 https://doi.org/10.24191/jeesr.v20i1.013

Predicting Engineering Students' Academic Performance using Ensemble Classifiers- A Preliminary Finding

A'zraa Afhzan Ab Rahim*, Norlida Buniyamin

***Abstract*— Current literature review indicates a void of an accurate predictive tool to assist educators and administrators in analyzing and monitoring student performance in Malaysia. Well-known data mining classifiers such as Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), Naïve Bayes (NB), and K-nearest neighbor (KNN) have been traditionally used but often suffer from the high variance and overfitting issues indicated by good performance on training data but relatively poor on unseen data. To address these problems, alternative ensemble classifiers such as Extreme Gradient Boosting (XGB), Random Forest (RF), and Heterogeneous Ensemble Method (HEM) are evaluated/proposed. This paper aims to compare the performance of single versus ensemble classifiers. In addition, another overarching research objective is to predict students' CGPA during their final semester grades by augmenting the more widely used cognitive with non-cognitive features to obtain a holistic solution. Not only will the accuracy among classifiers be compared, but another priority measure is their recall value to ensure each sample is classified correctly. It is found that ensemble classifiers outperform their single classifiers in terms of both accuracy and recall. Preliminary results indicate that augmenting cognitive features with non-cognitive features results in better accuracy in classifiers and can classify samples according to their respective classes with less variability.**

***Index Terms*—Academic Performance, Ensemble Model, Prediction**

I. Introduction

A persistent issue in tertiary education is students' poor academic performance which delays graduation or, even worse, leads to dropout. Ideally, all engineering undergraduate students who enrol in their chosen university should complete their studies on time, satisfy all minimum requirements, and obtain all learning outcomes within the stipulated time outlined by the university.

This manuscript is submitted on 16th June 2021 and accepted on 18th January 2022.

A'zraa Afhzan Ab Rahim and Norlida are with the School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor (e-mail: azraa@uitm.edu.my)

*Corresponding author
Email address: azraa@uitm.edu.my

Various methods have been suggested to address this issue, including personal tutoring, consultation with counselors, and pro-active academic advisors. However, all these interventions are only meaningful if the at-risk student can be identified earlier to enable positive interventions. Nevertheless, analyzing massive amounts of educational data and extracting essential information to predict whether a student requires intervention or not is a difficult task. Even more so, in the context of institutions of higher learning (IHLs), where an educator oversees a high number of students, it is ineffective to rely on individual expertise to examine data and identify patterns to detect poor-performing students.

To address the consequences of this phenomenon, numerous research has attempted to design models that predict the future academic performance of students enrolled in higher education [1], [2]. Educational Data Mining (EDM) such as Support Vector Machine (SVM) [3], Decision Trees (DT) [4], [5], Logistic Regression [6], K-nearest neighbor (KNN) [7], and Naïve Bayes (NB) [8] have been used to predict students' academic performance. Nonetheless, the voluminous educational data and the complexity of predicting students' academic performance require more sophisticated models. Therefore, this research postulate that ensemble classifiers, where more than one classifier is employed, can address this gap. The ensembles implemented in this paper include Extreme Gradient Boosting (XGB), Random Forest (RF), and the heterogeneous ensemble method (HEM). The ensemble architecture is posited to increase the accuracy of the prediction model and improve the ability of classifiers to classify the samples into their correct class. The ability to successfully predict and profile at-risk students provides intelligent insights to academic advisors, tutors, and counselors to tailor specific intervention plans and actions.

Engineering students face a myriad of challenges during their undergraduate endeavors, leading to poor academic performance, thus impacting their timeline to graduate or even causing premature withdrawal from the university [9], [10]. Several researchers have examined the conventional predictability of cognitive variables (academic and intellectual abilities) such as high school grades and standardized tests such as SAT that affect students' success [11], [12], [13], [14], [15], [16], [17], [18]. However, during the last decade, there is also a movement to include non-cognitive factors when predicting students' academic performance as demonstrated by [19], [20], [21], [22], [23], [24], [25], [26], [27]. But there is still a lack of synergy between social science and engineering due to faculties working in silos. Therefore, it is necessary to integrate social science and engineering tools to address the insufficiency of holistic predictors.

The purpose of this paper is to establish a framework to answer the following research questions (*RQ*):

RQ1: Which features among cognitive, non-cognitive, or combination of cognitive and non-cognitive produces the most accurate and highest recall value?

RQ2: Between single classifiers and ensemble, which will produce a better overall performance in predicting the correct class for each sample?

The ability to answer the two research questions indicates that the study has fulfilled the following two research objectives (*RO*):

RO1: To identify which among the cognitive, non-cognitive, and demographic are pivotal predictors when predicting students' academic performance.

RO2: To determine the best performing classifier with the highest accuracy and best recall for a multi-class classification problem. Correctly classifying students' future grades is essential so that educators and administrators can implement remedial interventions for struggling students.

The remainder of this work is organized as follows: Section 2 summarizes the scientific contributions of this work by describing the related work and its limitations; Section 3 describes the experimental methodology, highlights, and analyses the datasets used, and describes the performance metric used to evaluate the classifiers; Section 4 presents the experimental results and discussion; Section 5 concludes the paper and suggests some future research directions.

II. BACKGROUND

A. *Predictors of Academic Performance*

Engineering students encounter a wide spectrum of challenges in university [28]. The implication of the demanding nature of engineering curricula combined with the challenges of socially adapting to the university environment may result in students changing their major, abandon their studies altogether, or needs more time to complete their studies [28]–[30]. It is essential to investigate the paradox of highly credentialed and previously successful students that somehow fail to achieve the same success in university [31].

Students who are accepted into an engineering program after fulfilling the minimum academic criteria set by the university, indicating that they have strong academic credentials when entering university hence perceived equal potential for academic success in the undergraduate curriculum. However, while some flourish at the university, many flounder near the bottom of the pack. This observation suggests that factors other than cognitive ability play an essential role in determining students' success or failure to obtain an engineering degree [32].

In the pantheon of literature on students' academic performance prediction (SAPP), several researchers suggested the inclusion of non-cognitive factors to increase prediction accuracy [10], [33]. For example, Al Sheeb et al. [10] examined the experiences of 144 first-year engineering students and 175 sophomore engineering students on their first-year studies at Qatar University. They identified a positive link between engineering students' attitudes and motivation and successful transition into higher education, highlighting the importance of identifying vulnerable engineering students at risk of struggling academically early in the transition period. Based on their findings, this research attempts to identify the non-cognitive attributes that could help improvise prediction of student performance.

A comprehensive literature review undertaken in 2020/2021 summarized the most common non-cognitive factors as in Fig. 1. It provides an overview of the three main factors that influence academic performance: cognitive, non-cognitive, and demographic, along with their smaller components.

