based open-source machine learning software suite that includes visualization tools and algorithms for data analysis and predictive modeling [41-42].

Data harvesting. Data harvesting or most commonly known as "data collection," is a process that extracts and analyzes data collected from online sources [43], referred to as VLE in this study. Usually, it is achieved by running a SQL script or using any built-in features the back-end database provides. Then, data miners harvest activity and engagement logs of students from the VLE and consequently producing training and testing datasets for the predictive model. The model will learn to perform a task using a

training dataset, and the testing dataset will ensure that the model works correctly [44]. Moreover, the training dataset is used for model fitting or estimating the model's parameters. The test dataset is then used for final model evaluation, assessing the performance of the estimated model [45].

Data preprocessing. Data preprocessing involves cleaning noisy and missing values and handling outliers from the dataset. As a result, the predictive accuracy of the classifiers improves when appropriately managed. Weka supports numerous built-in preprocessing techniques such as converting numeric data to nominal data (discretize), synthetic re-balancing of the dataset (SMOTE), normalize, standardize, remove duplicates, and many more. Regardless of the preprocessing techniques implemented, the objective is to help the ML classifiers detect patterns and behavior accurately.

The dataset used in this study was preprocessed by scaling all the features out of 100. It includes replacing missing values with 0 and removing the Student Id attribute because it is insignificant to the predictive model. It also involves removing the Final Exam attribute because the study aims to identify at-risk students before taking their final exam. Without the final exam weighing 35%, the 65% stage of the student's course grade was calculated as the sum of the weights of Quiz01, Assign01, Midterm, Assign02, and Assign03, respectively. Then the result was divided by 65 and eventually multiplied by 100. Consequently, the new value of the target class takes on the 65% stage of the course grade.

Figure 2 shows the graphical view of the dataset features; the predicted class contains 464 good (G), eight fair (F), and 14 weak (W) students.

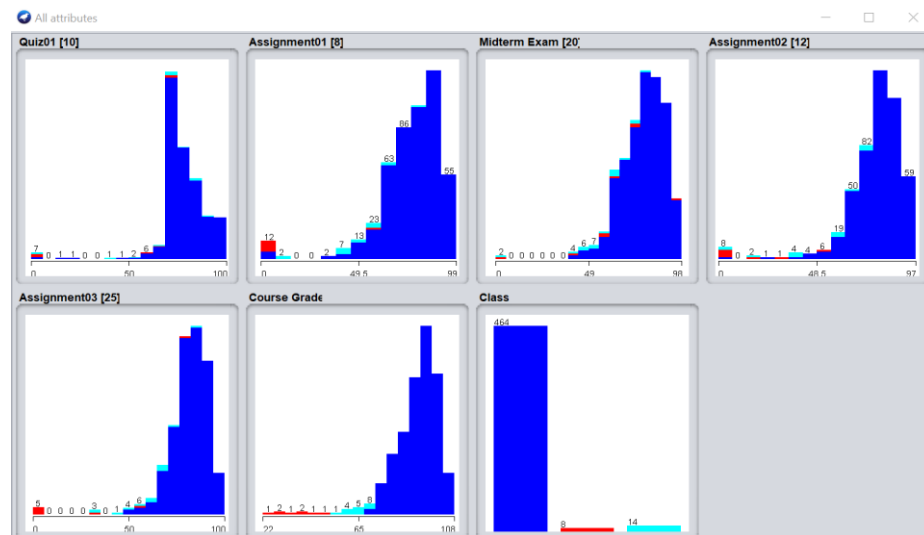


Fig. 2. Graphical view of dataset features

Best features extraction. Best features extraction, commonly known as feature or attribute selection, is a valuable method for dealing with high-dimensional data analysis by removing irrelevant and redundant data [46]. This method can shorten computation

time, improve learning accuracy, and better understand the learning model or data. This step will help remove some features in the dataset that do not significantly contribute to the model's predictive accuracy. Knowledge discovery during training becomes more complicated when information is irrelevant, redundant, noisy, or unreliable [47]. Therefore, its elimination frequently improves the performance of machine learning algorithms.

In data mining, adding too many features may result in overfitting; the opposite (very few features) can also result in underfitting. Therefore, there is a need to select the best attributes for the model to reduce the possibility of having poor predictive performance. There are two known feature extraction methods in ML; they are wrapper-based and filter-based methods. The wrapper-based methods seek to find the fewest discriminative features possible to achieve high classification accuracy. On the other hand, filter-based methods compute the 'best' subset of attributes based on some criteria [48].

Moreover, all the features in filter methods are scored and ranked based on specific statistical criteria. It chooses the attributes with the highest-ranking values and eliminates the low-ranking ones [49]. Compared to the filter methods, the computational costs of wrapper methods are higher. As a result, they are unsuitable for high-dimensional datasets; however, they are more effective at identifying the subset of features. Furthermore, the high accuracy of these methods for selecting a subset of features is noticeable [50].

