

# RHEM: A Robust Hybrid Ensemble Model for Students' Performance Assessment on Cloud Computing Course

Sapiah Sakri<sup>1</sup>, Ala Saleh Alluhaidan<sup>2</sup>

Department of Information Systems, College of Computer and Information Science  
Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia

**Abstract**—Creating tools, such as a prediction model to assist students in a traditional or virtual setting, is an essential activity in today's educational climate. The early stage towards incorporating these predictive models using techniques of machine learning focused on predicting the achievement of students in terms of the grades obtained. The research aim is to propose a robust hybrid ensemble model (RHEM) that can warn at-risks students (on Cloud Computing course) of their likely outcomes at the early semester assessment. We hybridised four renowned single algorithms – Naïve Bayes, Multilayer Perceptron, k-Nearest Neighbours, and Decision Table – with four well-established ensemble algorithms – Bagging, RandomSubSpace, MultiClassClassifier, and Rotation Forest – which produced 16 new hybrid ensemble classifier models. Hence, we have thoroughly and rigorously built, trained, and tested 24 models all together. The experiment concluded that the Rotation Forest + MultiLayer Perceptron model was the best performing model based on the model evaluation in terms of Accuracy (91.70%), Precision (86.1%), F-Score rate (87.3%), and Receiver Operating Characteristics Area detection (98.6%). Our research will help students identify their likely final grades in terms of whether they are excellent, very good, good, pass, or fail, and, thus, transform their academic conduct to achieve higher grades in the final exam accordingly.

**Keywords**—Academic performance; classification algorithms; cloud computing course; ensemble algorithms; hybrid ensemble classifier model; student academic performance tracking

## I. INTRODUCTION

Innumerable data are generated and gathered in numerous fields. The big data created need to be collected, organized, and analysed in order to extract useful information. In order to obtain valuable information, real-world environments and industries need to analyse vast quantities of generated data. To do so, Data Mining (DM) techniques are used to create a model that analyses the given dataset and identifies useful trends in the results. DM includes numerical data analysis techniques and the discovery of useful knowledge. One of the most required procedures in big data and data mining is prediction, which has been utilized in different domains to increase efficiency and reduce costs. This usage of algorithms in education is still in progress. To explain, the education provided at the university level is usually connected to the economy and development of a country. However, the quality and output of education at this level depends on the kind of students admitted and whether they are able to complete their

studies. The prediction of student academic performance helps in identifying weak students who will struggle with their studies. Science and IT majors are among the hardest at college level [1],[2]. Therefore, the management of computer and IT related institutions take essential steps to detect and correct the way for weak students. Many prediction and data mining algorithms have been used, such as clustering, classification, and association rule techniques, to extract knowledge from student datasets [3],[4]. This paper explores the effect of certain factors on student performance in advanced IT courses, such as cloud computing. Parameters, such as business course, maths course grade, science course grade, and core IT course grade can provide an indication of future students' performance in higher advanced courses. The current approaches have failed to analyse and monitor the progress of the student achievements [5]. Inappropriate methods or investigation procedures can also contribute to failure. This paper attempts to predict the educational performance of students based on motivational and academic factors. It introduces a hybrid prediction framework for measuring student performance in advanced computing courses, such as cloud infrastructure and services.

This paper is organized as follows. Section 2 explains the importance of this research. Section 3 presents the related work of the research. Section 4 elaborate the research approach. Section 5 describes the research preliminaries. Section 6 reports the experiment results. Section 7 discusses the findings in Section 6. Finally, Section 6 draws the conclusion and identifies future work.

## II. RESEARCH MOTIVATION

Predicting student performance at early stage of the semester would benefit both the university and students, particularly those students in their final year. For the university, high performing students would reflect the university's high quality of education system. For the students, knowing their level of performance at the early stage of the semester would avoid any decreased of grade in any courses which would have detrimental effect on their cumulative grade point average. Hence, we were motivated to propose a robust hybrid ensemble model (RHEM) that harnessed on machine learning algorithms which can predict at-risks students (on Cloud Computing course) of their likely outcomes at the early stage of the semester assessment. The prediction of exam performance through the context of student progress during

the course can increase efficiency and reduce the possibility of failing by offering pertinent advice and taking precautions. The marks obtained by a student in the examinations throughout the course duration can indicate the final exam results. Therefore, it becomes essential to predict whether the student will achieve an excellent, very good, good, pass or fail grade in the course. If the prediction indicates there is a high probability of the student failing the first exam prior to the final exam, then extra effort can be made to pass the exam.

### III. RELATED WORK

Machine Learning Algorithms are a group of useful tools that are used to create predictive models of student performance. Assessing the success prediction of the students is a very complex problem and uses different algorithms for this purpose. A systematic literature review to identify and collect the beneficial features for predicting student performance was discussed in [6], as well as the importance of Feature Selection (FS) to eliminate unrelated data that can produce a 10% difference in the prediction accuracy. Filter feature selection algorithms and classification algorithms were examined in [3]. The review shows that a variety of techniques have been used, but that there is no unified method that can be used for prediction in all cases. Specifically, the review uncovered a lack of quality and that there is a real need for more detailed reporting of the methods and results [6]. Referring to a closer topic to this research, student performance prediction by participating in an online discussion was mentioned in [2]. The sample was large as it comprised 76 second-year university students studying a Computer Hardware course. The study design was oriented to answer whether student performance prediction is possible and to compare different algorithms and features using classification and pre-processing techniques. The k-Nearest Neighbour algorithm accurately predicted unsuccessful students (89%). Moreover, students who were unsuccessful at the end of term could be predicted in the first 3 weeks with 74% accuracy.

