

Machine learning's model-agnostic interpretability on The Prediction of Students' Academic Performance in Video-Conference-Assisted Online Learning During the Covid-19 Pandemic

Eka Miranda

ekamiranda@binus.ac.id

Information Systems Department, School of Information Systems Bina Nusantara University Jakarta, Indonesia

Mediana Aryuni

Information Systems Department, School of Information Systems Bina Nusantara University Jakarta, Indonesia

Mia Ika Rahmawati

Sekolah Tinggi Ilmu Ekonomi Indonesia (STIESIA) Surabaya

Siti Elda Hiererra

Information Systems Department, School of Information Systems Bina Nusantara University Jakarta, Indonesia

Dian Sano

Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Research Article

Keywords: Students' academic performance, Video-conference-assisted online learning, (VCAOL), Machine learning, Model-agnostic

Posted Date: October 13th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3426498/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

The Covid-19 pandemic had an immediate impact on higher education. Although online technology has made contributions to higher education, its adoption has had a significant impact on learning activities during the Covid-19 pandemic. This paper proposed a predictive model for predicting students' academic performance in video-conference-assisted online learning (VCAOL) during Covid-19 pandemic based on machine learning approach. We investigated: Random Forest (RF), Support Vector Machine (SVM) and Gaussian Naive Bayes (GNB). There were 361 data gathered as a dataset from September 2022 to January 2023. The overall result revealed RF outperformed SVM and GNB with accuracy score of 60.27%, precision 59.46%, recall 60.27%, F1-score 59.51% and ROC AUC 87%. Understanding a machine learning model's black-box output was crucial for providing predictions that explain why and how they were developed. SHAP value of global interpretability to visualize global feature importance revealed that students' performance while using VCAOL (Performance) was the most critical attribute for predicting students' academic performance. The SHAP local interpretability bar plot revealed that 'student academic performance was still well achieved during the learning process, despite using video conferencing during the Covid-19 pandemic' (Performance), when Performance decreased it contributed negative impact on students' academic performance in VCAOL during Covid-19.

1. Introduction

Learning is a process of interaction between educators and students in a learning environment, aided by learning materials or resources. Educator aid and lead learners in gaining knowledge. Learning can take place in small or large groups. During this phase, students develop not just intellectually, but also psychomotor [1]. Face-to-face learning is used in official schools in Indonesia at the elementary, junior high, high school, and college levels. However, around three years ago, in 2019, a factor emerged that resulted in a learning change. This factor was the spread of the Covid-19 virus. Because of the rapid spread, the government took serious measures on March 16, 2020, by issuing a remote learning policy. All school activities were halted to limit the spread of the infection [2]. Significant changes in the global learning system, including in Indonesia, were caused by this policy. Learning must take place online, and educational institutions must adapt to this change. Online learning was being implemented in Indonesia at the same time starting from kindergarten, elementary school, junior high school, high school, and university [2]. It is necessary for students and teachers to adapt to the changes that occur. Lecturers must promptly adjust the system, syllabus, and learning process. Students, on the other hand, stammered because they had a pile of homework to finish while studying at home. In fact, online learning does not replace face-to-face instruction with digital applications, nor does it load students with a mountain of homework every day. Students should be encouraged to be creative in accessing as many sources of knowledge as possible, developing works, and honing insights through online learning [2].

One of the technologies used to support online learning is video conferencing. Video conferencing is a technology that enables two or more participants to engage simultaneously through video and voice. Video conference-assisted online learning (VCAOL) facilitates a wide range of learning opportunities, but

its implementation is fraught with challenges. Lecturer and students must supply enough equipment, such as PCs or laptops, as well as a stable internet network, to guarantee that the entire teaching and learning process runs properly. Lecturers must employ ways to maintain interaction with students during studying and to determine whether students understand the content being taught. Knowing students' academic performance while online learning and assisted by video conference platforms is necessary so that lecturers and educational institutions could make future changes [2].

Previous studies have demonstrated that the Covid-19 pandemic had an immediate impact on higher education. Although online technology has made contributions to higher education, its adoption has had a significant impact on learning activities during the Covid-19 pandemic. Research by [3] revealed students experienced negative emotional consequences during a video conference-assisted remote learning. The research looked at the following indicators: attitudes toward video conferencing, emotions, both positive and negative, and student evaluations measured (student and course learning outcomes, evaluation of instructor's teaching methods and practices). The PANAS-X questionnaire was utilized in this study, which analyses two general aspects, positive and negative effects, and both were then statistically examined. Research by [4] investigated the facilitating conditions and the students' perceptions toward using these interactive resources and video conferences to continue their remote learning journey during Covid-19. A partial least square (PLS) approach revealed the result. This research explored the link between UTAUT/UTAUT2's performance expectancy (that is synonymous with TAM's perceived usefulness) with attitudes toward technology. Another research by [5] employed the technology acceptance model (TAM) to explain how the usage of video conference technology improves learning outcomes resulting from involvement with the zoom video conference platform. The associations between the TAM variables were examined using structural equation modelling. The following factors were examined: ease of use with video conferencing (VC), usefulness with VC, attitude with VC, intention to use VC, and actual use with VC. Most of previous studies have focused on students' perceptions of video conference platforms during their learning journey, with little attention given to examining their academic performance.

In recent years, there has been a lot of interest in the use of machine learning (ML) in education. It is the field of revealing new and meaningful results from data. ML in education has evolved into a powerful tool for uncovering hidden patterns in educational data and forecasting students' academic performance [6, 7]. Previous researchers [8] developed machine learning approaches to predict the final test grades of undergraduate students. To predict students' final exam grades, random forests, nearest neighbor, support vector machines, logistic regression, Nave Bayes, and k-nearest neighbor algorithms were examined and evaluated. Another previous research [9] reported on the use of ML algorithms to predict student performance by taking into consideration both previous semester grades and socioeconomic characteristics. In this case, the K-nearest neighbors (KNN), Radom Forest (RF), and Support Vector Machines (SVM) methods were tested. Research by [10] has already demonstrated useful early prediction of student improvement using machine learning and eXplainable AI. Logistic Regression (LR), Decision Tree (DT), RF, Multi-Layer Perceptron, SVM, and XGBoost were the machine learning algorithms employed

in the research. Only twelve people, including educators, parents of K-12 students, and officials, took part in the focus group interview.