This study used both methods to ensure that these feature selection techniques have joint agreements on the best features used to build the predictive model. Weka tool supports both filter-based and wrapper-based methods in constructing predictive models. The framework used the CFS Subset Eval algorithm for filter-based method, using Best First and Greedy Stepwise Search as its search method. This algorithm assesses the value of a subset of attributes by considering each feature's individual predictive ability as well as the degree of redundancy between them [51]. On the other hand, the wrapper-based method will use the Classifier Subset Eval algorithm with various classifiers as its wrapper. This algorithm evaluates attribute subsets using training data and estimates the merit of a group of attributes.

Modelling and validation. When building a machine learning framework, the best features of the dataset should be modeled using various classification algorithms, evaluate its accuracy based on some performance metrics, and select the best algorithm to build the predictive model. Unfortunately, a common mistake of data miners is pre-selecting a specific algorithm and failing to compare it with the rest of the classifiers.

This study used the most common classification algorithms mentioned in the literature to train and test the model, such as Naive Bayes (NB), Decision Table (DT), Random Forest (RF), K-Nearest Neighbor (KNN), One Rule (OneR), Decision Tree (J48), Support Vector Machine (SVM), Neural Networks (MP), and Rule Induction (JRip). Moreover, the ensemble methods used in this study were bagging, boosting, and voting. Figure 3 illustrates the expanded view of the framework in which the modeling and validation step will use the best features extracted from the dataset. The ML framework introduced in this study is a hybrid approach for two reasons:

1. It uses filter-based and wrapper-based methods to extract the best features.
2. It determines the most suitable predictive model that best responds to the given dataset by training single classifiers and ensemble methods and then selects the best model.

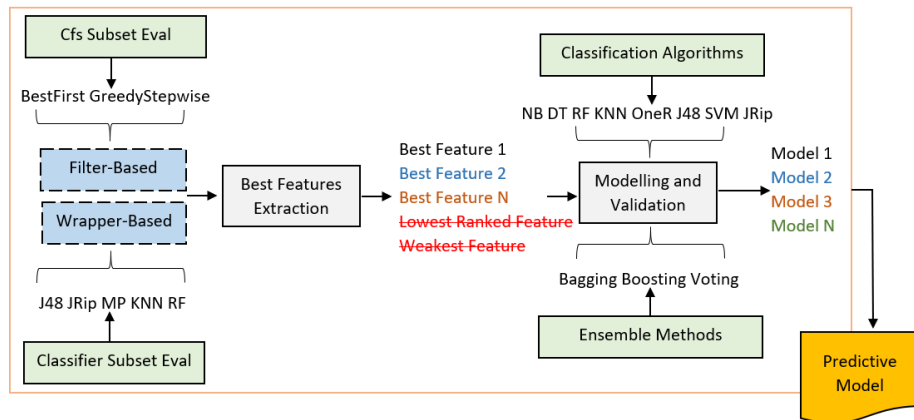


Fig. 3. Expanded view of the ML Framework

Further, the ensemble methods are based on the idea that a group of experts can make more accurate decisions than a single expert. Therefore, it combines classifiers to produce a single composite model with higher accuracy [31]. There are at least four ensemble methods to choose from: bagging, boosting, voting, and stacking. Voting constructs two or more sub-models, then each sub-model makes predictions and combines them somehow to get the mean or the mode of the predictions [52]. As shown in Figure 4, Weka supports ensemble methods by creating a single model of combined classification algorithms of your choice and predicts output based on their combined majority of voting for the target class.

Similarly, boosting, also known as a "meta-algorithm," is a chronological or sequential process in which each successive model attempts to remedy or correct the errors of the previous model [53]. On the contrary, bagging uses a bootstrap method to create multiple training sets. Multiple training sets are made using the bootstrap method by selecting random and repeatable samples from the original dataset [54]. Stacking, on the other hand, applies different learning algorithms to a single dataset. The predictions of the various classifiers are then combined and used by a meta-level classifier to generate a final hypothesis [55].

All the classification algorithms and ensemble methods used in this study would be trained and tested using 10-folds cross-validation. Cross-validation is a statistical method for evaluating and comparing learning algorithms that divide data into two segments: learning or training a model and validating the model [56]. Its most common form is ten-fold cross-validation, which splits the dataset into nine (9) sub folds for training, and one (1) fold for testing sets, then rotates the folds [57].