The data collected by institutions or learning management systems are used in the sense of learning analytics to forecast student performance and recognize important factors that may contribute to the successful completion of a course. As we are interested in estimating student results, we review the related research work in this field. Although different techniques have been implemented in terms of prediction within the education field, it is still possible to improve the current approaches and provide more accurate results in terms of the context in which it is implemented. In the following, we review the current approaches that have been established. Educational Data Mining (EDM) and Learning Analytics (LA) to reveal knowledge from educational data were used to predict student success using data from the various universities in Pakistan [7]. "Learning analytics, discriminative and generative classification models are used to determine whether or not a student will complete his or her degree" [7]. Outcomes reveal better accuracy due to the reference of family expenditure, such as a natural gas, electricity, telephone, water, and accommodation, and students' personal data, such as gender, marital status, and employment, etc. To enhance engineering students' performance, a study by [1] identified the factors that

can affect student success in this tough major. The study focused on the use of J48 and REP Tree algorithms to elicit the type of relationship between social parameters and student performance and predicting students' performance in their third semester. Analysis revealed that parents' education influences student performance and that previous semester grades greatly indicate the performance in the third semester. This finding helps in the early prediction of weak students to take the necessary decisions for improving students' performance. In terms of the algorithms used, J48 was more accurate than the REP Tree algorithm. Similarly, within science colleges, the author in [8] examined how the linking of prior knowledge and attitude for first year undergraduate chemistry students can affect their chemistry exam performance. Statistics showed that there are significant differences between the mean scores of students who have prior knowledge in chemistry and those who have not. Analysing the correlation and regression showed that previous knowledge affects the success of examinations. Two predictive models were suggested based on the regression analysis. In a similar vein, the research of [9] focused on how proficiency in certain courses can give an insight into student performance in programming courses. This is an IT concentrated research. The results of courses, such as introductory to physics and maths, can indicate performance in programming courses. Methods, such as Artificial Neural Network data mining, were used for prediction. The findings showed that having a background knowledge of mathematics and physics is vital for proficiency in programming. In looking for the most influential factors in student academic performance prediction, the authors in [10] aimed to present a predictive model for computer science students' study duration based on grades in the first two semesters. Naïve Bayes, decision tree, and Support Vector Machine (SVM) were used. The findings showed no significant difference between Naïve Bayes and decision tree in terms of efficiency, while SVM had the lowest performance. The influencing factors were grades, general subjects' grades, gender, and major subjects' grades.

A new method for prediction using Multi-Input Multi-Output, which relies on the Multi Adaptive Neuro-Fuzzy Inference System with Representative Sets, was introduced in [11]. To explain, authors used both global and a local training set with random parameters in the former, and premise and consequent parameters in the latter. Once the parameters have been refined, for the testing set, Fuzzy k-Nearest Neighbour is used to find which group it belongs to. This MANFIS-S model is validated against ANFIS, MANFIS, OneR, and Random Tree and is found to be more accurate. A dataset was collected from VNU University of Science, Vietnam, and three educational datasets were taken from KDD Cup. Another attempt using the Fuzzy Probabilistic Neural Network was mentioned in [12]. The Probabilistic Neural Network is a 4-layer, feed-forward, which includes an algorithm used for classification and mapping. It is based on Bayes' decision strategy and non-parametric kernel-based estimators of probability density functions. The experiments and results revealed that FPNN takes less time to be trained and the results are more accurate (average of 98.56%). The output consists of a class of three values (Good, Average, and Poor).

MATLAB was used to analyse 760 samples of the training dataset with over 18 factors as inputs (merit, interest, family background, class and study behaviour, interest and belief in learning). Various techniques and features have been designed to predict academic success from the literature reviewed; however, there is still a shortage of work predicting the achievement of higher courses in computer education. Therefore, this study aims to fill the gap by focusing on an advanced level course (highly required from Computer major students as it is the trend and the new approach of hosting and managing databases in the job market). The hybrid algorithm was designed to produce more accurate results.

#### IV. METHODOLOGY

The proposed research approach for this study is as shown in Fig. 1. Four phases are involved and four major experiments will be conducted.

##### A. Phase 1 – Data Pre-Processing

In phase 1, raw data was pre-processed by performing normalization, replacing missing values, and transforming the raw data into a new clean dataset appropriate for the experiment’s requirements. The dataset was split 70% for training and 30% for testing [13][14]. These two sets of data were used to train the models in the two main experiments: without hybridization and with hybridization.

##### B. Phase 2 – Train Models without Hybridisation

Phase 2 involved two parts. The first part was the building, training, and testing of ensemble-based models by using Bagging (BAG), Random SubSpace (RNDS), MultiClass Classifier (MCC), and Rotation of Forest (ROF) Algorithms [15]. The second part was the building, training and testing of the base learner or classification-based models using Naïve Bayes (NB), MultiLayer Perceptron (MLP), k-Nearest Neighbour (KNN), and Decision Table (DT) algorithms. The test option for both parts was to use a training set with 10-fold cross-validation during training and using the supplied test set with 10-fold cross-validation during the model testing.

##### C. Phase 3 – Train Models with Hybridisation

Phase 3 involved the building, training, and testing of all the hybrid ensemble-based models by hybridising ensemble algorithms with classification algorithms as base learners [14][16]. The models were BAG+NB, BAG+MLP, BAG+KNN and BAG+DT. Followed by RNDS+NB, RNDS+MLP, RNDS+KNN, and RNDS+DT. Next were MCC+NB, MCC+MLP, MCC+KNN, and MCC+DT. The last hybrid ensemble-based models were ROF+NB, ROF+MLP, ROF+KNN, and ROF+DT. The test option for both parts was to use a training set with 10-fold cross-validation during training and using the supplied test set with 10-fold cross-validation during the model testing.

##### D. Phase 4 – Perform Comparison Analysis

Phase 4 involved the comparison analysis of the performance metrics for all the models trained in phase 2 and phase 3. The metrics were in terms of accuracy, precision, recall, F-measure, and ROC area [14][17][18]. The models were the ensemble-based models, classification-based models, and hybrid ensemble-based models.