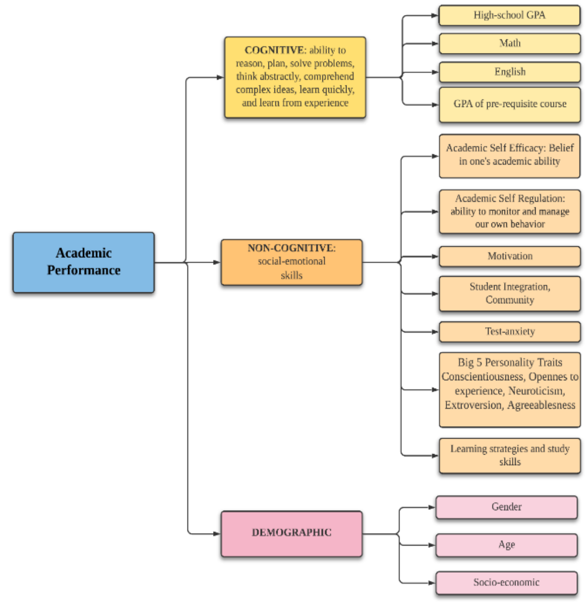


Fig. 1 Predictors of Academic Performance

B. *Students' Academic performance Prediction Algorithms*

Traditional statistical analysis approaches, including but not limited to logistic regression, have been used to estimate the students' retention in higher education.

Applying regression analysis in modeling and predicting student academic performance has long been a central research area in the tertiary education context. This section briefly introduces some of the valuable contributions under this domain.

A systematic mapping review on students' performance analysis using big data predictive models found that the predictive models used are varied [29]. Still, the majority of literature focused on one particular model, which is the regression model. Linear regression represents the data as a linear graph and primarily indicates the link between the independent variable with the output. This relationship provides an indication of strong, positive, weak, negative, or no relationship. Commonly, linear regression creates a straight-line fitness of the data. Still, in some cases where a straight line does not materialize, this will be classified as a nonlinear relationship [34].

Ting [35] modeled the relationship between standardized test scores and psychosocial variables to predict students' academic success using multiple stepwise regression. SAT and the Non-Cognitive Questionnaire (NCQ) were the predicting variables used in the regression to predict a GPA of 690 engineering freshmen (Caucasian and African American) at North Carolina State University.

Pinxten et al. [36] used statistical modeling and multiple linear regression to investigate the relation between first-year student success and the influence of cognitive measures, study strategies, and advice of the secondary school teacher board. A stepwise approach was administered wherein independent variables were included in the analysis following the empirical evidence obtained from the literature. Because academic factors were more influential than other independent variables, math and physics GPA were examined first. Factors other than academic measures were later augmented during the next stage. *Table I* summarizes the algorithms that have been used to predict students' academic performance.

TABLE I.
STUDENTS' ACADEMIC PERFORMANCE PREDICTION ALGORITHMS

	Algorithm	References
1	Regression	[37], [35], [38], [39], [40], [41], [42], [43], [44], [19], [45], [46], [21], [22], [17], [47], [26], [27], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [13], [31], [58], [59], [10], [29]
2	Decision Tree	[43], [60], [23], [26], [51], [53], [55], [61]
3	SVM	[45], [24], [17], [26], [53], [13]
4	NB	[23], [26], [51], [53], [55], [61]
5	K-NN	[26], [62], [51], [53], [61]
6	Ensemble	[53], [61], [63]

However, current research trends focus on educational data mining (EDM) methods to study the same issues. EDM methods have matured over the years and can now achieve high accuracy and be robust against missing data. Data mining involves a set of techniques for the extraction, identification, and understanding of patterns and trends in a large dataset. In data mining, machine learning, statistics, and visualization strategies are employed to discover and communicate easily understandable information to humans. When harnessed to its full potential, data mining algorithms enable profiling, prediction of future trends, and behaviors that assist the stakeholders in enhancing the quality of the decision-making process [64][65][66].

Several data mining techniques can be applied to analyze the voluminous educational data available over the years in IHL's repositories. Among them are classification and regression, which are commonly used when dealing with performance prediction problems. Classification is used where the predicted variable is categorical. On the other hand, when the predicted variable is continuous, regression is used. The most widely used methods for classification and regression include decision trees, Logistic Regression (LR), Decision Tree (DT), Naive Bayes (NB), K-Nearest Neighbour (KNN), and Support Vector Machines (SVM) [67].

However, the increase of student data and the complexity of predictive features led us to propose the ensemble architecture to overcome the limitation of high variance and overfitting imposed by single prediction models. A combination of different base classifier algorithms (heterogeneous) such as the Naïve Bayes (NB), Logistic Regression (LR), Decision Tree (DT), K-nearest neighbor (KNN), and Support Vector Machine (SVM) is employed to enhance predictive accuracy and recall value. These algorithms are treated as weak classifiers, which are then fed into a meta-classifier. The prediction performance of the ensembles is compared with the traditional single classifiers: NB, LR, DT, KNN, and SVM.

To ensure that an ensemble model exceeds the accuracy of all of its single base classifiers, the base classifiers need to be accurate and distinct enough to elicit the underlying structure of the data. The task of building such an ensemble model is non-trivial as it requires assessing a diverse range of possible base classifiers and how they complement each other's strengths and weaknesses. One example of a good ensemble model is the one used by the winning team of the 2017 KDD cup, where they combined 13 base classifiers made up of DT, NN, and linear models to form an ensemble architecture [68].

III. METHODOLOGY

This paper attempts to compare the performance of different single classifiers such as in *Table 1* against ensemble classifiers such as Extreme Gradient Boosting (XGB), Random Forest (RF), and Heterogeneous Ensemble Method (HEM) in predicting engineering students' academic performance. Assessing the effect of different input features such as using only cognitive (CGPA), only non-cognitive (MSLQ), and the combination of cognitive and non-cognitive (CGPAMSLQ) is another target for this research. The methodology used to attain these goals is depicted in Fig. 2.

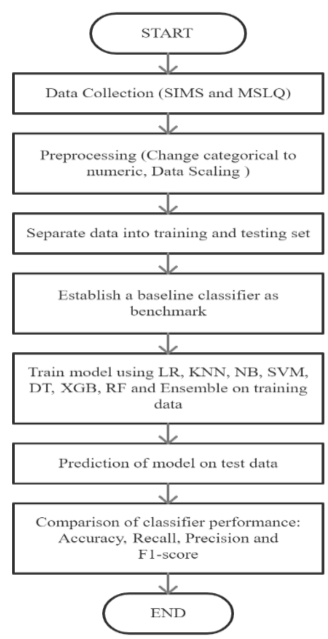


Fig. 2 Research Methodology

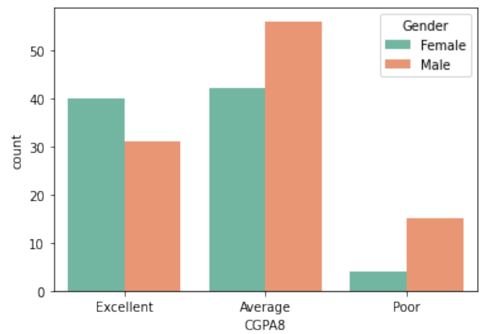
A. Data Collection

Data is obtained from 188 Faculty of Electrical Engineering's students from two programs, EE241 (Electronics) and EE242 (Electrical), where 86 (46%) are female and 102 (54%) are males. All 188 students are from the March 2018 intake. Their final semester's CGPA, obtained from the Students Information and Management System (SIMS), is used as the predicted output and referred to as *CGPA8*. *CGPA8* are se gregated into three classes, *'Poor'* ($CGPA < 3$), 'Average' ($3 \leq CGPA < 3.5$) and 'Excellent' ($CGPA \geq 3.5$) making this a multi-class classification problem. Table II shows the number of samples for each class.