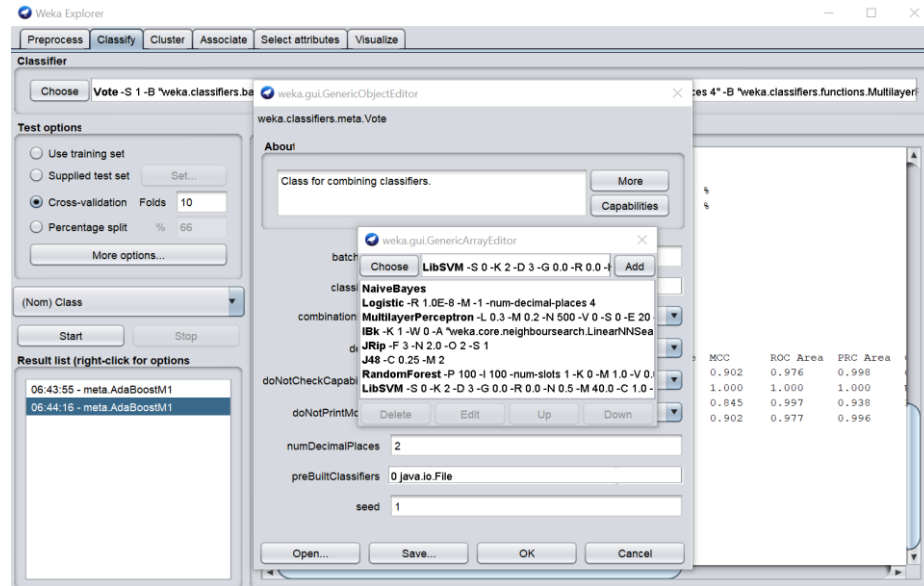


Fig. 4. Adding of classifiers into Voting Ensemble Method using Weka

The following performance evaluation metrics were used to evaluate the obtained results of the models: accuracy, F-measure, precision, and recall. Their calculations were based on a confusion matrix of binary classifiers, as illustrated in Figure 5 and described mathematically in Equations (1), (2), (3), and (4), respectively.

		DETECTED	
		Positive	Negative
ACTUAL	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Fig. 5. The Confusion Matrix

$$Recall = \frac{TP}{TP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (3)$$

$$F - Measure = \frac{2 \times (Recall \times Precision)}{Recall+Precision} \quad (4)$$

Precision is the percentage of correct predictions among positively predicted cases, whereas recall is accurate predictions among actual positive cases. Thus, accuracy, precision, recall, and F-measure values are within (0,1), and higher values indicate better predictions [58]. Furthermore, accuracy is the likelihood that a randomly selected instance (positive or negative, relevant or irrelevant) will be correct. In contrast, F-measure is the weighted harmonic mean of recall and precision [59]. In this study, the accuracy of the model and its corresponding F-measure value would be monitored to determine the best performing model. A model with the highest accuracy and an F-measure value close to 1 is our target model.

4 Results and discussion

In this study, the best features of the dataset were extracted using filter-based and wrapper-based methods. In addition, the selected attributes were trained and evaluated using a hybrid of single classifiers and ensemble methods. As a result, the model with the highest recorded accuracy and F-measure would predict students' performance in a future dataset. All experiments were carried out using the Weka machine learning tool.

4.1 Best features extracted

Table 2 shows the selected subset of Classifier Subset Eval wrapper method using single classifiers such as J48, JRip, MP, KNN, and RF. The best-first search was used as the search method, while accuracy served as the performance evaluation measure in selecting attributes. The table reveals that 3 out of 5 algorithms (J48, JRip, and MP) agree that all features are relevant and correlated. Only KNN and RF algorithms dropped Assign02 in the selected subset.

Table 2. Feature selection using Classifier Subset Eval wrapper method

J48	JRip	MP	KNN	RF
Assign03	Assign03	Assign03	Assign03	Assign03
Quiz01	Quiz01	Quiz01	Quiz01	Quiz01
Assign01	Assign01	Midterm	Assign01	Assign01
Assign02	Assign02	Assign02	Midterm	Midterm
Midterm	Midterm	Assign01		

Similarly, the ranked features using the CFS Subset Eval filter method having best first and greedy stepwise search are shown in Table 3. It can be observed in the table that the ranking of features is the same for both search methods having Assign03 and Midterm as the highest-ranked and lowest-ranked attributes, respectively.

Table 3. Feature selection using CFS Subset Eval filter method

Rank	Best First Search	Greedy Stepwise Search
5	Assign03	Assign03
4	Quiz01	Quiz01
3	Assign01	Assign01
2	Assign02	Assign02
1	Midterm	Midterm

The objective of using both methods is to identify the weakest attribute/s in wrapper methods (Assign02) and to determine the lowest-ranked attribute/s in the filter method (Midterm). With this, the best features would include Assign03, Quiz01, and Assign01, respectively. These would be used for experimental purposes in training the models.

4.2 Performance of the trained models

The comparative analysis of the performance of the trained models using single classifiers and ensemble methods is demonstrated in Table 4. Based on the experiments shown in Table 4, it could be observed that using ensemble methods improved the trained models' performance compared to single classifiers. The predictive accuracy of ensemble methods was higher than any of the single classifiers regardless of whether all or best dataset features were used. Boosting gained the highest accuracy of 98.56% when all features were trained, while Bagging recorded 98.35% when only the best features were used. Indeed, the ensemble methods (Bagging, Boosting, Voting) consistently performed better than any single classifiers in training the models for both experiments (all features and best features). The table further reveals that the trained models for single classifiers and ensemble methods gained high F-measure values, almost close to 1. It only means that low false positives and false negatives had been attained; hence, the trained models correctly identified the predicted class.