Most of the past research has focused on the performance of ML models or the significance of features, with little emphasis on completely comprehending and explaining predictions using interpretable methodologies. Understanding the black-box output of a ML model was critical for generating forecasts that explain why and how they were generated. Model-independent ML approaches have been developed to discover and evaluate informative features. A model-agnostic interpretation approach, such as the SHapley Additive exPlanations (SHAP) framework, takes as inputs a dataset and several prediction models, applies these models to the data, and then determines the characteristics of data features inside each prediction model. [11].

The objectives of this study were as follows: (1) constructing a predictive model for the prediction of students' academic performance in VCAOL during Covid-19 pandemic using a machine learning (ML) approach. Prediction features covered 4 perspectives namely video conference application (VC), internet connection (IC), students' ability to learn (SL) and student knowledge (SK); (2) constructing model-agnostic ML algorithms for discovering and understanding predictive characteristics. To address the research objectives, there were two research questions (RQ): RQ1: Using a ML technique, how to create a predictive model for predicting students' academic performance in VCAOL during the Covid-19 pandemic? RQ2: how to build model-agnostic ML to interpret classification model for comprehending predictive properties. To answer RQ1 we constructed a predictive model for the prediction of students' academic performance in VCAOL, and we assessed the effectiveness of RF, SVM, and Gaussian Naive Bayes (GNB) models. We investigated RF, SVM, and GNB as these algorithms have previously demonstrated potential in predicting models for students' academic performance [6–10]. Subsequently, we employed SHAP framework to answer RQ2. SHAP enable to compute and study the influence of features on individual and overall predictions, as well as evaluate informative features and investigate their interpretability and characteristics.

Few studies have attempted to address the concerns mentioned in our paper. We expect that our methodology will give evidence and pave the way for future research on computational agents and machine learning for detecting, predicting, and evaluating prediction models for student academic achievement in VCAOL.

2. Related Work

In this research, we conducted a literature review of related works. Many researchers have contributed to the prediction of students' academic performance using various ML methods. A variety of ML techniques used to the prediction of students' academic performance prediction are outlined in Table 1.

Table 1
Related work

Study	ML technique, results, gap
Khurram Jawad [12]	RF, XGBoost and Logistic Regression (LR). Accuracy for testing data was 84.2%, 84.3% and 80.9% respectively. Dataset: Open University Learning Analytics
	Gap: More data on student involvement and evaluation scores allow for more accurate forecasts.
Thao-Trang Huynh-Cam [13]	RF, C50, CART, and multilayer perceptron (MLP) algorithms. CART outperforms C50, RF, and MLP algorithms with accuracy 80%. Dataset: school register system and school grading system the first semester of the academic year 2020–2021 at one technical and vocational university in Taiwan.
	Gap: focus on prediction accuracy value only.
Silvia Gaftandzhieva [14]	Random Forest, XGBoost, KNN, and SVM. RF algorithm obtained the highest prediction accuracy—78%. Dataset; The data set includes data on the performance of 105 students in the course “Object Oriented Programming”.
	Gap: exclude capture data from student's experience participating in LMS and video-conferencing
Abdulkream A. Alsulami [15]	DT, Naive Bayes, and RF. DT method with boosting outperformed with accuracy of 77%. Dataset: The data set for this study was obtained from the Kalboard 360 E-Learning system. Consists of 480 records with 17 attributes.
	Gap: improve the performance of the proposed technique to several online educational data sets to support decision-makers for high-impact e-learning development.
Barnabás Holicza [16]	SVM, DT, RF and k-nearest neighbor's algorithms. RF classifier performed well on both offline and online datasets, with accuracy values of 0.95 and 1.00, respectively. Dataset: Kaggle higher education students performance Evaluation. dataset contains 33 attributes, and originally had 145 records.
	Gap: quality of the data used
Muhammad Hammad Musaddiq [17]	Long Short-Term Memory (LSTM), SVM, RF, naïve bayes, Multilayer Perceptron (MLP). Long Short-Term Memory (LSTM) achieved higher accuracy of 90.9% as compared to other machine learning methods. Dataset: The dataset of 200 students in their first four semesters of bachelor's degree in the discipline of Computer Science. The data features included academic features, locational features, and semester behavioural features,
	Gap: exclude students' academic performance
Ali Al-Zawqari [18]	RF and artificial neural network with 86% and 88%, respectively. Dataset: The Open University Learning Analytics dataset (OULAD).
	Gap: the students' demographic information helps predict their performance, but their misuse.

Many previous studies focused on developing prediction models and evaluating results using the ML technique, with little attention focused on comprehending classification models for understanding

predictive features [6–18]. Understanding the black-box output of a machine-learning model was crucial for computing and examining the influence of features on individual and overall predictions, as well as evaluating useful features and investigating their interpretability and characteristics.

3. Material and Methods

a. Study Design

The purpose of this study was to examine students' academic performance in VCAOL during the Covid-19 pandemic. Academic performance was chosen as the subject of study by researchers because many students were affected by their academic performance, or it took time to adjust to online learning during the Covid-19 outbreak [2]. From September 2022 to January 2023, data was collected from university students at numerous universities in the Jakarta area using Google Form questionnaire. The questionnaire has 28 statements filled up. The 28 statements examined students' perceptions of VCAOL during the Covid-19 pandemic, including 4 perspectives namely video conference application (VC), internet connection (IC), student ability to study (SL), learning method (LM), and student knowledge (SK). There were 361 data gathered as a dataset. Following that, 28 statements on questionnaire were regarded as features for developing machine learning prediction models. Because supervised prediction required attribute as label prediction, as a label prediction, one statement, students' academic performance in VCAOL during Covid-19, was chosen, with options 'very degraded', 'decreased', 'stable', 'good', and 'very good'. This label represented a student's statement for academic performance during the Covid-19 pandemic. Features for prediction model are shown in Table 2. Subsequently the flowchart of the proposed work displays in Fig. 1.

Table 2 Features perspective

Perspective	Features	
Video conference application (VC) [19]	Benefit	UserInterface
	Efficient	Bored
	ListApplicationVicon	Interactive
	Frequency	Project
	EaseOfLearning	Feature
Internet connection (IC) [20]	EaseWorkGroup	Tools
	NetworkNot Affect	Often Disconnected
Students' ability to learn (SL) [19]	NetworkAffects Speed	Internet Problems
	ConstraintType	Completing Project
	FrequencyConstraint	
Learning method (LM) [19]	LearningAsUsual	SupportingMaterial
	AdequateMethod	IncreaseValue
		Ability
Student knowledge (SK) [20,21]	Performance	Competence
	Knowledge	Positive effect

b. Data Pre-processing

The data pre-processing step includes data quality assessment, data cleansing, data transformation, and data reduction [22]. Because all the statements in the questionnaire were previously designated as required inquiries, all inquiries must be answered, as a result, the missing value was not found during this activity. Subsequently, each statement was distinct, no duplicate statements were found, and hence no duplicate values were found as well. We determined the data type of each feature during data pre-processing; we detected 2 features on ordinal data type, 20 features on numerical data type, and 6 features on categorical data type. Following that, we transformed features for data compatibility. We executed mandatory changes for data compatibility by converting non-numeric features to numeric ones. We executed the change prior to training. Table 3 shows the converting value for each feature.