#### V. RESEARCH PRELIMINARIES

##### A. Dataset Descriptions

Real data were collected based on more rational attributes that were suggested by the previously conducted relevant research. An online questionnaire, generated using Google forms, was circulated on social media to different groups targeting university students taking the Cloud Computing course to gather the necessary data. A total of 319 students filled out the questionnaire, which was considered an appropriate dataset size to be used in building and training single classifier-based models, ensemble classifier-based models, and classifier-based hybrid ensemble models. The questionnaire was designed to include students’ demographic and students’ motivational behaviour questions for the course cloud computing. The independent variables can be easily transformed to dependent variables or attributes that may predict the class of final examination results (Excellent, Very Good, Good, Past, Fail). The list of collected attributes is illustrated in Table I.

##### B. Performance Metric Descriptions

1) *Multi class confusion matrix evaluation:* The prediction model’s method of evaluating fitness was by analysing the confusion matrix. The confusion matrix, as shown in Table II, contained information about the proposed classifier’s actual and predicted classification. With the aid of the Academician expert, the proposed model was verified to check the prediction model’s accuracy.

2) *Accuracy detailed evaluation:* The performance metrics that we apply to assess the proposed model’s performance were in terms of classification accuracy, recall, precision, F-measure, and ROC area [19]. Table III shows the classification measures representations.

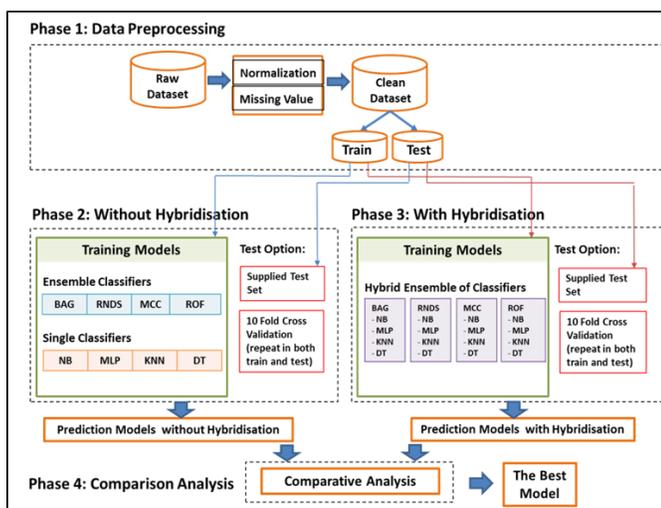


Fig. 1. Proposed Research Approach.

TABLE I. DATASET DESCRIPTION

Attribute	Description	Possible Value
<b>Demographic Information</b>		
Age	Student's Age	18-20, 21-24, > 25
FSize	Student's Family Size	Less or equal than 3, more than 3
HStatus	Student's health status	Healthy, Not Healthy
MStatus	Student's marital status	Single, Married, Other
WStatus	Student's work status	Full-time job, part-time job, online job
Family Highest Education Level	Student's family highest education level	Tertiary, Secondary, Primary
<b>Motivational Behaviour Information</b>		
Level of Self Confidence	Confidence towards this course	High, Medium, Low
Level of interest	Interest towards this course	High, Medium, Low
Level of positive thinking	Positive thinking towards this course	High, Medium, Low
Level of focus in class	Focus in the class	High, Medium, Low
Have Personal Goal	Personal Goal towards this course	Yes, No
<b>Academic Information</b>		
Major Program	Name of program	CS, IS, IT
No. of Absence	Hours of absence in the class	1-2, 3-4, > 4
Current GPA	Current GPA	1.0 – 1.9, 2.0 – 2.9, 3.0 – 3.9, 4.0 – 5.0
Grade acquire in First Exam	Exam 1 grade (20%)	Excellent, V. Good, Good, Pass, Fail
Data Class	Categories	Excellent, V. Good, Good, Pass, Fail

TABLE II. MULTI-CLASS CONFUSION MATRIX EVALUATION

Actual		Predicted				
		Very Good	Excellent	Fail	Pass	Good
		a	b	c	d	e
Very Good	a	<b>TP</b>	Error <sub>ab</sub>	Error <sub>ac</sub>	Error <sub>ad</sub>	Error <sub>ae</sub>
Excellent	b	Error <sub>ba</sub>	<b>TP</b>	Error <sub>bc</sub>	Error <sub>bd</sub>	Error <sub>be</sub>
Fail	c	Error <sub>ca</sub>	Error <sub>cb</sub>	<b>TP</b>	Error <sub>cd</sub>	Error <sub>ce</sub>
Pass	d	Error <sub>da</sub>	Error <sub>db</sub>	Error <sub>dc</sub>	<b>TP</b>	Error <sub>de</sub>
Good	e	Error <sub>ea</sub>	Error <sub>eb</sub>	Error <sub>ec</sub>	Error <sub>ed</sub>	<b>TP</b>

TP – True Positive E- Error

TABLE III. ACCURACY DETAILED EVALUATION

Accuracy Detailed	Explanation
Classification Accuracy	Accuracy measures the proportion of instances that are a correctly classified. Formula: (TP + TN)/(TP + FP + FN + TN)
Sensitivity (Recall)	Sensitivity is the proportion of positive factors that the classifier correctly identifies as positive. Formula: TP/(TP + FN)
Precision	This is a calculation of the relevant collected instances Formula: TP/(TP + FP)
F-Score / F-Measure	The F1 Score is needed to seek a balance between Precision and Recall. Formula: F1 = 2 (Precision * Recall)/(Precision + Recall)
ROC Area	The AUC-ROC curve is a classification problem quality calculation at various threshold settings. ROC is the curve of probability and AUC is the degree metric. It indicates how much a model between classes can be differentiated. The higher the AUC, the better the model is at predicting 0s as 0s and 1s as 1s. Formula: FP/(FP + TN).