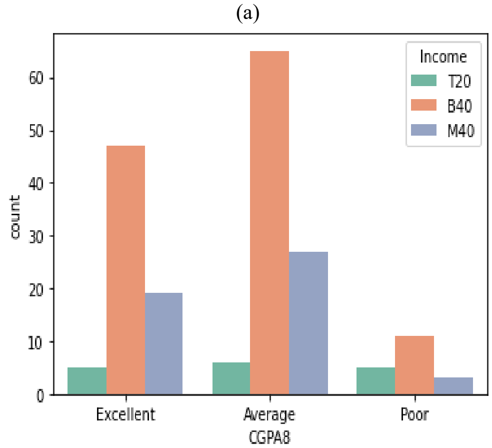
TABLE II
NUMBER OF SAMPLES FOR EACH CLASS

Class	Number of samples
Poor	19
Average	98
Excellent	71

More than 50% of the samples belong to the *'Average'* class, while the '*Poor'* class comprises five times fewer samples. The disparity between classes creates data imbalance which will later prove to be significant when evaluating the performance of classifiers.



(a)



(b)

Fig. 3 Breakdown of each class according to (a) Gender and (b) Income

To gain more insight into the samples, Fig. 3*a* illustrates the composition of male and female students belonging to each class where female students achieve higher *CGPA8* than their male peers. As for their demographic background, which considers their family's monthly household income (HI), each class has a similar distribution of T20 (HI >RM10,971), M40 ($RM4850 \leq HI \leq RM10{,}971$), and B40 ($HI < RM4850$), but it is obvious from Fig. 3b that students from the B40 family dominate. This makes it more essential to properly categorize students so that B40 students are ensured full support from the faculty should they be among the *'Average'* and *'Poor'* performing classes.

This research is unique since it also considers the non-cognitive aspect by measuring their motivational inclination via The Motivated Strategies for Learning Questionnaire (MSLQ). The survey was distributed to 243 Faculty of Electrical Engineering students through email, where a link to the google form of the MSLQ survey was attached. Students' participation is voluntary. A high response rate of 77%, where 188 students responded to the survey, ensures that the sample is representative of the population.

Table III presents the features used to classify students' academic achievement, indicated by *CGPA8*. The classifications are separately carried out when using only cognitive features (GPA), only non-cognitive features (MSLQ), and finally, a combination of both GPA and MSLQ.

TABLE III
FEATURES OF THE DATASET

FEATURE NAME	DATA TYPE
DEMOGRAPHIC	
Gender	Categorical
Income	Categorical
GPA	
Program	Categorical
ECE431	Numeric
ECE521	Numeric
ELC590	Numeric
MAT455	Numeric
MAT575	Numeric
CGPA1	Numeric
GPA1	Numeric
CGPA2	Numeric
GPA2	Numeric
GPA8	Numeric
CGPA8	Categorical
MSLQ	
Intrinsic Goal Orientation	Numeric
Extrinsic Goal Orientation	Numeric
Task Value	Numeric
Control Of Learning Beliefs	Numeric
Self-Efficacy for Learning and Performance	Numeric
Test Anxiety	Numeric
Rehearsal	Numeric
Elaboration	Numeric
Organization	Numeric
Critical Thinking	Numeric
Metacognitive Self-Regulation	Numeric
Time and Study Environment	Numeric
Effort Regulation	Numeric
Peer Learning	Numeric
Help Seeking	Numeric

B. Data Pre-processing

Raw data contains both categorical (nominal and ordinal) and numeric data. However, since *Python* is the programming language used, all categorical variables must first be converted to numerical values, done via one-hot-encoding in Python.

Categorical data refers to variables composed of label values, such as gender, which can be either female or male. The majority of machine learning methods demand that any input or output variables have a numeric value. This requires that all categorical data be transformed into integers.

One-hot-encoding is a technique that can perform such transformation. We use one-hot encoding to convert each categorical value to a new categorical column and assign those columns a binary value of 1 or 0.

C. Data Separation

Data has to be separated into training and testing to avoid overfitting. We allocated 70% of our data for training, while the rest were used as test data. Furthermore, the training dataset is divided into K equivalent subsets analogous to the technique of K-fold cross-validation. Base classifiers are trained using the (K − 1) subsets, whereas the Kth subset is kept for validation. Once training is complete, each base classifier is validated using the Kth subset and the testing data. The decision of individual classifiers is then used as a new set of training and testing data for the meta-classifier [69].

D. Baseline Classifier

It is pertinent to establish a baseline classifier as a benchmark so that all the classification results can be compared to the baseline to determine whether the classifier performance improved or degrade. The baseline classifier, in our case, just predicts the majority class.

E. Ensemble Models

Traditionally, EDM models using single classifiers or regressors has been a favorite method in designing models to predict students' Grade Point Average (GPA) and retention, as evident by a high number of research publications adopting this method such as [35][70][38][39][40][41][42][44][19][46]. However, their result sometimes led to unsatisfactory accuracy or generalizability. Furthermore, they are limited to solving relatively simple formulated problems where only one or at most two out of the three factors (cognitive, non-cognitive, demographic) are considered. Our research addresses this problem by including non-cognitive factors alongside the more commonly used cognitive measures.

Ensembles combine more than one ML algorithm to achieve higher accuracy and improve the robustness of a model. The ensemble learning method aims to improve a model's predictive capability by integrating many ML algorithms instead of the traditional single classifier. It is formed using a collective set of base (single) learning algorithms and has been applied in numerous classification exercises. Ensemble learning methods can be homogeneous (same base classifiers), such as Bagging (bootstrap aggregation) and random forest. It can also be heterogeneous (different base classifiers) for instance voting, and stacking [71][69][72][73]. *Fig. 4* depicts the different types of ensemble models, which will be explained in the subsequent sub-sections.

Bagging

Bagging ensemble is also called bootstrap aggregation [74]. The same type of classifiers are used as weak learners where they learn in parallel, independent of one another, and then their average is calculated. The samples are bootstrapped from the original dataset and are used to train each independent classifier. Consequently, the combination of estimates from multiple homogeneous weak learners contributes to variance reduction. Random Forest (RF) is an example of Bagging using Decision Tree (DT) as the weak learner. The original training data is randomly partitioned to be fitted on the DTs.