Table 3 Data transformation

Data type	Data transformation	Data type
Ordinal (n = 2)	<p>Value:</p> <p>Benefits attribute contains values: Very useful, moderately useful, Helpful, less useful.</p> <p>Efficient attribute contains the value:</p> <p>Very efficient, moderately efficient, Efficient, less efficient</p> <p>Data transformation:</p> <p>The value is converted to a number between 1 and 4, with 1 representing the lowest order level and 4 representing the highest order level.</p>	Ordinal (n = 2)
Numerical (n = 1)	<p>Value:</p> <p>Frequency contains integers: 0, 1, 2, 3 and >3</p> <p>Data Transformation:</p> <p>Numbers 0, 1, 2, and 3 are unchanged, while number > 3 is changed to number 4.</p>	Numerical (n = 1)
Categorical (n = 6)	<p>Value:</p> <p>Data in the form of options with categorical data types allows the user to select more than one option.</p> <p>Data transformation:</p> <p>Values are translated into numbers based on the number of choices selected in each statement.</p>	Categorical (n = 6)
Numerical (n = 19)	<p>Value:</p> <p>The data are derived from the numbers on a Likert scale that ranges from 1 to 5.</p> <p>Data transformation:</p> <p>Keep the numbers entered in each statement unchanged.</p>	Numerical (n = 19)

As a label prediction, one statement, students' academic performance in VCAOL during Covid-19, was chosen, with options 'very degraded', 'decreased', 'stable', 'good', and 'very good'. We converted each option into number: 'very degraded': 0, 'decreased': 1, 'stable': 2, 'good': 3, 'very good': 4.

c. Data splitting

To train any ML model, irrespective of the nature of the dataset used, the dataset must be split into training and testing data. In a data split, the training data set is used to train and construct models.

Training sets are frequently used to estimate different parameters or to compare different model performances [23]. We split the data using the hold-out method, which typically uses 80% of data for training and the remaining 20% of the data for testing.

d. Machine Learning Algorithms for Prediction

In the classification experiments, RF, SVM, and GNB were used. Based on machine learning, these algorithms have the potential to be used to predict students' academic performance [12-18]. RF is an ensemble of high-performing trees that have been combined into a single model. This approach outperforms the decision tree technique in terms of performance [24]. The support vector machine (SVM) is a class of supervised learning methods that can be used for classification, regression, and outlier detection. SVM techniques employ many kernel functions, the most prevalent of which are linear, nonlinear, polynomial, radial basis functions, and sigmoid functions. This kernel function is analogous to a two-layer perceptron model of a neural network, which functions as a neuron activation function. As a result, applying such a function aid in the modification of the prediction process in the same way as neural networks do [25]. The Naive Bayes (NB) method is a supervised classification algorithm that takes feature independence into account. It's very useful for datasets with a lot of features. The algorithm considers all features, even those with tiny effects on prediction. NB has received a lot of attention due to its simple classification model and excellent classification results [26].

Tables 4 and 5 show the parameter settings for the RF, SVM, GNB, and SHAP explainer configurations.

Table 4 Parameter setting for the algorithm.

Algorithm	Parameter setting	Definition
RF	n_estimators =100	The number of trees in the forest.
	Criterion = entropy	The function to measure the quality of a split.
	max_depth = none	The maximum depth of the tree
	min_samples_split = 2	The minimum number of samples required to split an internal node
	min_samples_leaf = 1	The minimum number of samples required to be at a leaf node.
	min_weight_fraction_leaf = 0.0	The minimum weighted fraction of the sum total of weights (of all the input samples) required to be at a leaf node.
	max_features = "sqrt"	The number of features to consider when looking for the best split
SVM	Cfloat =1.0	The regularization parameter (lambda) determines the relevance of misclassifications.
	Kernel: 'linear'	The kernel method is a mathematical methodology used in machine learning to analyze data. Linear Kernel is used when the data is linearly separable.
	degreeint =3	Polynomial kernel function degree ('poly'). It must not be negative.
	random_state = none	The random state is a model hyperparameter that regulates the randomness in machine learning models. None: this enables the function to make use of the global random state instance.
GNB	priorsarray-like of shape (n_classes,) =None	The classes' prior probability. If this option is specified, the priors are not modified based on the data.

Table 5 SHAP explainer configurations.

Bar plot	Parameter setting	Definition
Global bar plot	shap.plots.bar(shap_values, max_display=28)	The global importance of each feature is defined as its mean absolute value across all samples. The features are sorted in this order based on their impact on the prediction. The bar plot displays a maximum of ten bars by default, however, this can be changed using the max_display parameter = 28 (number of feature predictions in this research).
Local bar plot	shap.plots.bar(shap_value[0])	This plot depicts the primary features influencing the prediction of a single observation, as well as the magnitude of the SHAP value for each feature.

e. Performance Measure

Afterwards, we evaluated our model using a variety of relative metrics. In our evaluation, we used a confusion matrix. These measures were calculated so far and were based on True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). If the dataset is not balanced, accuracy may not be a good measure [16]. We calculated the following measure scores: accuracy, precision, F1-score, and recall. We also computed the area under the receiver operating characteristic curve (AUC). The receiver operating characteristic (ROC) curve measures a classifier's prediction quality. The optimal position is thus in the upper left corner of the plot, where false positives equal 1 and genuine positives equal 0. The AUC is a measure or degree of separability. This demonstrates how well the model can distinguish between classes. A higher AUC indicates that the model more accurately predicts class 0 as 0 and class 1 as 1 [27].