### C. Single Algorithm Descriptions

1) *Naïve Bayes Algorithm (NB)*: NB algorithm is a supervised method of learning as well as a quantitative approach of classification proposed by Thomas Bayes [21]. This algorithm assumes a proactive inhibition model, which results in doubt about the system by specifying the likelihood of outcomes in theory. Diagnostic and predictive problems can be addressed. The Bayesian classification provides realistic algorithms for learning and prior knowledge incorporating observed data.

2) *Multilayer Perceptron (MLP)*: The feed-in class to the artificial neural network is an MLP. The MLP contains at least three layers of nodes: an input layer, a hidden layer, and an output layer [21]. The node, excluding the input nodes, is a neuron that uses a nonlinear activation function. MLP uses the guided learning method for instruction, called back propagation. MLP differentiates from a linear perceptron by its multiple layers and non-linear activation. Data that cannot be separated linearly can be differentiated.

3) *k-Nearest Neighbour Algorithm (kNN)*: kNN is an algorithm of gradation that is widely used in the identification of statistical patterns [21]. Every class has a few sample models and a set of pattern vectors. When a vector must be named, it will be one of the model vectors that is its nearest k neighbour. The majority rule is the tag category. To prevent relations to class and overlap areas, the value of k should be odd. This rule is sophisticated but plain, and, in practice, has a low error rate.

4) *Decision Table Algorithm (DT)*: DT are classification models used in forecasting, such as Decision Trees and neural networks [22]. They are induced by algorithms for machine learning. A table of decisions includes a hierarchy of the values in which each entry in a higher table is broken down to form a table by the values of a pair of additional attributes.

#### D. Ensemble Algorithms Description

1) *Bagging Algorithm (BAG)*: An Ensemble Meta-Stimulator for BAG is an ensemble that fits the base classifiers into a random subset of the original dataset and then aggregates its individual forecasts (by vote or by means of an average of) into a final prediction [22]. Normally, such a meta-estimator can be used as a means of reducing the variance of a blackbox estimator (e.g. the decision tree), randomizing its design process, and then creating an ensemble from it.

2) *Random Subspace Algorithm (RNDS)*: The base classifier model is based on a set constructed from the initial set of functionalities using the RNDS approach proposed by Ho [23]. Through a simple majority vote procedure, the outcomes of the individual graders are merged in a final decision.

3) *Multiclass Classifier Algorithm (MCC)*: MCC is a metaclassifier with 2-class classifiers for managing multi-class datasets [22]. This can also add error to correct a metaclassifier output code in order to improve accuracy.

4) *Rotation Forest Algorithm (ROF)*: ROF is a way to produce classifier assemblies based on the extraction of features [22]. The feature set is randomly divided into K (K is an algorithm parameter), and Principal Component Analysis (PCA) is applied to each subset to create training data for a base categorizer. In order to preserve variation information in the results, all the principal components are retained. Therefore, the K axis rotation forms the new features for a simple classification system.

### VI. EXPERIMENTAL RESULTS

In this study, four main experiments were conducted sequentially with the aim to assess the students' performance using cloud computing course dataset by training various single, ensemble and hybrid ensemble classifiers. Followed by conducting comparative analysis to detect any performance improvement in all the different types of models. These experiments eventually identify the best performing model in predicting student's performance on cloud computing course [20].

#### A. Experiment 1: Training Models without Hybridisation

The aim of this experiment is to observe the effect of the four ensemble classifiers and the four single classifiers without the process of hybridisation between the two classifiers type. In total, eight models were evaluated in this experiment. Fig. 2 shows the results of the evaluation which indicate that each classifiers have achieved their highest performance for different metrics. ROF model achieved the highest accuracy value at 90.90% and also the highest ROC metric value at 98.10%. MCC model obtained the highest precision value at 83.8% and also the highest F-score value at 86.10%. Whereas, RNDS model achieved the highest recall value at 94.30%.

Fig. 3 shows the experiment results for the single-based model evaluation. It shows that MLP out-performed the rest of the models by obtaining the highest accuracy value at 90.50%,

the highest precision value at 81.4%. The highest F-score value at 84.90% and the highest ROC value at 97.60%. However, in terms of the recall metric, NB and DT models achieved the highest value at 91.40%.

#### B. Experiment 2: Training Models with Hybridisation

The aim of this experiment is to hybrid the ensemble classifiers with the single classifiers as the base learners. In this experiment, we thoroughly evaluated 16 hybrid ensemble models. The results were shown in Fig. 4, Fig. 5, Fig. 6 and Fig. 7.

Fig. 4 shows the evaluation results of the hybrid BAG-based models which indicated that BAG+MLP model achieved the highest performance in all the evaluation metrics. This model obtained the highest accuracy metric (89.30%), the highest precision metric (76.20%), the highest F-score metric (83.90%) and the highest ROC metric (98.30%). However in terms of the recall metric, this model shared the highest value with BAG+NB, and BAG+DT models at 91.40%.

Fig. 5 shows the experiment results for the hybrid RNDS-based models' evaluation. The results shows that, RNDS+KNN model obtained the highest accuracy value at 87.60%, the highest precision value at 76.20% and the highest ROC value at 98.60%. However, in terms of recall metric, RNDS+DT model achieved the highest value at 100% and RNDS+MLP model demonstrated the highest F-Score value at 78.50%.

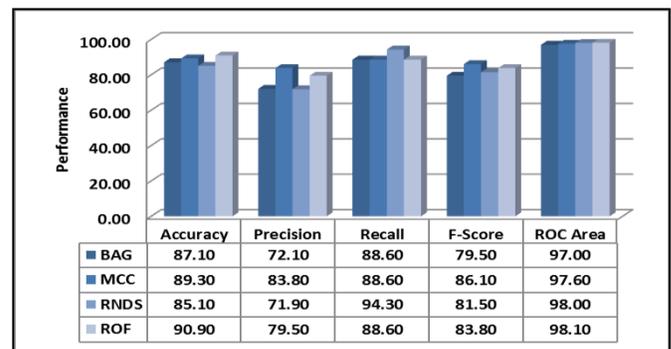


Fig. 2. Ensemble-Based Models Evaluation.