Table 4. Performance of the Predictive Models

Approach	Algorithm	All Features		Best Features	
		Accuracy	F-Measure	Accuracy	F-Measure
Single Classifiers	NB	95.88	0.97	95.47	0.96
	DT	95.88	0.95	96.50	0.96
	RF	97.33	0.97	97.53	0.97
	KNN	97.74	0.98	97.33	0.97
	OneR	95.68	0.94	95.27	0.94
	J48	95.88	0.95	96.71	0.96
	SVM	95.47	0.98	95.47	0.96
	JRip	96.30	0.96	97.12	0.97
Ensemble Methods	Bagging	98.15	0.98	98.35	0.98
	Boosting	98.56	0.99	98.15	0.98
	Voting	97.94	0.98	98.15	0.98

As seen in Figure 6, eight (8) single classifiers and three (3) ensemble methods were used in this study. Due to this hybrid approach of training the models, the machine learning framework provided by this study gave a complete overview of which algorithm performed better between single classifiers or ensemble methods to predict students' performance in a future dataset. The algorithm with the highest predictive accuracy would be chosen to build the final predictive model.

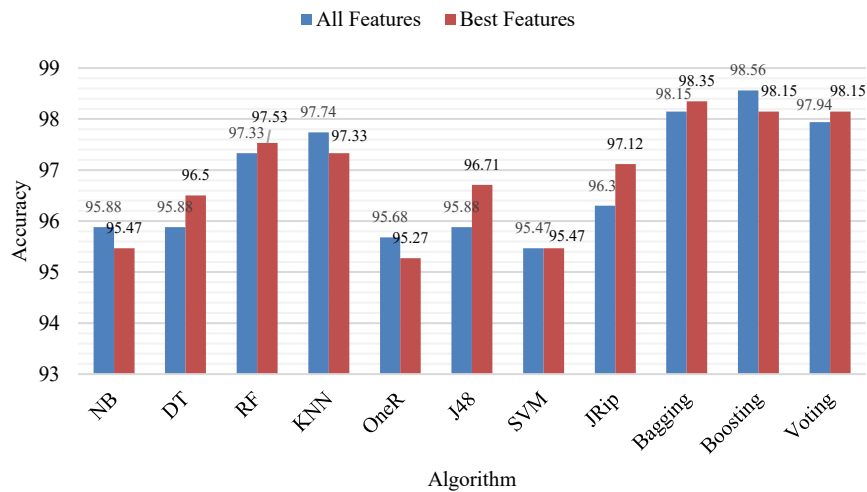


Fig. 6. Comparison of model's accuracy in term of the number of features

To illustrate this concept, among the single classifiers, KNN gained the highest accuracy of 97.74% and 97.33% for all features and best features, respectively. Similarly, boosting algorithm topped the other ensemble methods earning a recorded accuracy of 98.56% and 98.15% on the same experiments. Since boosting algorithm was higher than KNN on this experiment; therefore, it would be chosen to build the model for predicting future datasets if the plan was to use all features. Furthermore, it is valid provided all features are proven to be correlated to the predicted class.

The performance of the predictive models improved better for DT, RF, J48, JRip, Bagging, and Voting algorithms when it trained using the best features of the dataset. The best attributes of the dataset were selected using the combined approach of Classifier Subset Eval and CFS Subset Eval feature selection techniques. Feature selection improves classification performance because it helps obtain optimal accuracy; however, this is dependent on the feature selection method used [60].

Figure 6 further reveals that algorithms like NB, KNN, OneR, and Boosting did not improve their accuracy using the best features. It could be due to the feature selection methods used, which may not fit into these algorithms. Based on the experimental results, when the best features of the dataset were used, the bagging algorithm gained the highest accuracy of 98.35%. Therefore, it would be chosen to build the final predictive model if only the plan was to use the best features of the dataset. In machine learning,

the best attributes selected by feature selection techniques are more preferred than using all the dataset features because it reduces training time, decreases over-fitting, and improves the accuracy if the right subset was chosen.

4.3 Research questions

This section addresses the following research questions:

Do feature selection techniques improve the accuracy of the predictive model?

Feature selection techniques improve the accuracy of the predictive models by selecting the best subset correlated to the target class. As illustrated in Figure 6, most of the classifiers (DT, RF, J48, JRip, Bagging, Voting) used in this study improved the model's accuracy. However, the right method should be selected appropriately to achieve this. Weka supports various feature selection techniques involving wrapper-based and filter-based methods.