f. Model-Agnostic Interpretation

ML models are typically viewed as black boxes that receive specific features and produce predictions. Approaches based on interpretability can help overcome the problems associated with black-box models. Although machine learning algorithms may learn complex associations and enhance forecast accuracy, their inner workings are complex. Interpretability approaches come into play here by providing a lens through which to view these complex models [28]. Fully complicated models can be interpreted either globally or locally using model-agnostic interpretability strategies. Global interpretability explains the model's overall behaviour throughout the entire population. Local interpretability, on the other hand, provides explanations for a given model prediction [29]. SHapley Additive exPlanations (SHAP) was employed for model interpretation. SHAP is widely used to interpret various classification and regression models. In this method, attributes are ranked according to their contribution to the model, and the relationship between attributes and results can be visualized. Its absolute value influences the effect of the attribute, and its positive or negative value reveals the attribute's positive or negative impact on the prediction. A SHAP bar plot will use the mean absolute value of each attribute across all instances (rows)

of the dataset by default [28]. SHAP presents a simple approach to f by explaining the contribution of each attribute value. Model g is as follows: (see Eq 1) [28]. In this case, p is the number of attributes, $z = [z_1, z_2, \dots, z_p]$ is a simplification in the input x , where z represents the data prediction attributes and is 1 and the unused attribute has a z value of 0. Furthermore, $\phi_i \in \mathbb{R}$ reflects each attribute's contribution to the model.

$$g(z) = \phi_0 + \sum_{i=1}^p \phi_i z_i \quad (1)$$

We used SHAP global and local interpretation to build a model-agnostic classification model for understanding predictive features. Global interpretation combines SHAP values across numerous instances to understand the general behaviour of a machine learning model and identify the most essential features influencing its predictions. Local interpretation focuses on understanding the causes influencing individual predictions using SHAP values, which provide instance-specific interpretations [28].

4. Results

a. Machine learning algorithm's performance for prediction

We examined 361 records during our investigation. Fig. 2 shows the demographic data. Most of our data subjects were STEM (Science, Technology, Engineering, and Math) students. Our respondent was a domineering male. Most of participants were computer science students, followed by information systems, humanities, and engineering students.

The label in our model was the student's academic performance statement. Fig. 3 displays the frequency distribution of each class. The dataset must be separated into training and testing data for the ML model to be trained. Using the hold-out method, we randomly divided the dataset into two parts, allocating 80% for training ($n = 298$) and the remaining 20% for testing ($n = 72$). The confusion matrix of the testing data for each algorithm is shown in Fig. 4. The number of positive and negative observations correctly identified is referred to as accuracy. Precision addresses the subject of how many affirmative identifications were right. The F1-score is the harmonic mean of precision and recall, not only the accuracy value. The recall answers the topic of how many true positives were correctly detected [22]. Table 6 shows the performance of prediction model using RF, SVM, and GNB.

Table 6 The Prediction Model Performance

Model	Accuracy	Precision	Recall	F1-score	ROC AUC
RF	60.27%	59.46%	60.27%	59.51%	87%
SVM	43.84%	44.92%	43.83%	43.80%	86%
GNB	38.35%	39.87%	38.35%	38.50%	81%

Fig. 5 shows the ROC-AUC curve for RF, SVM and GNB. A ROC curve shows the true positive rate on the Y-axis and the false positive rate on the X-axis, both overall and by class. The best position is in the plot's upper left corner, where false positives are 0 and genuine positives are 1. The AUC measures the difference between false positives and true positives. A higher AUC indicates a more accurate overall model [30].

The difference between macro and micro averaging is that macro equally weights each class, whereas micro equally weights each sample. If each class has an equal number of samples, macro and micro will have the same score [31]. To compute the average metric, a micro-average will aggregate the contributions of all classes. If there is a possibility of class imbalance in a multi-class classification system, micro-average is better. When the classes are imbalanced, the micro average gives more weight to the majority class [31]. We discovered that our dataset was imbalanced since most of the data belonged to class 3 (see Fig.3), hence the micro average gave higher weight to the majority class (class 3, students' academic performance in VCAOL during Covid-19 was good). Macro average, on the other hand, calculates each class's performance separately and then takes the unweighted mean of the class-wise performance. Macro average gives each class equal weight and is useful when all classes are of equal importance [31]. We discovered that there was no equal weight for each, and that class 0 produced NAN for ROC AUC, therefore the macro average ROC AUC was also NAN.

b. Model-agnostic interpretation

Based on Table 4, RF outperformed almost all other classification and prediction metrics, so we built model-agnostic ML to interpret classification model for comprehending predictive properties. We used SHAP on global interpretation to identify the most important features influencing its predictions model and local interpretation to explain the predictions of a single instance by evaluating the contribution of each feature to the single prediction observation. The SHAP Python package was used to compute SHAP values and generate visualizations. A global feature importance plot was created by feeding a matrix of SHAP values into the bar plot function. This approach allocated the global relevance of each feature to coincide with the mean absolute value of that feature across all samples. The x-axis displays the average absolute SHAP value of each feature. The features are listed in descending order based on their impact on the model's prediction. This arrangement considers the absolute SHAP value, thus whether the feature affects the prediction positively or adversely is unimportant [32]. Fig. 6 (a) shows a global feature importance plot for the RF method, which was used to predict students' academic performance. The plot was constructed with the Python libraries bar type. The most significant attribute for predicting students' academic performance in VCAOL during Covid-19 was 'student academic performance was still well achieved during the learning process, despite using video conferencing during the Covid-19 pandemic' (Performance), followed by 'the VCAOL learning method increasing students' score' (IncreaseValue), 'VCAOL learning method ability to help students' (Ability), 'the number of times in one day that a video conferencing application was used as a learning medium during lectures' (Frequency), and 'student ability

to complete a project' (CompletingProject). Following that, we used plots of individual data points to assess the implications on an individual basis (local interpretability). The plot shows the key features influencing the prediction of a single observation, as well as the magnitude of the SHAP value for each feature. To emphasize the contributions of different variables, the bar plot is zero-centered. Each bar represents the SHAP value of a single feature. Fig. 6 (b) displays local interpretability for the RF algorithm, which was used to predict students' academic performance in VCAOL during Covid-19. Blue bars on negative shift represent features, indicating that decreased occurrence counts for the feature have a negative impact on the result while red bars on positive shift represent features, indicating that increased occurrence counts for the feature have a positive impact on the result of a single prediction observation.