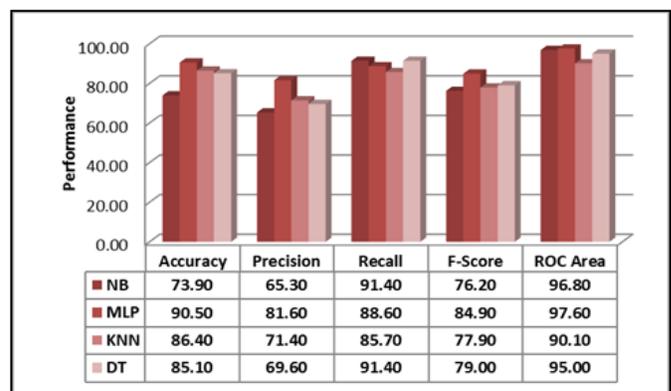


Fig. 3. Single-Based Models Evaluation.

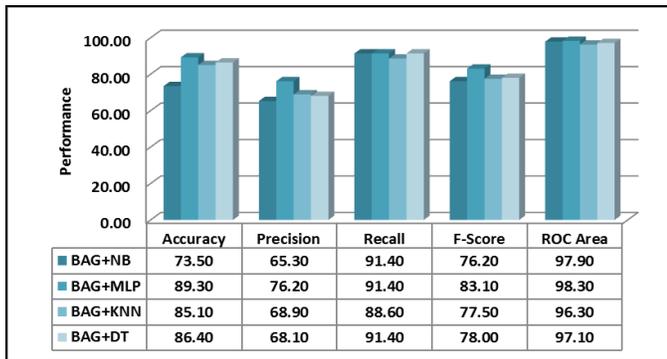


Fig. 4. Hybrid BAG-Based Models Evaluation.

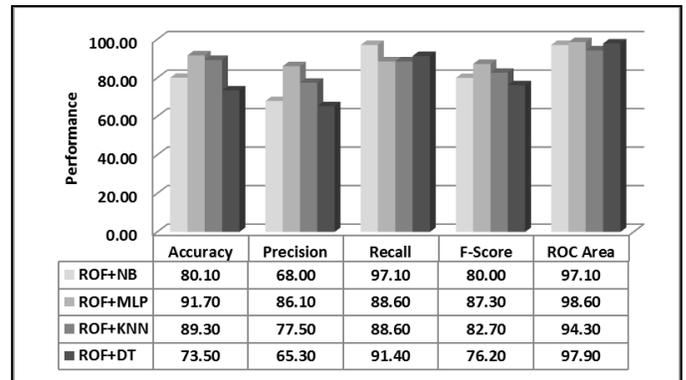


Fig. 7. Hybrid ROF-Based Models Evaluation.

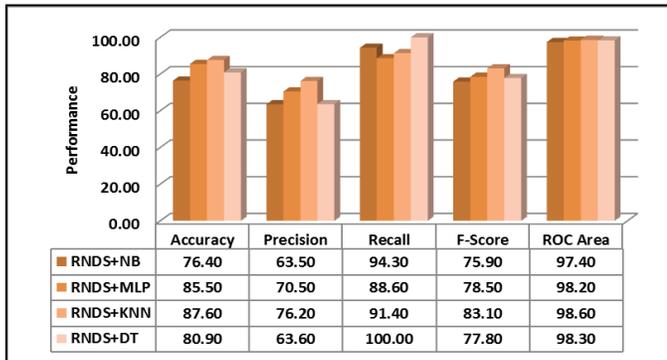


Fig. 5. Hybrid RNDS-Based Models Evaluation.

Fig. 6 presents the results for the hybrid MCC-based models' evaluation. The results clearly demonstrated that MCC+MLP model out-performed the rest in terms of the accuracy metric (90.90%), the precision metric (82.10%), the F-score metric (86.50%) and the ROC metric (98.20%). However, in terms of the recall metric, MCC+NB model achieved the highest value at 94.30%.

Fig. 7 shows the experiment results for the hybrid ROF-based models' evaluation which highlight the best performance of ROF+MLP model in terms of accuracy (91.70%), precision (86.10%) and ROC (98.60%). However, in terms of recall and F-score metric, ROF+NB model achieved the highest value at 97.10% and 88.00%, respectively.

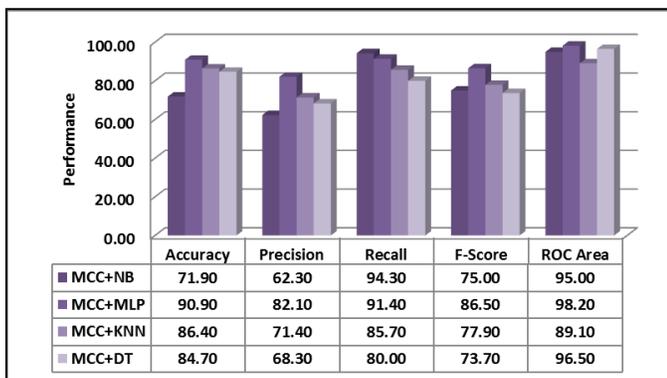


Fig. 6. Hybrid MCC-Based Models Evaluation.

### C. Experiment 3: Comparative Analysis

In this analysis, our aim is to observe any performance improvement in the single-based, ensemble-based and hybrid ensemble-based models by comparing the evaluation results of the hybrid and the non-hybrid models. The first experiment is to compare between the non-hybrid models. The second experiment is to compare between the hybrid models. The third experiment is to evaluate the confusion matrix and AUC ROC that corresponds to the best-fitted model.