What is the best machine learning classification algorithm to predict students' performance in VLEs? A common mistake of most data miners is to pre-select a specific algorithm to solve an ML problem due to their existing algorithm assumptions. For example, various researchers use Decision Trees or Neural Networks algorithms since they predict better without testing other available algorithms. Before settling on a final model, it is critical to compare the predictive accuracy of various algorithms first. As shown in Table 4, various algorithms using the available dataset harvested from VLEs were trained, their accuracy and F-measure was recorded, then the best classification algorithm was selected. In this dataset, boosting algorithm outperformed the other known classifiers used in the study; such as K-Nearest Neighbor and Random Forest.

Does the use of ensemble methods help the predictive models to achieve better performance? Ensemble methods are a general meta-machine learning approach that seeks to improve predictive performance by combining predictions from multiple models [52]. Figure 6 shows that out of 11 classification algorithms and meta algorithms used in this study, the ensemble methods such as bagging, boosting, and voting achieved better predictive accuracy than all the other classification algorithms. For example, boosting algorithm gained the highest accuracy of 98.56% when all features were used, while the bagging algorithm recorded the highest accuracy of 98.35% when the best features were used.

5 Conclusion and future works

The early prediction of students' performance is an essential tool for educational institutions to provide necessary intervention to at-risk students. Machine learning (ML) is one of the methods used for student profile modeling to create knowledge from data automatically [63]. ML techniques in predicting student performance have been proven to help identify poor performers and allow tutors to take early corrective measures [64]. To this end, an ML framework is proposed for predicting students' performance in a Virtual Learning Environment (VLE) using the Weka machine learning tool. Unlike

similar studies conducted in the past, this study is a hybrid approach of training models based on comparing single classifiers and ensemble methods. Moreover, the selection of the best features of the dataset was determined using filter-based and wrapper-based methods. The study made use of an open-access dataset from the repository of Optimized Computing and Communications (OC2) containing grades of the 486 students in the different assignments, quizzes, and exams. In addition, experiments were conducted to predict a multi-class case (Good, Fair, Weak) dataset and identified students' performance before taking their final exams.

This study used classification algorithms such as Naive Bayes, Decision Table, Random Forest, K-Nearest Neighbor, One Rule, J48, Support Vector Machine, Multi-layer Perceptron, and JRip. Moreover, the ensemble methods used in this study were bagging, boosting, and voting. Experimental results revealed an increased predictive accuracy of the trained model for all ensemble methods used compared to the single classification algorithms. Furthermore, the performance of the trained models improved among the majority of algorithms when the best features of the dataset were used.

Future works include using the framework to engage in other areas of predicting performance, such as students' engagement and determining students' at-risk of dropping a course. In addition, the author plans to test the framework in a much larger dataset to optimize its performance and to perform any needed tweaking on its processes.

6 Acknowledgement

The author wishes to thank Zayed University (Abu Dhabi Campus), the College of Technological Innovation, and the Department of Information Systems and Technology Management for their exceptional support and start-up grant (R21042) provided in completing this study.

7 References

- [1] Pokhrel S., Chhetri R. A Literature Review on Impact of COVID-19 Pandemic on Teaching and Learning. *Higher Education for the Future*. 2021,8(1):133-141. <https://doi.org/10.1177/2347631120983481>
- [2] Said, E. How Did the COVID-19 Pandemic Affect Higher Education Learning Experience? An Empirical Investigation of Learners' Academic Performance at a University in a Developing Country. *Advances in Human-Computer Interaction*, vol. 2021, Article ID 6649524, 10 pages, 2021. <https://doi.org/10.1155/2021/6649524>
- [3] Bahman, R. VLEs in a Post-COVID World: Kuwait's Universities. *Challenges and Opportunities for the Global Implementation of E-Learning Frameworks*, 2021, 19 pages.
- [4] Mosquera, L. Impact of Implementing a Virtual Learning Environment (VLE) in the EFL Classroom. October 2017, *Íkala Revista de Lenguaje y Cultura* 22(3):479-498. <https://doi.org/10.17533/udea.ikala.v22n03a07>
- [5] Dhawan, S. Online Learning: A Panacea in the Time of COVID-19 Crisis. *Journal of Educational Technology Systems*. 2020 Sep, 49(1): 5–22. <https://doi.org/10.1177/0047239520934018>