5. Discussion

a. Findings

We created a machine learning model to predict students' academic performance in VCAOL during the Covid-19 pandemic. We evaluated three models: RF, SVM, and GNB. In our experiment, RF outperformed the other models on practically every classification and prediction metric (accuracy, precision, recall, F1-score and ROC AUC) (RQ1). Following that, we performed a SHAP analysis to uncover insights and patterns that were not obvious from the initial RF features. A global interpretability analysis using the SHAP summary_plot method with a bar plot type revealed that 'student academic performance was still well achieved during the learning process, despite using video conferencing during the Covid-19 pandemic' (Performance) was the most critical attribute for predicting students' academic performance in VCAOL during Covid-19. Subsequently, to evaluate the contribution of each feature to the single prediction observation, we used the local interpretability SHAP bar plot approach to explore the contribution of individual features to individual predictions. In this context, most all features are represented on the blue bars in negative shift, indicating that decreased occurrence counts for the feature have a negative impact on the result. It can be explained for 'student academic performance was still well achieved during the learning process, despite using video conferencing during the Covid-19 pandemic' (Performance), when Performance decreased it contributed negative impact on students' academic performance in VCAOL during Covid-19. The same finding explanation applies to Increase Value, Project, and all features with blue bars in a negative shift. (RQ2). A deep investigation needs to be done on the ConstraintType feature which is represented by the red bars on positive shifts represent features, which indicates that increasing the number of occurrences of a ConstraintType has a positive impact on the observed single-predictive results. ConstraintType refers to the number of type issues that students confront when using VCAOL. Our findings have the following implications: (1) we created a model to predict students' academic performance in VCAOL during Covid-19 pandemic when the learning style switched from face-to-face to online mode, and it could influence students' academic performance. (2) we interpreted the prediction model and evaluated informative aspects to aid comprehension. We also explored their interpretability and features to explain an instance's predictions. Previous research has confirmed some of our findings, verifying our perspectives and features. Student ability to learn (SL),

student knowledge (SK), video conference application (VC), students' perceptions, attitudes toward VC, intention to use VC, and actual use of VC were considered for analysis when attempting to identify and predict students' academic performance in VCAOL during Covid-19 [3-5]. Our finding filled the gap while several previous studies have investigated RF, SVM, and GNB for academic prediction in students, generating better results than ours, but they did not develop model-agnostic ML to evaluate classification models for comprehending predictive features [11-17].

b. Limitations

Certain limitations in this study open the way for future research, namely: (1) data was collected from September 2022 to January 2023 during the semester break of the university term, with only a few students available to serve as respondents, (2) only a few students from a few universities in the Jakarta area participated, (3) this work did not take into consideration the individual characteristics of the courses and faculties at the university under examination, and (4) another limitation of this work was the imbalanced nature of the dataset. The strategies for dealing with it must be implemented.

6. Conclusion

The study's goals were to develop machine-learning models that can help predict students' academic performance in VCAOL during Covid-19. ML approaches were used to train prediction models as well as to develop model-agnostic for recognizing and comprehending predictive characteristics. During this phase, the SHAP framework was used. By using the SHAP framework we leveraged its global interpretation to discover the most important factors influencing its prediction model, and its local interpretation to explain the predictions of a single instance by assessing each feature's contribution to the single prediction observation. Our proposed work has the potential to help assess and reveal insights into students' academic performance in the VCAOL context and in the future. A more in-depth investigation of the effects of VCAOL for the types of courses (STEM or non-STEM) in each faculty is necessary. In the future, we intend to add more students and universities in the Jakarta area as respondents to collect more students' experiences and academic performance data in VCAOL during the Covid-19 pandemic. Universities should strive to implement programs to mitigate these consequences, and the efficacy of any activities should be monitored and assessed. Handling imbalanced data is another potential target for this research. Imbalanced data is a common issue in machine learning, where one class has far more observations than the other. This can lead to biased models and poor performance in the minority class. Techniques for coping with skewed prediction findings must be established to avoid them.

Declarations

Ethics Approval

No need ethics approval.

Consent to Participate

No need consent to participate

Availability of data and Material

Data is attached in the form of rar data (png and .xlsx)

Competing interests

We don't have a financial and non-financial competing interests must be declared in this section.

Funding

This work is supported by Directorate General of Higher Education, Research, and Technology Ministry of Education, Culture, Research, and Technology and Research and Technology Transfer Office, Bina Nusantara University

Authors' Contribution

Eka Miranda : Data Collection, writing the manuscript, analysis and interpretation data

Mediana Aryuni, Albert Verasius Dian Sano : Programmer

Mia Ika Rahmawati, Siti Elda Hiererra : Data Collection, writing the manuscript

Acknowledgement

The research was sponsored by Directorate General of Higher Education, Research, and Technology Ministry of Education, Culture, Research, and Technology According to the Research Contract for Fiscal Year 2023 No 1402/LL3/AL.04/2023, dated June 26, 2023 With a contract at Bina Nusantara University No: 149/VR.RTT/VII/2023. with the title Prediction Model for Academic Performance Evaluation of Student Online Learning using a Machine Learning Approach in the Fundamental Research Plan - Regular for the Fiscal Year 2023.

References

1. Sun HL, Sun Y, Sha FY, Gu XY, Hou XR, Zhu FY, Fang PT. The influence of teacher-student interaction on the effects of online learning: based on a serial mediating model. *Front. Psychol.* 2022;13:1-10.
2. Ministry of Education, Culture, Research, and technology online learning in the midst of the covid-19 pandemic, a maturity challenge, Accessed: August 2023. 2020.
<https://pusdatin.kemdikbud.go.id/pembelajaran-online-di tengah-pandemi-covid-19-tantangan-yang-mendewasakan>.