1) *Experiment 3-1: Comparative Analysis of the Non-Hybrid Models:* The aim of this experiment is to observe the performing achievement between the non-hybrid models. The experiment results, as shown in Fig. 8 demonstrated that ensemble-classifiers out-performed the single classifiers-based model. ROF-based models perform better in accuracy metric (90.9%) and in ROC metric (98.10%). Whereas, MCC-based model achieved the highest precision value at 83.8%. However for the recall metric RNDS-based model achieved the highest value at 94.3%. Nevertheless, MLP-based model which represent the single-classifier, achieved the highest F-score value at 84.9%.

2) *Experiment 3-2: Comparative Analysis of the Hybrid Models:* The aim of this experiment is to identify the best performing hybrid model by evaluating and comparing the hybrid models' performance accordingly. Due to the complexity of the experiments, the results representation were divided into three parts as shown in Fig. 9, Fig. 10 and Fig. 11. In Fig. 9, it was observed that ROF+MLP performed better in the accuracy metric (91.7%) and in the precision metric (86.10%). In other words, this model can predict student performance for the excellent class with 91.7% accuracy as compared to the rest of the hybrid models. The result in terms of the precision metric can be interpreted as the model's ability to precisely predict that 86.1% of the data were relevant to the 'excellent', 'very good', 'good', 'pass', and 'fail' class. The results clearly indicate that the hybrid ensemble-based model improves the accuracy and precision of the prediction model.

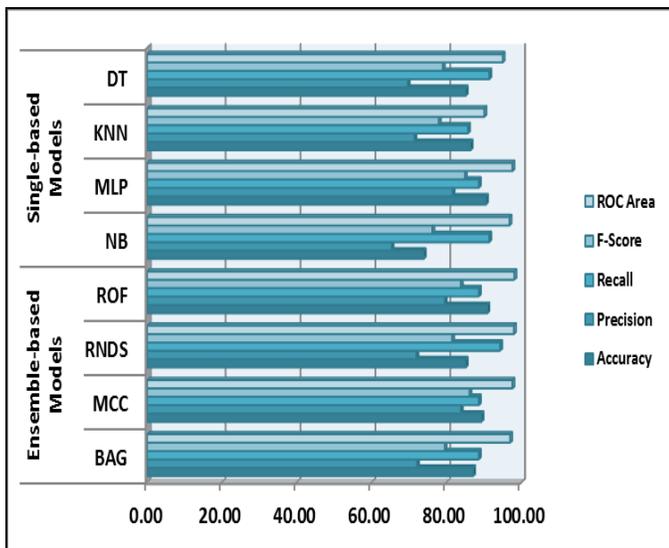


Fig. 8. Comparison between Non-Hybrid Models.

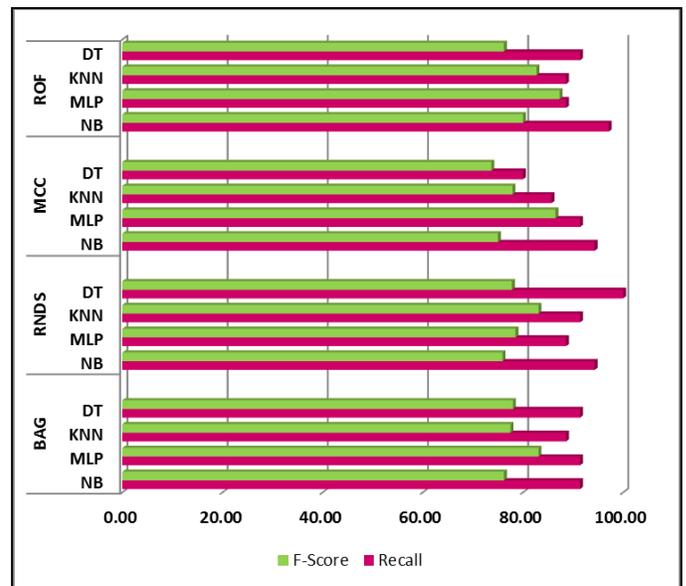


Fig. 10. Comparison of F-Score and Recall.

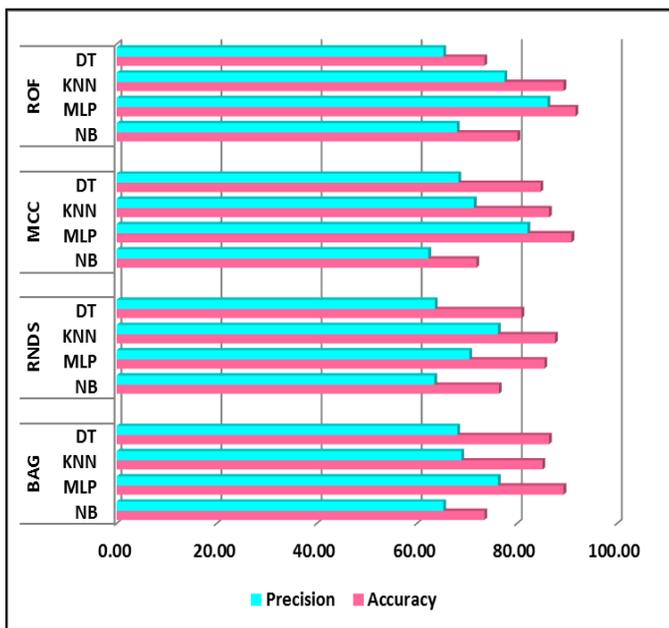


Fig. 9. Comparison of Precision and Accuracy.

Fig. 10 observed the comparison analysis between the hybrid ensemble-based models in terms of recall and f-score metric. The results indicate that RNDS+DT model achieved the highest recall value of 100%. Whereas, ROF+MLP model achieved the highest F-score value at 87.3%.