- [6] Sawant, P., Gupta, S., Sharma, Y., Singh, A. Classification Approach for Evaluating Students Performance in Covid 19 Pandemic. International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249-8958, Volume-10 Issue-4, April 2021. <https://doi.org/10.35940/ijeat.D2374.0410421>
- [7] Dias, S.B., Hadjileontiadou, S.J., Diniz, J. et al. DeepLMS: a deep learning predictive model for supporting online learning in the Covid-19 era. Sci Rep 10, 19888 (2020). <https://doi.org/10.1038/s41598-020-76740-9>
- [8] Gull, H., Saqib, M., Iqbal, S., Saeed, S. Improving Learning Experience of Students by Early Prediction of Student Performance using Machine Learning. IEEE International Conference for Innovation in Technology (INOCN), 2020, pp. 1-4. <https://doi.org/10.1109/INOCN50539.2020.9298266>
- [9] Chibani, S., Coudert, F. Machine learning approaches for the prediction of materials properties. APL Materials, AIP Publishing 2020, 8 (8), pp.080701. <https://doi.org/10.1063/5.0018384>
- [10] Amrieh, E.A., Hamtini, T.M., Aljarah, I. Mining Educational Data to Predict Student's academic Performance using Ensemble Methods. International journal of database theory and application, 2016, 9, 119-136. <https://doi.org/10.14257/ijdt.2016.9.8.13>
- [11] Aman, F., Rauf, A., Ali, R., Iqbal, F., Khattak, A. M. A Predictive Model for Predicting Students Academic Performance. 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA), 2019, pp. 1-4. <https://doi.org/10.1109/IISA.2019.8900760>
- [12] Sadiq, M. H., Ahmed, N. S. Classifying and Predicting Students' Performance using Improved Decision Tree C4.5 in Higher Education Institutes. Journal of Computer Science, 2019, 15(9), 1291-1306. <https://doi.org/10.3844/jcssp.2019.1291.1306>
- [13] Yan, L., Liu, Y. An Ensemble Prediction Model for Potential Student Recommendation Using Machine Learning. Symmetry 2020, 12, 728. <https://doi.org/10.3390/sym12050728>
- [14] Parikh, D., Polikar, R. An ensemble-based incremental learning approach to data fusion. Syst. Man Cybern. 2007, 37, 437–450. <https://doi.org/10.1109/TSMCB.2006.883873>
- [15] Almasri, A., Celebi, E., Alkhawaldeh, R. EMT: Ensemble Meta-Based Tree Model for Predicting Student Performance. Scientific Programming, vol. 2019, Article ID 3610248, 13 pages, 2019. <https://doi.org/10.1155/2019/3610248>
- [16] Stapel, M., Zheng, Z., Pinkwart, N. An Ensemble Method to Predict Student Performance in an Online Math Learning Environment. 2016. Proceedings of the 9th International Conference on Educational Data Mining.
- [17] Seni, G., JohnF.Elder, I. Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions. Ensemble Methods in Data Mining. 2010. <https://doi.org/10.2200/S00240ED1V01Y200912DMK002>
- [18] Ismail L., Materwala H., Hennebelle A. Comparative Analysis of Machine Learning Models for Students' Performance Prediction. In: Antipova T. (eds) Advances in Digital Science. ICADS 2021. Advances in Intelligent Systems and Computing, 2021, vol 1352. Springer, Cham. https://doi.org/10.1007/978-3-030-71782-7_14
- [19] Zohair, L.M. Prediction of Student's performance by modelling small dataset size. Int J Educ Technol High Educ, 2019, 16, 27. <https://doi.org/10.1186/s41239-019-0160-3>
- [20] Yadav, S., Bharadwaj, B., Pal, S. Mining Education Data to Predict Student's Retention: A comparative Study. 2012. IJCSIS) International Journal of Computer Science and Information Security, Vol. 10, No. 2, 2012.
- [21] Yan, N., Au, O. Online learning behavior analysis based on machine learning. Asian Association of Open Universities Journal, 2019, 14, 97-106. <https://doi.org/10.1108/AAOUJ-08-2019-0029>