3. Okabe-Miyamoto K, Durnell E, Howell RT, Zizi M. Video conferencing during emergency distance learning impacted student emotions during Covid-19. *Comput Hum Behav Rep.* 2022;7:1-8.
4. Camilleri MA, Camilleri AC, Remote learning via video conferencing technologies: Implications for research and practice. *Technol Soc.* 2022;68:1-10.
5. Bailey DR, Almusharraf N, Almusharraf A. Video conferencing in the e-learning context: explaining learning outcome with the technology acceptance model. *Educ Inf Technol* 2022; 27: 7679–7698.
6. Tarik A, Aissa H, Yousef F. Artificial intelligence and machine learning to predict student performance during the Covid-19. *Procedia Comput Sci.* 2021;184:835-840.
7. Alruwais N, Zakariah M. Evaluating student knowledge assessment using machine learning techniques. *Sustainability.* 2023;15 (7):1-25.
8. Yağc M. Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments.* 2022;9 (11):1-19.
9. Dervenis C, Kyriatzis V, Stoufis S, Fitsilis P. Predicting students' performance using machine learning algorithms, in IOP Conf. Series: Materials Science and Engineering, 2022, pp. 1-7.
10. Jang Y, Choi S, Jung H, Kim H. Practical early prediction of students' performance using machine learning and eXplainable AI. *Education and Information Technologies.* 2022;27:12855–12889.
11. Cho E, Chang TW, Hwang G. Data pre-processing combination to improve the performance of quality classification in the manufacturing process. *Electronics* 2022; 11(3):1-15.
12. Jawad K, Shah MA, Tahir M. Students' academic performance and engagement prediction in a virtual learning environment using random forest with data balancing. *Sustainability.* 2022; 4(22):1-15.
13. Huynh-Cam TR, Chen LS, Le H. Algorithms, using decision trees and random forest algorithms to predict and determine factors contributing to first-year university students' learning performance. *Algorithms.* 2021;4(11):1-17.
14. Gaftandzhieva S, Talukder A, Gohain N, Hussain S, Theodorou P, Salal Y, Doneva R. Exploring online activities to predict the final grade of student. *Mathematics.* 2022;10(20):1-20.
15. Alsulami AA, AL-Ghamdi ASL, Ragab M. Enhancement of e-learning student's performance based on ensemble techniques. *Electronics.* 2023; 12(6):1-18.
16. Holicza B, Kiss A. Academic performance using machine learning algorithms predicting and comparing students' online and offline. *Behav Sci (Basel).* 2023; 13(4):1-21.
17. Musaddiq MH, Sarfraz MS, Shafi N, Maqsood R, Azam A, Ahmad M. Predicting the impact of academic key factors and spatial behaviors on students' performance. *Appl. Sci.* 2022;12(19):1-21.
18. Al-Zawqari A, Peumans D, Vandersteen G. A flexible feature selection approach for predicting students' academic performance in online courses. *Computers and Education: Artificial Intelligence.* 2022;3:1-14.
19. Kerzic D, Alex JK, Alvarado RP, Bezerra DD, Cheraghi M, Dobrowolska B, Aleksander A. Academic student satisfaction and perceived performance in the e-learning environment during the Covid-19

- pandemic: evidence across ten countries. *Plos One*. 2021;1-23.
20. Wei HC, Chou C. Online learning performance and satisfaction: do perceptions and readiness matter? *Distance Education*. 2020; 41(1): 1-13.
 21. Salzman J, Williamson M, Rey AE, Kibble J, Kauffman C. Effects of voluntary attendance patterns on first-year medical students' wellness and academic performance during Covid-19. *Adv Physiol Educ*. 2021;45(3):634-643.
 22. Cho E, Chang TW, Hwang G. Data pre-processing combination to improve the performance of quality classification in the manufacturing process. *Electronics*. 2022;11:1-15.
 23. Limberg C, Wersing H, Ritter H. Beyond cross-validation—accuracy estimation for incremental and active learning models. *Mach Learn Knowl Extr*. 2022;2:327–346.
 24. Johnson A, Cooper GF, Visweswaran S. A novel personalized random forest algorithm for clinical outcome prediction. *Stud Health Technol Inform*. 2022;290:248- 252.
 25. Cemiloglu A, Zhu L, Arslan S, Xu J, Yuan X, Azarafza M, Derakhshani R. Support Vector Machine (SVM) application for Uniaxial Compression Strength (UCS) prediction: a case study for maragheh limestone. *Appl. Sci.* 2023;13(4):1-14.
 26. Arif MS, Mukheimer A, Asif D. Enhancing the early detection of chronic kidney disease: a robust machine learning model. *Big Data Cogn. Comput*. 2023;7(3):1-16.
 27. He S, Qu L, He X, Zhang D, Xie N. Comparative evaluation of 15-minute rapid diagnosis of ischemic heart disease by high-sensitivity quantification of cardiac biomarkers. *Exp Ther Med*. 2020;20(2):1702-1708.
 28. Lu S, Chen R, Wei W, Belovsky M, Lu X. Understanding heart failure patients EHR clinical features via SHAP interpretation of tree-based machine learning model predictions. in *AMIA Annu Symp Proc*, 2022, pp. 813-822.
 29. Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: a review of machine learning interpretability methods. *Entropy*. 2021;23(1):1-45.
 30. Budholiya K, Shrivastava SK, Sharma V. An optimized XGBoost based diagnostic system for effective prediction of heart disease. *J King Saud Univ Comput Inf Sci*. 2022;34(7):4514-4523.
 31. Han H. The Utility of Receiver Operating Characteristic Curve in Educational Assessment: Performance Prediction. *Mathematics* 2022;10(9):1-11.
 32. Futagami K, Fukazawa Y, Kapoor N, Kito T. Pairwise acquisition prediction with SHAP value interpretation. *The Journal of Finance and Data Science*. 2021; 7:22-44.