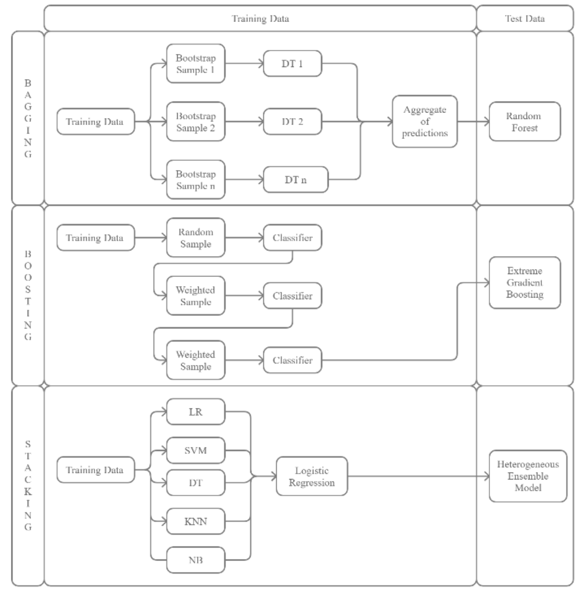


Fig. 4 Different types of Ensemble Models

Boosting

Boosting is a sequential process in which each succeeding model seeks to correct the mistakes of the preceding model. In Boosting, subsequent models are impacted by their predecessors. Learners are taught sequentially using this technique, with early learners fitting simple models to the data and examining errors. When a hypothesis poorly classifies an input, the weight of the hypothesis is increased to ensure that the next hypothesis correctly classifies it. This way, weak learners become better classifiers by integrating all of the weak learners.

Stacking

Stacking is a heterogeneous ensemble method gaining popularity, especially among the data mining competition community [71]. The stacking architecture targets to reduce both variance and bias by searching for the optimal blending of base learners, resulting in the best balance between variance and bias [73]. In a two-tier stacking architecture, the first tier comprises of base learners, while the second tier consists of a meta-learner. Meta-learning uses the decisions from the base learner as inputs to train another machine learning algorithm to produce a second-tier classification. The idea of using stacking to enhance classification performance is based on the premise that all base learners have good prediction performance [71][63]. Overall, the ensemble stacking method flourishes when the base algorithms are diverse [71][75][63]. *Table IV* denotes the hyperparameters used to train each classifier in *Python*.

Model Validation

Cross-validation is a model validation approach used to determine the generalizability of statistical analysis results to an independent dataset. This article employs two widely used cross-validation techniques: random hold-out (which randomly assigns 80% of the data to the training set and 20% to the test set) and the 5-fold cross-validation [76].

It should be emphasized that the cross-validation method is only utilized on the training data to prevent data leakage into the test set, leading to overfitting.

TABLE IV
HYPERPARAMETER USED FOR EACH CLASSIFIER

Model	Hyperparameter
LR	penalty=l2, C=1.0, solver=lbfgs
KNN	n_neighbors=5, leaf_size=30
NB	No parameter
SVM	C=1.0, kernel=rbf, degree=3, gamma=scale
DT	criterion=gini, splitter=best
RF	n_estimators=100, criterion='gini', bootstrap=True
XGB	base_score=0.5, booster=gbtree, importance_type=gain, learning_rate=0.300000012, tree_method=exact
HEM	Default for base and meta-learner

Model Performance Evaluation

There are many performance indicators to ascertain the performance of each classifier. Therefore, it is quintessential to select the most relevant performance metric to the problem that is being studied. A thorough explanation of the metrics used to assess classification models can be found in work done by Tharwat [77]. The Confusion Matrix is commonly used to evaluate classification models. The following are the terms used in developing a Confusion Matrix [78]:

True Positive (TP): When the classifier correctly predicts positive observation.

True Negative (TN): When the classifier correctly predicts negative observation.

False Positive (FP): When the classifier wrongly predicts positive observation.

False Negative (FN): When the classifier wrongly predicts negative observation.

Based on these four terms, the accuracy of the prediction, which is the ratio of correctly classified output to the total number of observations, refer to *(1)*

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN} \quad (1)$$

However, when dealing with imbalanced data, accuracy alone is insufficient to measure the goodness of a model [83]; therefore, we use another measure, the recall value. Recall answers the question of how many true positives the classifier was able to predict correctly, refer to *(2)*.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

IV. RESULT AND DISCUSSION

A baseline classifier sometimes referred to as a naïve classifier, is first established to evaluate the performance of other classifiers. Our baseline classifier predicts the majority observation. In this case, it predicts all samples to belong to the *'Average'* class, as shown in *Fig. 5* since the *'Average'* class has the most number of samples.

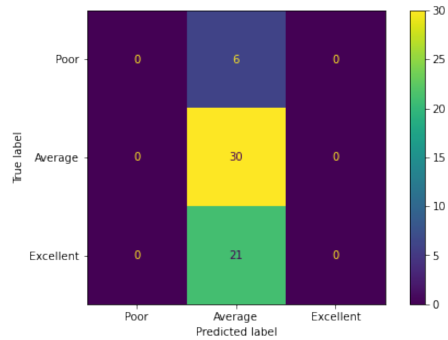


Fig. 5 Confusion matrix of baseline classifier

Subsequent classifiers are compared to this baseline performance to determine if the model improves the classification performance.

The boxplot in *Fig. 6* summarizes the average accuracy across all classifiers when using only GPA, only MSLQ, and the combination of GPA and MSLQ. According to the boxplots in *Fig. 6*, the highest accuracy is consistently achieved when classifiers use only MSLQ as their input feature since almost 100% of their prediction achieves at least 80% accuracy. Classifiers that only used GPA as their input feature resulted in the worst accuracy since nearly 75% of their prediction achieved less than 80% accuracy. However, using GPA and MSLQ together produced prediction accuracy with the least variability, as shown by the shorter length of the box and whiskers.

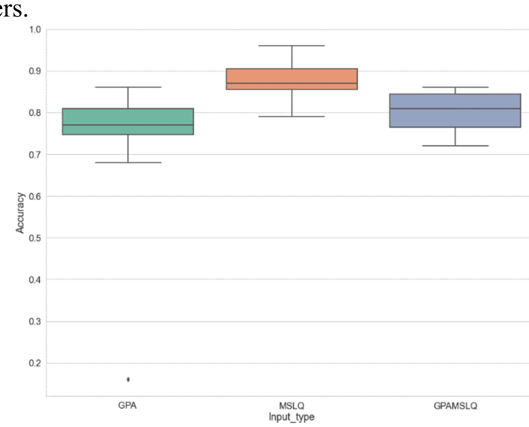


Fig. 6 Comparison of accuracy between using only GPA, only MSLQ, and combination of GPA and MSLQ

However, care needs to be taken when judging the performance of a classifier solely by its accuracy. For example, the accuracy for the Logistic Regression using only GPA as input features resulted in 77% accuracy. Even though it is an acceptable value since the base classifier's accuracy was 53%, but from Fig. 7, it is observed that the recall for the *'Poor'* class is only at 0.17%, predicting only one out of six correctly. Therefore, it is pertinent to also consider recall values to evaluate classification performance.