- [22] Macarini, L., Cechinel, C., Machado, M., Ramos, V., Munoz, R. Predicting Students Success in Blended Learning-Evaluating Different Interactions Inside Learning Management Systems. *Applied Sciences*. 2019, 9, 5523. <https://doi.org/10.3390/app9245523>
- [23] Clow, D. An overview of learning analytics. *Teaching in Higher Education*, 2013, Vol. 18 No. 6, pp. 683-695. <https://doi.org/10.1080/13562517.2013.827653>
- [24] Campbell, J.P., DeBlois, P.B., Oblinger, D.G. Academic analytics. *Education Review*, Vol. 42, October 2017, pp. 40-57.
- [25] Madnaik, S. Predicting Students' Performance by Learning Analytics. *Artificial Intelligence and Robotics Commons, and the Other Computer Sciences Commons*, 2020, ETD Projects 941, etd.6jib-ua9w.
- [26] Soni, A., Kumar, V., Kaur, R., Hemavathi, D. Predicting Student Performance Using Data Mining Techniques. *International Journal of Pure and Applied Mathematics*, Volume 119 No. 12, 2018, 221-227.
- [27] Ünal, F. Data Mining for Student Performance Prediction in Education. *Data Mining - Methods, Applications and Systems*, 2020, Computer Science. <https://doi.org/10.5772/intechopen.91449>
- [28] Adnan, M., Habib, A., Ashraf, J., Mussadiq, S., Raza, A.A., Abid, M., Bashir, M., Khan, S.U. Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models. *IEEE Access*, 2021, 9, 7519-7539. <https://doi.org/10.1109/ACCESS.2021.3049446>
- [29] Abuassba, A.O., Zhang, D., Luo, X., Shaheryar, A., Ali, H. Improving Classification Performance through an Advanced Ensemble Based Heterogeneous Extreme Learning Machines. 2017. *Computational Intelligence and Neuroscience*, 2017. <https://doi.org/10.1155/2017/3405463>
- [30] Löfström, T. On Effectively Creating Ensembles of Classifiers: Studies on Creation Strategies, Diversity and Predicting with Confidence. 2015. Stockholm University, Ph.D. thesis, 2015.
- [31] Pandey, M., Taruna, S. Comparative Study of Ensemble Methods for Students' Performance Modeling. *International Journal of Computer Applications*, 2014, 103, 26-32. <https://doi.org/10.5120/18095-9151>
- [32] Injadat, M., Moubayed, A., Nassif, A. B., Shami, A. Systematic ensemble model selection approach for educational data mining. *Knowledge-based Systems*, vol. 200, p. 105992, Jul. 2020. <https://doi.org/10.1016/j.knosys.2020.105992>
- [33] Injadat, M., Moubayed, A., Nassif, A. B., Shami, A. Multi-split optimized bagging ensemble model selection for multiclass educational data mining. *Applied Intelligence*, 50, pp. 4506–4528, Jul. 2020. <https://doi.org/10.1007/s10489-020-01776-3>
- [34] Moubayed, A., Injadat, M., Nassif, A. B., Lutfiyya, H., Shami, A. E-Learning: Challenges and Research Opportunities Using Machine Learning & Data Analytics. In *IEEE Access*, vol. 6, pp. 39117-39138, 2018. <https://doi.org/10.1109/ACCESS.2018.2851790>
- [35] Moubayed, A., Injadat, M., Shami, A., Lutfiyya, H. Student Engagement Level in an e-Learning Environment: Clustering Using K-means. *American Journal of Distance Education*, 34:2, pp. 137-156, Mar. 2020. <https://doi.org/10.1080/08923647.2020.1696140>
- [36] Moubayed, A., Injadat, M., Shami, A., Lutfiyya, H. Relationship Between Student Engagement and Performance in E-Learning Environment Using Association Rules. 2018 *IEEE World Engineering Education Conference (EDUNINE)*, Buenos Aires, 2018, pp. 1-6. <https://doi.org/10.1109/EDUNINE.2018.8451005>
- [37] Moubayed, A., Injadat, M., Shami, A., Nassif, A., Lutfiyya, H. Student Performance and Engagement Prediction in eLearning datasets. 2020. *IEEE Dataport*.

- [38] Student Performance and Engagement Prediction in Elearning Datasets, Western OC2 Lab. Available online: <https://github.com/Western-OC2-Lab/Student-Performance-and-Engagement-Prediction-eLearning-datasets> (accessed on 28 July 2021).
- [39] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. CRISP-DM 1.0: Step-by-step data mining guide, 2000.
- [40] Marbán, Ó., Mariscal, G., Segovia, J. A Data Mining & Knowledge Discovery Process Model, Data Mining and Knowledge Discovery in Real Life Applications. 2009. Julio Ponce and Adem Karahoca, IntechOpen. <https://doi.org/10.5772/6438>
- [41] Evangelista E. 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). ISSN: 2278-3075, Volume-8, Issue-6S3, April 2019.
- [42] Naik, A., Samant, L. Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime. Procedia Computer Science. Vol. 85. Elsevier B.V., 2016. 662–668. Procedia Computer Science. <https://doi.org/10.1016/j.procs.2016.05.251>
- [43] Mathai, P.P., Balan, R.V., Babu, I. An Efficient Approach for Item Set Mining Using Both Utility and Frequency Based Methods. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 12 (2017) pp. 3470-3473.
- [44] R. Medar, V. S. Rajpurohit and B. Rashmi, "Impact of Training and Testing Data Splits on Accuracy of Time Series Forecasting in Machine Learning," 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA), 2017, pp. 1-6. <https://doi.org/10.1109/ICCUBEA.2017.8463779>
- [45] Raykar V.C., Saha A. Data Split Strategies for Evolving Predictive Models. In: Appice A., Rodrigues P., Santos Costa V., Soares C., Gama J., Jorge A. (eds) Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2015. Lecture Notes in Computer Science, vol 9284. Springer, Cham. https://doi.org/10.1007/978-3-319-23528-8_1
- [46] Cai, J., Luo, J., Wang, S., Yang, S. Feature selection in machine learning: A new perspective. Neurocomputing, Volume 300, 2018, Pages 70-79, ISSN 0925-2312. <https://doi.org/10.1016/j.neucom.2017.11.077>
- [47] Hall, M., & Smith, L.A. Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper. FLAIRS Conference, 1999.
- [48] Leng, J., Valli, C., Armstrong, L. A Wrapper-based Feature Selection for Analysis of Large Data Sets. Proceedings of 2010 3rd International Conference on Computer and Electrical Engineering (ICCEE 2010). (pp. 167-170).
- [49] Kumari, B., Swarnkar, T. Filter versus Wrapper Feature Subset Selection in Large Dimensionality Micro Array: A Review. International Journal of Computer Science and Information Technologies, Vol. 2 (3), 2011, 1048-1053.
- [50] Pirgazi, J., Alimoradi, M., Esmaeili Abharian, T. An Efficient hybrid filter-wrapper metaheuristic-based gene selection method for high dimensional datasets. Sci Rep 9, 18580 (2019). <https://doi.org/10.1038/s41598-019-54987-1>
- [51] Bouckaert, R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., Scuse, D. WEKA Manual for Version 3-7-4. Hamilton, New Zealand: University of Waikato, 2011.
- [52] Brownlee, J. Machine Learning Mastery with Weka: Analyze Data, Develop Models, and Work Through Projects. Machine Learning Mastery, v1.1, 2016.
- [53] Nti, I.K., Adekoya, A.F., Weyori, B. A comprehensive evaluation of ensemble learning for stock-market prediction. Journal of Big Data, 2020, 7, 1-40. <https://doi.org/10.1186/s40537-020-00299-5>