Figures

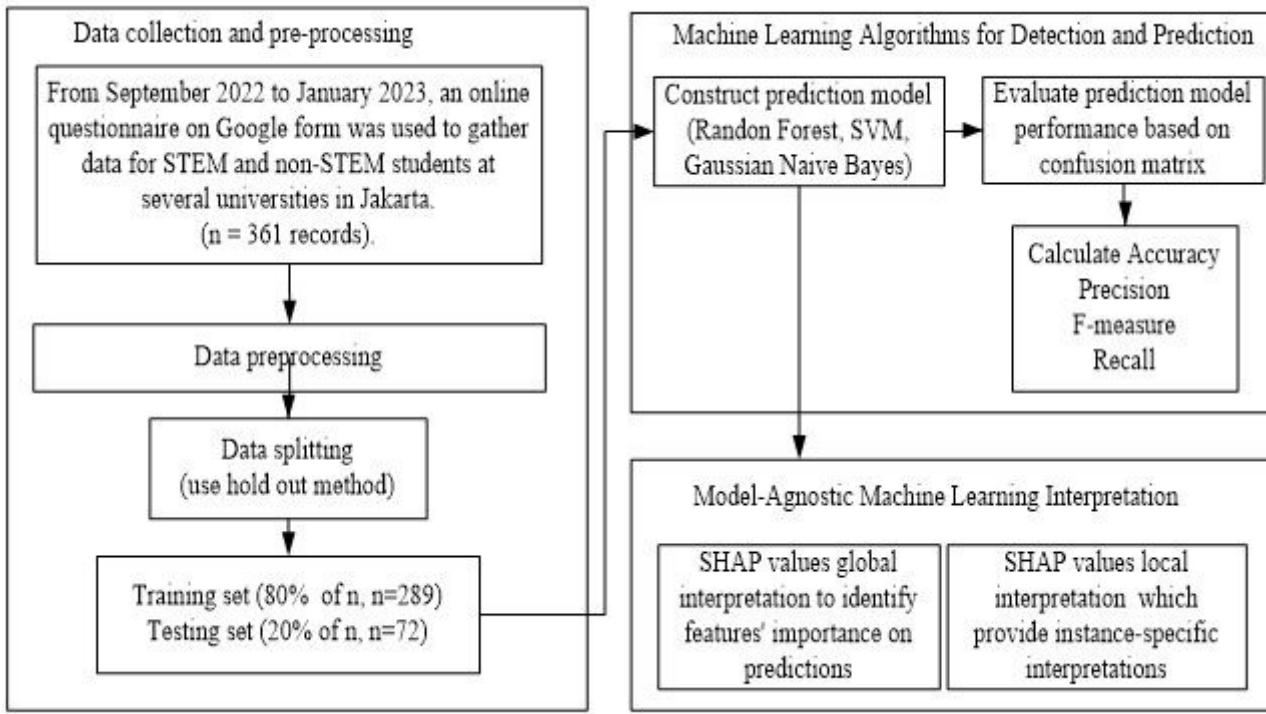


Figure 1

The flowchart of the proposed work

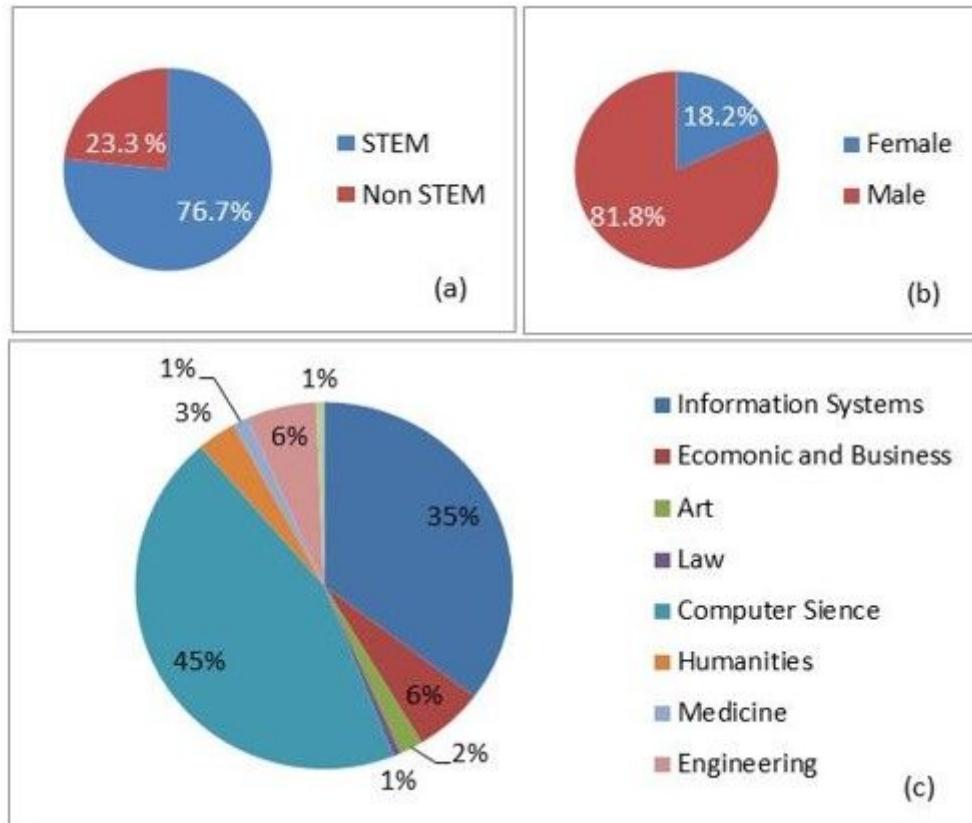


Figure 2

Demography data (%): (a) STEM and Non-STEM students, (b) Gender and (c) Faculty

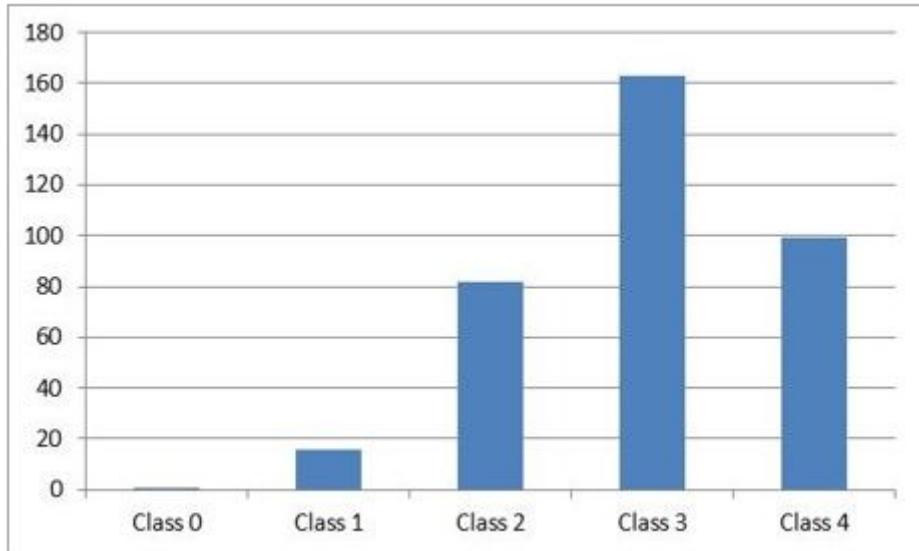


Figure 3