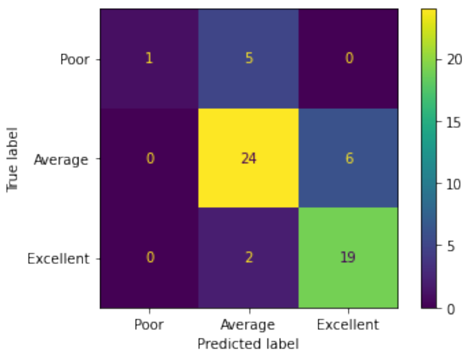


Fig. 7 Confusion matrix for Logistic Regression (LR) classifier using only GPA as input feature

Fig. 8 represents a boxplot with the distribution of recall generated in modeling *CGPA8,* where GPA, MSLQ, and CPGAMSLQ were separately incorporated as the predictor variables. There appears to be quantitative evidence that incorporating only MSLQ leads to at least 75% of the samples achieving greater than 80% recall. Given the widely spread errors and large outliers indicating large error magnitudes, adopting only the non-cognitive factor far exceeds that of exclusively using GPA in all modeling scenarios. The outliers in boxplots when strictly using GPA as the input depict extreme errors encountered in predicting students' *CGPA8* category. However, differences between MSLQ and GPAMSLQ are less conspicuous.

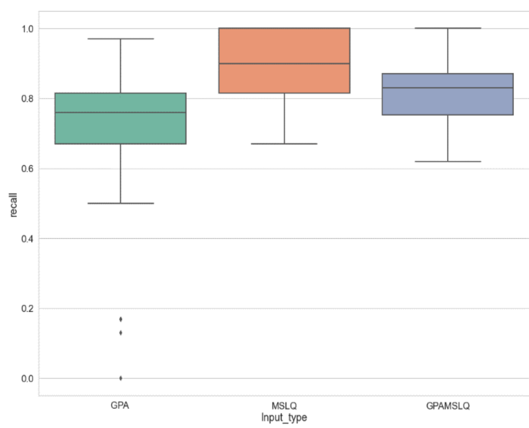


Fig. 8 Comparison of accuracy between using only GPA, only MSLQ, and combination of GPA and MSLQ

The recall performance of all classifiers using three different types of input features is illustrated in Fig. 9. Since the edge of the boxplot denotes the upper and the lower quartile errors, and the central margin shows the median value of the error, Fig. 9 confirms that exclusively using GPA as an input feature resulted in the worst recall across all models. This is evident by the median of all classifiers using only GPA as input to be lower than the median of their competing counterparts. Judging by the length of the box, the worst recall is the SVM classifier using only GPA as the input feature, which has the largest dispersion.

It can also be deduced that the recall for ensemble-type classifiers (XGB, RF, and Ensemble) are relatively better than their single classifier counterparts' models, as the former's quartiles and medians are significantly smaller, as evidenced in Fig. 9.

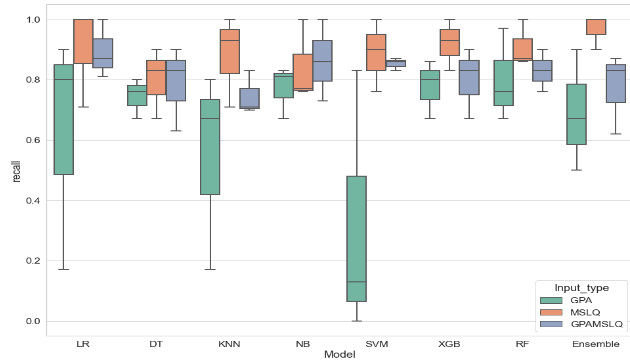


Fig. 9 Recall of all classifiers using GPA, MSLQ, and GPAMSLQ different types of input features

To evaluate the performance of each classifier in predicting the correct class labels for *'Poor'*, *'Average',* and *'Excellent'*, we refer to the boxplot in Fig.10. The three ensemble classifiers can predict the 'Poor' group better than single classifiers with a median recall above 70%. Ensemble classifiers also perform better when predicting the *'Average'* group judging by the recall above 90% and the lower variability indicated by the relatively shorter boxes.

From Fig. 10, it is observed that single classifiers can predict the *'Average'* and *'Excellent'* classes—all single classifiers have a median recall of more than 70%. However, SVM has high variability, from failing to predict the *'Excellent'* class and predicting up to approximately 90% recall.

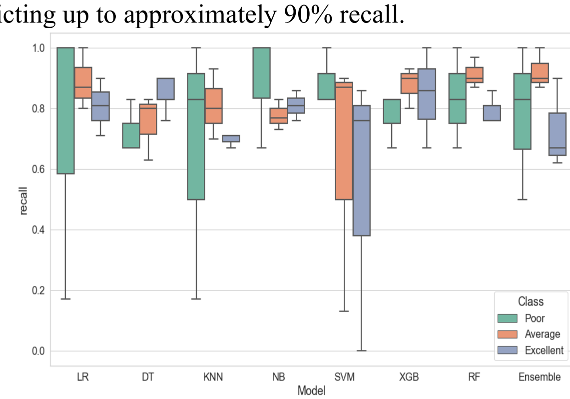


Fig. 10 Recall of all classifiers in predicting the multi-class output

Table V shows that the highest mean recalls belong to the ensemble classifiers, all with low standard deviations. On the other hand, single classifiers such as LR, KNN, and SVM's recall values have higher dispersion as indicated by the higher standard deviations.

TABLE V
MEAN AND STANDARD DEVIATIONS OF RECALL FOR DIFFERENT CLASSIFIERS

Model	*Mean*	*Standard Deviation*
RF	0.847	0.105
XGB	0.832	0.110
HEM	0.827	0.115
NB	0.826	0.113
LR	0.807	0.260
DT	0.790	0.099
KNN	0.724	0.236
SVM	0.687	0.360

The recall performance comparison between cognitive, non-cognitive, and the combination of cognitive and non-cognitive is highlighted in Table VI. As previously mentioned, the recall performance when using only MSLQ and combining GPA and MSLQ as the input feature is quite similar, as seen by the similar mean and standard deviations. However, using only GPA proves to be inferior compared to the other two. Table VI highlights that GPA has the lowest value for mean recall and the highest standard deviations among all three.

TABLE VI
MEAN AND STANDARD DEVIATIONS OF RECALL FOR DIFFERENT INPUT FEATURES

Model	*Mean*	*Standard Deviation*
MSLQ	0.889	0.121
GPAMSLQ	0.848	0.100
GPA	0.659	0.268

V. CONCLUSION AND FUTURE WORK

The main problem investigated in this research was the inconsistencies of findings when implementing EDM using traditional single classifiers. After comparing the performance of single classifiers against their ensemble opponent, it was found that ensemble classifiers such as XGB, RF, and HEM emerged as better classifiers, both in terms of accuracy and recall. The other concern addressed revolves around which predictors among the cognitive, non-cognitive, or the combination of both predicts student's academic performance better. The findings from our research suggest that using cognitive features such as the GPA of first-year courses is unable to successfully predict students' final semester's CGPA accurately. On the side, using only non-cognitive features resulted in higher accuracies and recall but there is also a lot of variabilities since students are individuals with various motivational spectrum.

However, the amalgamation of cognitive and non-cognitive led to overall better accuracy and recall with less variability. The potential to accurately predict student performance in an engineering undergraduate program using heterogeneous ensemble methods (HEM) paves the way for improving their educational outcomes. Accurate performance prediction allows university administration to identify struggling at-risk students, as indicated by lackluster GPA, to implement targeted remedial intervention to avoid immature withdrawals. Optimal allocation of resources and instructions will benefit the university and students to achieve their best potential. Intervention is critical for students with low academic performance. They can be properly helped with correct academic and personal means designed to their needs to enhance their achievement and sense of satisfaction, increasing their chances to succeed in engineering. Finally, the problem of student attrition and the late completion of engineering majors can be alleviated.