- [54] Taser, P.Y. Application of Bagging and Boosting Approaches Using Decision Tree-Based Algorithms in Diabetes Risk Prediction. *Proceedings* 2021, 74, 6. <https://doi.org/10.3390/proceedings2021074006>
- [55] Džeroski, S., & Zenko, B. Is Combining Classifiers with Stacking Better than Selecting the Best One? *Machine Learning*, 2004, 54, 255-273. <https://doi.org/10.1023/B:MACH.0000015881.36452.6e>
- [56] Refaeilzadeh, P., Tang, L., Liu, H. Cross-Validation. In: LIU L., ÖZSU M.T. (eds) *Encyclopedia of Database Systems*, 2009. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-39940-9_565
- [57] Sokkhey, P., Okazaki, T. Hybrid Machine Learning Algorithms for Predicting Academic Performance. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 11(1), 2020. <https://doi.org/10.14569/IJACSA.2020.0110104>
- [58] Ma, J., Ding, Y., Cheng, J.C., Tan, Y., Gan, V.J., Zhang, J. Analyzing the Leading Causes of Traffic Fatalities Using XGBoost and Grid-Based Analysis: A City Management Perspective. *IEEE Access*, 2019, 7, 148059-148072. <https://doi.org/10.1109/ACCESS.2019.2946401>
- [59] Powers, D. What the F-measure doesn't measure: Features, Flaws, Fallacies and Fixes. *ArXiv*, 2015, abs/1503.06410.
- [60] Chu, C., Hsu, A., Chou, K., Bandettini, P., Lin, C. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *NeuroImage*, 2012, 60, 59-70. <https://doi.org/10.1016/j.neuroimage.2011.11.066>
- [61] Kausar, S., Oyelere, S., Salal, Y., Hussain, S., Cifci, M., Hilcenko, S., Iqbal, M., Wenhao, Z., & Huahu, X. (2020). Mining Smart Learning Analytics Data Using Ensemble Classifiers. *International Journal Of Emerging Technologies In Learning (IJET)*, 15(12), pp. 81-102. <https://doi.org/10.3991/ijet.v15i12.13455>
- [62] Cui, Y., Wang, S., & Zhao, R. (2021). Machine Learning-Based Student Emotion Recognition for Business English Class. *International Journal Of Emerging Technologies In Learning (IJET)*, 16(12), pp. 94-107. <https://doi.org/10.3991/ijet.v16i12.23313>
- [63] Hamim, T., Benabbou, F., & Sael, N. (2021). Survey of Machine Learning Techniques for Student Profile Modeling. *International Journal Of Emerging Technologies In Learning (IJET)*, 16(04), pp. 136-151. <https://doi.org/10.3991/ijet.v16i04.18643>
- [64] Hussain, S., Muhsion, Z., Salal, Y., Theodorou, P., Kurtoglu, F., & Hazarika, G. (2019). Prediction Model on Student Performance based on Internal Assessment using Deep Learning. *International Journal Of Emerging Technologies In Learning (IJET)*, 14(08), pp. 4-22. <https://doi.org/10.3991/ijet.v14i08.10001>

8 Author

Edmund Evangelista is an Assistant Professor at the College of Technological Innovation at Zayed University, Abu Dhabi Campus, United Arab Emirates. 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).

Article submitted 2021-08-10. Resubmitted 2021-10-25. Final acceptance 2021-10-27. Final version published as submitted by the author.