The following experiment is to evaluate and compare in terms of ROC area metric as shown in Fig. 11. The results demonstrated that ROF+MLP and RNDS+KNN model have the highest ROC value of 98.6%. In other words, by analogy, the higher the ROC, the better the model is at distinguishing between students' grades which were classified as 'excellent', 'very good', 'good', 'pass', or 'fail'.

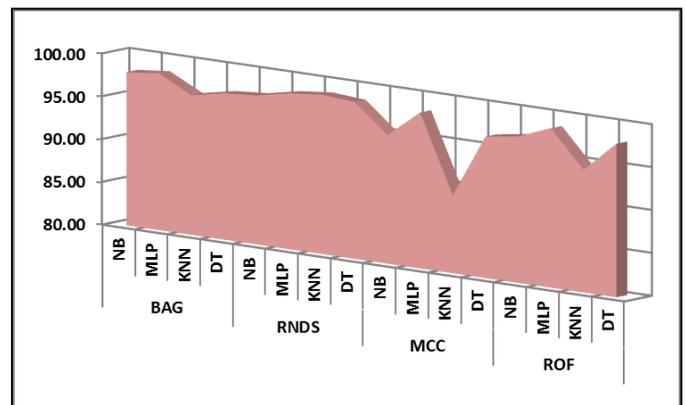


Fig. 11. Comparison of the AUC ROC Area.

## VII. DISCUSSION AND ANALYSIS

After comparing all the models, we have sufficient evidence to show that the hybridised ensemble models outperformed the non-hybridised ensemble-based models and also the single-based models. Table IV shows the summary of the comparative analysis.

### A. Experiment 4: The Confusion Matrix of ROF+MLP

Based on Table IV, there is clear evidence that ROF+MLP model is the best fitted model for predicting student academic performance in the cloud computing course. Thus, this experiment aim to prove that ROF+MLP as the best performing model by using the confusion matrix and observing the ROC area for all the classes. Confusion Matrix results as shown in Table V, confirms the above findings.

1) *Confusion Matrix for ROF+MLP Model:* Table V shows the confusion matrix for the ROF+MLP model. The results indicate that the model can correctly predicts 31 students or 93% are 'Excellent' students. The model also correctly predicts the rest of the class as follows: 30 students or

81% are ‘Very Good’ students, 78 students or 95% are ‘Good’ students, 27 students or 96% are ‘Good’ students, and 54 students or 95% are ‘Fail’ students. The confusion matrix indicates that the ROF+MLP model has excellent ability to correctly predict student performance with less than a 19% error.

TABLE IV. THE SUMMARY OF THE COMPARATIVE ANALYSIS

Performance Metrics	Single-based Model	Ensemble-based Model	Hybrid Ensemble-based Model
Classification Accuracy	MLP = 90.5%	ROF=90.9%	ROF+MLP = 91.7%
Precision	MLP = 81.6%	MCC=83.8%	ROF+MLP = 86.1%
Recall	NB = 91.4% DT = 91.4%	RNDS=94.3%	RNDS+DT = 100%
F-Score	MLP = 84.9%	MCC=86.1%	ROF+MLP = 87.3%
ROC (AUC)	MLP = 97.6%	ROF=98.1%	ROF+MLP = 98.6% RNDS+KNN = 98.6%

TABLE V. CONFUSION MATRIX FOR ROF+MLP MODEL

Actual		Predicted				
		V. Good	Excellent	Fail	Pass	Good
		a	b	c	d	e
Very Good	a	<b>30</b>	3	0	0	1
Excellent	b	3	<b>31</b>	0	0	1
Fail	c	0	0	<b>54</b>	2	0
Pass	d	2	1	2	<b>27</b>	2
Good	e	2	1	1	1	<b>78</b>

2) *ROC for Each Class in ROF+MLP Model:* The aim of this experiment is to observe the performance of ROF+MLP model in distinguishing the value between classes in the model. Fig. 12 shows the experiment results with regards to the Area under ROC or the threshold curve for the individual classes in the ROF+MLP model. The results indicate that all the classes have a high value of ‘Area Under ROC’. In other words, ROF+MLP model is good at distinguishing between class = “Excellent”, class = “Very Good”, class = “Good”, class = “Pass”, and class = “Fail” students. The highest value is obtained by the class = “Fail” with 99.9%. Followed by class = “Good” with 99.6%. The third place is class = “Excellence” with 97.95%. The fourth place is class = “Very Good” with value of 97.56%. The lowest value under ROC is class = “Pass” with value 86.7%.

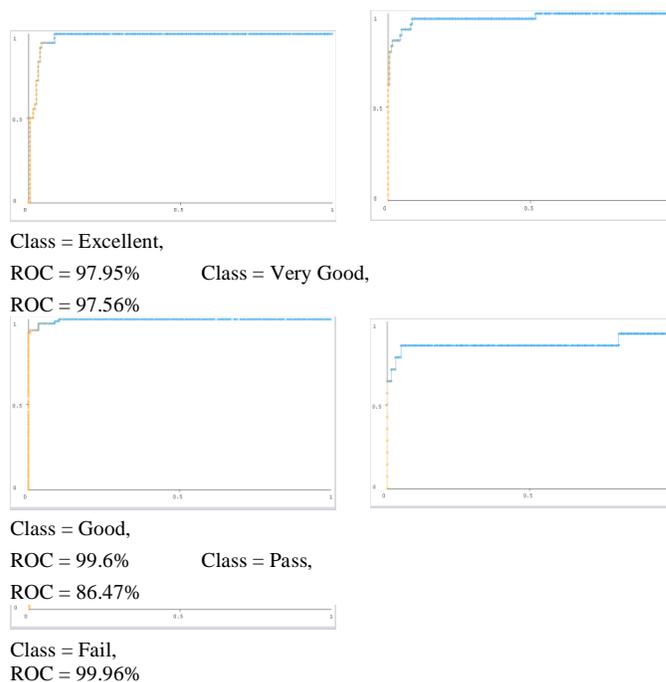


Fig. 12. AUC ROC for Each Classes in ROF+MLP Model.