Since traditional predictors tend to focus more on cognitive ability, this research investigates more comprehensive factors, considering the joint contribution of non-cognitive factors and giving a more holistic view of what it takes for a student to succeed.

This research can be expanded in various ways, and future work will extend to the following directions. Since ensemble classifiers have been shown to perform better than traditional single classifiers on a small-sized dataset, it is recommended that they also be evaluated on a larger dataset to determine the classifier's generalizability. Additionally, data resampling methods such as Synthetic Minority Oversampling Technique (SMOTE) can be applied to overcome the constraint of an imbalanced dataset. Furthermore, hyperparameter tuning strategies can be used to improve the accuracy and recall of models.

REFERENCES

[1] R. K. Mothilal, T. de Laet, T. Broos, and M. Pinxten, "Predicting First-Year Engineering Student Success: From Traditional Statistics to Machine Learning," in *SEFI Annual Conference 2018: Creativity, Innovation and Entrepreneurship for Engineering Education Excellence*, 2019, pp. 322–329.

[2] E. Lauría, "Framing Early Alert of Struggling Students as an Anomaly Detection Problem: An Exploration," Apr. 2021, pp. 26–35. doi: 10.5220/0010471900260035.

[3] N. Naicker, T. Adeliyi, and J. Wing, "Linear Support Vector Machines for Prediction of Student Performance in School-Based Education," *Mathematical Problems in Engineering*, vol. 2020, 2020, doi: 10.1155/2020/4761468.

[4] H. Hamsa, S. Indiradevi, and J. J. Kizhakkethottam, "Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm," *Procedia Technology*, vol. 25, pp. 326–332, 2016, doi: 10.1016/j.protcy.2016.08.114.

[5] A. Shanthini, G. Vinodhini, and R. M. Chandrasekaran, "Predicting Students' Academic Performance in the University Using Meta Decision Tree Classifiers," *Journal of Computer Science*, vol. 14, no. 5, pp. 654–662, 2018, doi: 10.3844/jcssp.2018.654.662.

[6] M. E. Urrutia-Aguilar, R. Fuentes-García, V. D. Mirel Martínez, E. Beck, S. O. León, and R. Guevara-Guzmán, "Logistic Regression Model for the Academic Performance of First-Year Medical Students in the Biomedical Area," *Creative Education*, vol. 07, no. 15, pp. 2202–2211, 2016, doi: 10.4236/ce.2016.715217.

[7] M. Mardolkar and N. Kumaran, "Forecasting and Avoiding Student Dropout Using the K-Nearest Neighbor Approach," *SN Computer Science*, vol. 1, no. 2, Mar. 2020, doi: 10.1007/s42979-020-0102-0.

[8] F. Razaque *et al.*, "Using Naïve Bayes Algorithm to Students' bachelor Academic Performances Analysis."

[9] M. Besterfield-Sacre, C. J. Atman, and L. J. Shuman, "Characteristics of Freshman Engineering Students: Models for Determining Student Attrition in Engineering," *Journal of Engineering Education*, vol. 86, no. 2, pp. 139–149, 1997, doi: 10.1002/j.2168-9830.1997.tb00277.x.

[10] B. Al Sheeb, G. M. Abdella, A. M. Hamouda, M. S. Abdulwahed, and To, "Predictive Modeling of First-Year Student Performance in Engineering Education Using Sequential Penalization-Based Regression," *Journal of Statistics and Management Systems*, vol. 22, no. 1, pp. 31-50, 2019, doi: 10.1080/09720510.2018.1509817.

[11] U. Mat, N. Buniyamin, and P. M. Arshad, "Educational Data Mining Classifier for Semester One Performance to Improve Engineering Students Achievement," *Middle-East Journal of Scientific Research*, vol. 24, no. 2, pp. 338–346, 2016, doi: 10.5829/idosi.mejsr.2016.24.02.23023.

[12] N. Buniyamin, U. Mat, and P. M. Arshad, "Educational Data Mining for Prediction and Classification of Engineering Students Achievement," in *IEEE International Conference on Engineering Education (ICEED)*, 2015, pp. 49–53. doi: 10.1109/ICEED.2015.7451491.

[13] E. Y. Obsie and S. A. Adem, "Prediction of Student Academic Performance using Neural Network, Linear Regression and Support Vector Regression: A Case Study," *International Journal of Computer Applications*, vol. 180, no. 40, pp. 39–47, 2018, doi: 10.5120/ijca2018917057.

[14] O. E. Aru, N. Chukwuchekwa, and F. K. Opara, "Exploiting the Core Academic Performance Prediction Parameters Using a Neuro-Fuzzy Model," *American Journal of Engineering Research*, vol. 6, no. 6, pp. 232–245, 2017.

[15] E. Isong, U. Kingsley, and G. Ansa, "Cognitive Factors in Students' Academic Performance Evaluation using Artificial Neural Networks," *Information and Knowledge Management*, vol. 8, no. 5, pp. 57–71, 2018.

[16] S. A. Naser, I. Zaqout, M. A. Ghosh, R. Atallah, and E. Alajrami, "Predicting Student Performance Using Artificial Neural Network: in the Faculty of Engineering and Information Technology," *International Journal of Hybrid Information Technology*, vol. 8, no. 2, pp. 221–228, 2015, doi: 10.14257/ijhit.2015.8.2.20.

[17] S. Huang and N. Fang, "Predicting Student Academic Performance in an Engineering Dynamics Course: A Comparison of Four Types of Predictive Mathematical Models," *Computers & Education*, vol. 61, pp. 133–145, 2013, doi: 10.1016/j.compedu.2012.08.015.

[18] M. Aryee, "College Students' Persistence and Degree Completion in Science , Technology , Engineering , and Mathematics (STEM): The Role of Non-Cognitive Attributes Of Self-Efficacy , Outcome Expectations , And Interest," Seton Hall University, 2017.

[19] A. N. Paimin, R. Hadgraft, J. K. Prpic, and M. Alias, "An Examination of Learning Strategy, Interest, Intention and Academic Performance: Case Studies of Australia and Malaysia," 2011.

[20] Q. Jin, P. K. Imbrie, J. J. J. Lin, and X. Chen, "A Multi-Outcome Hybrid Model for Predicting Student Success in Engineering," 2011.

[21] P. H. Hsieh, J. R. Sullivan, D. A. Sass, and N. S. Guerra, "Undergraduate Engineering Students' Beliefs, Coping Strategies, and Academic Performance: An Evaluation of Theoretical Models," *Journal of Experimental Education*, vol. 80, no. 2, pp. 196–218, 2012, doi: 10.1080/00220973.2011.596853.

[22] A. M. Shaharoun, M. Puteh, and H. Norehan Haron, "Does Motivation Affect Students' Understanding and Performance in Engineering Statics ?," *Procedia - Social and Behavioral Sciences*, 2012, doi: 10.1016/j.sbspro.2012.09.646.

[23] E. Osmanbegovi and M. Suljic, "Data Mining Approach for Predicting Student Performance," *Economic Review : Journal of Economics and Business*, vol. X, no. 1, pp. 3–12, 2012.

[24] S. A. Kumar and M. N. Vijayalakshmi, "Appraising the Significance of Self Regulated Learning in Higher Education Using Neural Networks," *International Journal of Engineering Research and Development*, vol. Volume 1, no. Issue 1, pp. 9–15, 2012.

[25] M. F. Musso, E. Kyndt, E. C. Cascallar, and F. Dochy, "Predicting General Academic Performance and Identifying the Differential Contribution of Participating Variables Using Artificial Neural Networks," *Frontline Learning Research*, vol. 1, pp. 42–71, 2013.

[26] G. Gray, C. McGuinness, and P. Owende, "An Application of Classification Models to Predict Learner Progression in Tertiary Education," in *IEEE International Advance Computing Conference*, 2014, pp. 549–554. doi: 10.1109/IAdCC.2014.6779384.

[27] L. M. Larson, K. M. Pesch, S. Surapaneni, V. S. Bonitz, T. F. Wu, and J. D. Werbel, "Predicting Graduation: The Role of Mathematics/Science Self-Efficacy," *Journal of Career Assessment*, vol. 23, no. 3, pp. 399–409, 2015, doi: 10.1177/1069072714547322.

[28] A. W. Astin, *What Matters in College: Four Critical Years Revisited*. San Francisco, CA: Jossey Bass Higher & Adult Education Series, 1993.

[29] B. A. Al-Sheeb, A. M. Hamouda, and G. M. Abdella, "Modeling of student academic achievement in engineering education using cognitive and non-cognitive factors," *Journal of Applied Research in Higher Education*, vol. 11, no. 2, pp. 178–198, 2019, doi: 10.1108/JARHE-10-2017-0120.

[30] C. P. Veenstra, E. L. Dey, and G. D. Herrin, "A Model for Freshman Engineering Retention," *Advances in Engineering Education*, vol. 1, no. 3, pp. 1–31, 2009.

[31] M. Scheidt *et al.*, "Validity Evidence for the SUCCESS survey: Measuring Non-Cognitive and Affective Traits of Engineering and Computing Students," 2018.

[32] K. Meyers *et al.*, "Some Perspectives on First Year Engineering Education," 2008.

[33] D. G. Young, "The Case for an Integrated Approach to Transition Programmes at South Africa's Higher Education Institutions," *Journal of Student Affairs in Africa*, vol. 4, no. 1, pp. 17–32, 2016, doi: 10.14426/jsaa.v4i1.142.

[34] A. Al Mazidi and E. Abusham, "Study of General Education Diploma Students' Performance and Prediction in Sultanate of Oman, Based on Data Mining Approaches," *International Journal of Engineering Business Management*, vol. 10, pp. 1–11, 2018, doi: 10.1177/1847979018807020.

[35] S. M. R. Ting, "Predicting Academic Success of First-Year Engineering Students from Standardized Test Scores and Psychosocial Variables," *International Journal of Engineering Education*, vol. 17, no. 1, pp. 75–80, 2001.

[36] M. Pinxten, C. Van Soom, C. Peeters, T. De Laet, and G. Langie, "At-Risk at The Gate: Prediction of Study Success of First-Year Science and Engineering Students in an Open-Admission University in Flanders—Any Incremental Validity of Study Strategies?," *European Journal of Psychology of Education*, vol. 34, pp. 45–66, 2019, doi: 10.1053/euhj.2000.2250.

[37] W. L. Gorr, D. Nagin, and J. Szczypula, "Comparative Study of Artificial Neural Network and Statistical Models for Predicting Student Grade Point Averages," *International Journal of Forecasting*, vol. 10, pp. 17–34, 1994, doi: 10.1016/0169-2070(94)90046-9.

[38] B. F. French, J. C. Immekus, and W. Oakes, "A Structural Model of Engineering Students Success and Persistence," in *Frontiers in Education Conference*, 2003, vol. 1, p. T2A19-T2A24. doi: 10.1109/FIE.2003.1263286.

[39] B. F. French, J. C. Immekus, and W. C. Oakes, "An Examination of Indicators of Engineering Students' Success and Persistence," *Journal of Engineering Education*, vol. 94, no. 4, pp. 419–425, 2005, doi: 10.1002/j.2168-9830.2005.tb00869.x.

[40] A. M. Kosnin, "Self-regulated Learning and Academic Achievement in Malaysian Undergraduates," *International Education Journal*, vol. 8, no. 1, pp. 221–228, 2007.

[41] L. E. Bernold, J. E. Spurlin, and C. M. Anson, "Understanding Our Students : A Longitudinal-Study of Success and Failure in Engineering With Implications for Increased Retention," *Journal of Engineering Education*, vol. 96, no. 3, pp. 263–274, 2007.

[42] P. Fenollar, S. Román, and P. J. Cuestas, "University Students' Academic Performance: An Integrative Conceptual Framework and Empirical Analysis," *British Journal of Educational Psychology*, vol. 77, no. 4, pp. 873–891, 2007, doi: 10.1348/000709907X189118.

[43] Z. Ibrahim and D. Rusli, "Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision Tree and Linear Regression," in *Proceedings of the 21st Annual SAS Malaysia Forum*, 2007, pp. 1–6.

[44] A. Olani, "Predicting First Year University Students' Academic Success," *Electronic Journal of Research in Educational Psychology*, vol. 7, no. 3, pp. 1053–1072, 2009.

[45] S. Huang, "Predictive Modeling and Analysis of Student Academic Performance in an Engineering Dynamics Course," Utah State University, 2011.

[46] J. De Winter and D. Dodou, "Predicting Academic Performance in Engineering Using High School Exam Scores," *International Journal of Engineering Education*, vol. 27, no. 6, pp. 1343–1351, 2011.

[47] P. M. Arsad, N. Buniyamin, and J. L. A. Manan, "Prediction of Engineering Students' Academic Performance Using Artificial Neural Network and Linear Regression: A Comparison," in *IEEE International Conference on Engineering Education*, 2013, pp. 43–48. doi: 10.1109/ICEED.2013.6908300.

[48] Z. Li, C. Shang, and Q. Shen, "Fuzzy-Clustering Embedded Regression for Predicting Student Academic Performance," in *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2016, pp. 344–351. doi: 10.1109/FUZZ-IEEE.2016.7737707.

[49] P. Bazelais, D. J. Lemay, and T. Doleck, "How does grit impact college students' academic achievement in science?," *European Journal of Science and Mathematics Education V*, vol. 4, no. 1, pp. 33–43, 2016.

[50] R. R. Senkpeil and E. J. Berger, "Impact of Non-Cognitive Factors on First-Year Performance," 2016.

[51] G. Gray, C. Mcguinness, and P. Owende, "Non-Cognitive Factors of Learning as Early Indicators of Students At-Risk of Failing in Tertiary Education," in *Non-cognitive Skills and Factors in Educational Attainment*, Sense`Publishers, 2016, pp. 199–237.