### VIII. CONCLUSION AND FUTURE WORK

Cloud computing is considered to be a very tough course for most students. Hence early warning of the assessment outcome would be beneficial to at-risk students who have problems in sustaining their grades in that course through-out the whole semester. A robust hybrid ensemble model (RHEM) is highly useful in the prediction of assessment course outcome, assisting the students in deciding to continue or to drop the course at early semester. Based on the summary of the comparative analysis depicted in Table IV, it clearly demonstrated that the hybrid ensemble classifiers were able to improvise the ensemble and also the single classifiers. After many iterations of thorough and rigorous training that were carried out using all 24 models, the analysis indicated that Rotation Forest ensemble classifier hybrid with Multilayer Perceptron classifier as the base learner (ROF+MLP), appears to be the best robust hybrid ensemble model or RHEM that out-performed the rest of the models to predict students’ performance in cloud computing course at early stage of the semester.

A logical extension of this work would be the creation of a meta-analysis system for future study, which can be regarded as a decision support method based on the model that will achieve the highest efficiency and effectiveness.

#### ACKNOWLEDGMENT

This research was funded by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University through the Fast-track Research Funding Program.

#### REFERENCES

- [1] M. Imran, S. Latif, D. Mehmood, and M. S. Shah, "Student academic performance prediction using supervised learning techniques," *Int. J. Emerg. Technol. Learn.*, vol. 14, no. 14, p. 92, 2019.
- [2] G. Akçapınar, A. Altun, and P. Aşkar, "Using learning analytics to develop early-warning system for at-risk students," *Int. J. Educ. Technol. High. Educ.*, vol. 16, no. 1, 2019.
- [3] M. Zaffar, M. Ahmed, K. S. Savita, and S. Sajjad, "A study of feature selection algorithms for predicting students academic performance," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 5, 2018.
- [4] P. Nuankaew, "Dropout Situation of Business Computer Students, University of Phayao," *Int. J. Emerg. Technol. Learn.*, vol. 14, no. 19, p. 115, 2019.
- [5] A. J. P. Delima, Ariel, and Ruji, "Variable reduction-based prediction through modified genetic algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, 2019.
- [6] A. Hellas, P. Ihanola, A. Petersen, V.V. Ajanovski, M. Gutica, T. Hynninen, A. Knutas, J. Leinonen, C. Messom, and S.N. Liao, "Predicting academic performance: a systematic literature review," in *Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education - ITiCSE 2018 Companion*, 2018.
- [7] A. Daud, N. R. Aljohani, R. A. Abbasi, M. D. Lytras, F. Abbas, and J. S. Alowibdi, "Predicting student performance using advanced learning analytics," in *Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion*, 2017.
- [8] J. Ross, E. Guerra, and S. Gonzalez-Ramos, "Linking a hierarchy of attitude effect to student engagement and chemistry achievement," *Chem. Educ. Res. Pr.*, vol. 21, no. 1, pp. 357–370, 2020.
- [9] N. Sael, T. Hamim, and F. Benabbou, "Implementation of the analytic hierarchy process for student profile analysis," *Int. J. Emerg. Technol. Learn.*, vol. 14, no. 15, p. 78, 2019.
- [10] E. Abou Gamie, S. Abou El-Seoud, and M. A. Salama, "Comparative analysis for boosting classifiers in the context of higher education," *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 10, p. 16, 2020.
- [11] L. H. Son and H. Fujita, "Neural-fuzzy with representative sets for prediction of student performance," *Appl. Intell.*, vol. 49, no. 1, pp. 172–187, 2019.
- [12] X. Zhang, R. Xue, B. Liu, W. Lu, and Y. Zhang, "Grade prediction of student academic performance with multiple classification models," in *2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 2018.
- [13] O. W. Adejo and T. Connolly, "Predicting student academic performance using multi-model heterogeneous ensemble approach," *J. appl. res. high. educ.*, vol. 10, no. 1, pp. 61–75, 2018.
- [14] B. K. Francis and S. S. Babu, "Predicting academic performance of students using a hybrid data mining approach," *J. Med. Syst.*, vol. 43, no. 6, p. 162, 2019.
- [15] A. Alhassan, B. Zafar, and A. Mueen, "Predict Students' Academic Performance based on their Assessment Grades and Online Activity Data," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 4, 2020.
- [16] P. Sökkhey and T. Okazaki, "Hybrid machine learning algorithms for predicting academic performance," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 1, 2020.
- [17] A. Almasri, E. Celebi, and R. S. Alkhaldeh, "EMT: Ensemble meta-based tree model for predicting student performance," *Sci. Program.*, vol. 2019, pp. 1–13, 2019.
- [18] R. Hasan, S. Palaniappan, S. Mahmood, K. U. Sarker, and A. Abbas, "Modelling and predicting student's academic performance using classification data mining techniques," *Int. J. Bus. Inf. Syst.*, vol. 34, no. 3, p. 403, 2020.
- [19] X. Xu, J. Wang, H. Peng, and R. Wu, "Prediction of academic performance associated with internet usage behaviors using machine learning algorithms," *Comput. Human Behav.*, vol. 98, pp. 166–173, 2019.
- [20] 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," *Educ. Inf. Technol.*, vol. 24, no. 2, pp. 1527–1543, 2019.
- [21] M. Injadat, A. Moubayed, A. B. Nassif, and A. Shami, "Systematic ensemble model selection approach for educational data mining," *Knowl. Based Syst.*, vol. 200, no. 105992, p. 105992, 2020.
- [22] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 4, p. e1249, 2018.
- [23] D. Ndirangu, W. Mwangi, and L. Nderu, "Improving multiclass classification and outlier detection method through ensemble technique," in *Proceedings of the 4th International Conference on Communication and Information Processing - ICCIP '18*, 2018.