[52] D. F. Shell, L. K. Soh, A. E. Flanigan, and M. S. Peteranetz, "Students' Initial Course Motivation and Their Achievement and Retention in College CS1 Courses," in *ACM Technical Symposium on Computing Science Education*, 2016, pp. 639–644. doi: 10.1145/2839509.2844606.

[53] G. Gray, C. McGuinness, P. Owende, and M. Hofmann, "Learning Factor Models of Students at Risk of Failing in the Early Stage of Tertiary Education," *Journal of Learning Analytics*, vol. 3, no. 2, pp. 330–372, 2016, doi: 10.18608/jla.2016.32.20.

[54] S. M. Broghammer, "Grit As a Predictor of Academic Success for First-Time Undergraduate Students," University of Northern Colorado, 2017.

[55] S. Sultana, S. Khan, and M. A. Abbas, "Predicting Performance of Electrical Engineering Students Using Cognitive and Non-Cognitive Features for Identification of Potential Dropouts," *International Journal of Electrical Engineering Education*, vol. 54, no. 2, pp. 105–118, 2017, doi: 10.1177/0020720916688484.

[56] L. Fonteyne, W. Duyck, and F. De Fruyt, "Program-Specific Prediction of Academic Achievement on the Basis of Cognitive and Non-Cognitive Factors," *Learning and Individual Differences*, no. May, pp. 34–48, 2017, doi: 10.1016/j.lindif.2017.05.003.

[57] C. Mason, J. Twomey, D. Wright, and L. Whitman, "Predicting Engineering Student Attrition Risk Using a Probabilistic Neural Network and Comparing Results with a Backpropagation Neural Network and Logistic Regression," *Research in Higher Education*, vol. 59, no. 3, pp. 382–400, 2018, doi: 10.1007/s11162-017-9473-z.

[58] A. I. Adekitan and E. Noma-Osaghae, "Data Mining Approach to Predicting the Performance of First Year Student in a University Using the Admission Requirements," *Education and Information Technologies*, vol. 24, pp. 1527–1543, 2019, doi: 10.1007/s10639-018-9839-7.

[59] R. K. Mothilal, T. De Laet, T. Broos, and M. Pinxten, "Predicting First-Year Engineering Student Success: From Traditional Statistics to Machine Learning," in *SEFI Annual Conference 2018: Creativity, Innovation and Entrepreneurship for Engineering Education Excellence*, 2019, pp. 322–329.

[60] M. Wook, Y. H. Yahaya, N. Wahab, M. R. M. Isa, N. F. Awang, and H. Y. Seong, "Predicting NDUM Student's Academic Performance Using Data Mining Techniques," in *International Conference on Computer and Electrical Engineering, ICCEE*, 2009, vol. 2, pp. 357–361. doi: 10.1109/ICCEE.2009.168.

[61] N. Hutagaol and Suharjito, "Predictive Modelling of Student Dropout Using Ensemble Classifier Method in Higher Education," *Advances in Science, Technology and Engineering Systems*, vol. 4, no. 4, pp. 206–211, 2019, doi: 10.25046/aj040425.

[62] C. Y. Weston, "On the K-Nearest Neighbor Approach to the Generation of Fuzzy Rules for College Student Performance Prediction," Morgan State University, 2015.

[63] E. J. M. Lauría, E. Presutti, M. Kapogiannis, and A. Kamath, "Stacking Classifiers for Early Detection of Students at Risk," in *CSEDU 2018 - Proceedings of the 10th International Conference on Computer Supported Education*, 2018, pp. 390–397. doi: 10.5220/0006781203900397.

[64] R. Alkhasawneh and R. H. Hargraves, "Developing a Hybrid Model to Predict Student First Year Retention and Academic Success in STEM Disciplines using Neural Network," *Journal of STEM Education: Innovations and Research*, vol. 15, no. 3, pp. 35–42, 2014.

[65] A. I. Adekitan and E. Noma-Osaghae, "Data Mining Approach to Predicting the Performance of First Year Student in a University Using the Admission Requirements," *Education and Information Technologies*, vol. 24, pp. 1527–1543, 2019, doi: 10.1007/s10639-018-9839-7.

[66] A. C. K. Hoe *et al.*, "Analyzing Students Records to Identify Patterns of Students' Performance," *International Conference on Research and Innovation in Information Systems, ICRIIS*, pp. 544–547, 2013, doi: 10.1109/ICRIIS.2013.6716767.

[67] V. L. Miguéis, A. Freitas, P. J. V. Garcia, and A. Silva, "Early Segmentation of Students According to Their Academic Performance: A Predictive Modelling Approach," *Decision Support Systems*, vol. 115, no. September, pp. 36–51, 2018, doi: 10.1016/j.dss.2018.09.001.

[68] M. Shahhosseini, G. Hu, and H. Pham, "Optimizing Ensemble Weights and Hyperparameters of Machine Learning Models for Regression Problems," *arXiv:1908.05287*, 2019.

[69] 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, 2020, doi: 10.1007/s13202-020-00839-y.

[70] S. Ransdell, "Predicting college success: The importance of ability and non-cognitive variables," *International Journal of Educational Research*, vol. 35, no. 4, pp. 357–364, 2001, doi: 10.1016/S0883-0355(01)00032-5.

[71] L. Yan and Y. Liu, "An Ensemble Prediction Model for Potential Student Recommendation Using Machine Learning," *Symmetry*, vol. 12, no. 728, pp. 1–17, 2020.

[72] H. Hassan, S. Anuar, and N. B. Ahmad, "Students' Performance Prediction Model using Meta-classifier Approach," in *Engineering Applications of Neural Networks*, Springer International Publishing, 2019, pp. 221–231. doi: 10.1007/978-3-030-20257-6.

[73] M. Shahhosseini, G. Hu, and H. Pham, "Optimizing Ensemble Weights and Hyperparameters of Machine Learning Models for Regression Problems," *arXiv:1908.05287*, 2019.

[74] Cha Zhang and Yunqian Ma, Eds., *Ensemble Machine Learning: Methods and Applications*. Springer, Boston, MA, 2012.

[75] T. G. Dietterich, "Ensemble Methods in Machine Learning," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 1857 LNCS, pp. 1–15, 2000, doi: 10.1007/3-540-45014-9_1.

[76] R. Ghorbani and R. Ghousi, "Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques," *IEEE Access*, vol. 8, pp. 67899–67911, 2020, doi: 10.1109/ACCESS.2020.2986809.

[77] A. Tharwat, "Classification Assessment Methods," *Applied Computing and Informatics*, 2018, doi: 10.1016/j.aci.2018.08.003.

[78] E. K. Ampomah, Z. Qin, and G. Nyame, "Evaluation of Tree-Based Ensemble Machine Learning Models in Predicting Stock Price Direction of Movement," *Information (Switzerland)*, vol. 11, no. 332, 2020, doi: 10.3390/info11060332.

[79] L. A. Jeni, J. F. Cohn, and F. de La Torre, "Facing Imbalanced Data - Recommendations for the Use of Performance Metrics," in *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, 2013, pp. 245–251. doi: 10.1109/ACII.2013.47.