Frequency distribution of each class

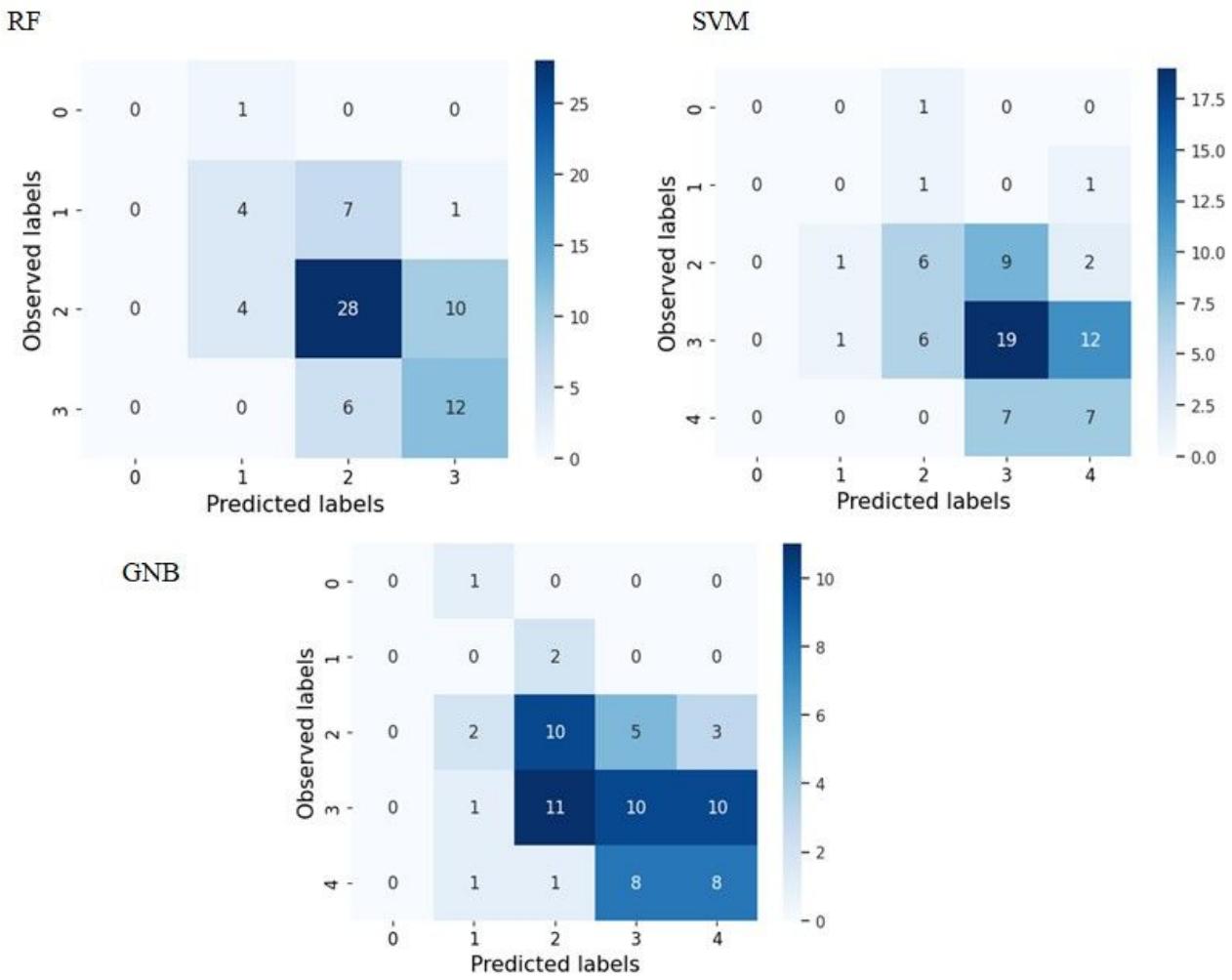


Figure 4

Confusion matrix for testing data

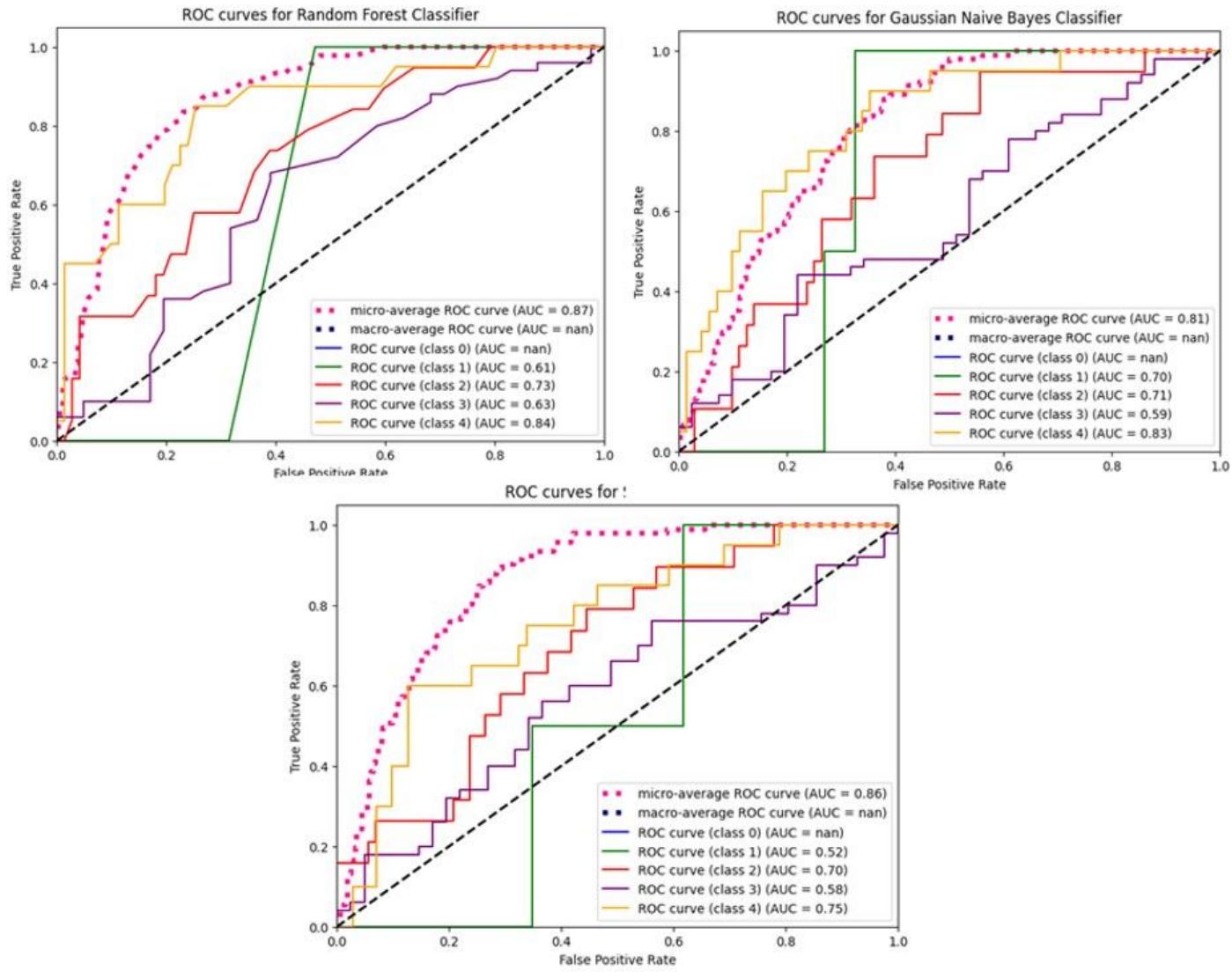


Figure 5

Receiver operating characteristic (ROC) curve and area under the curve (AUC) for RF, SVM and GNB

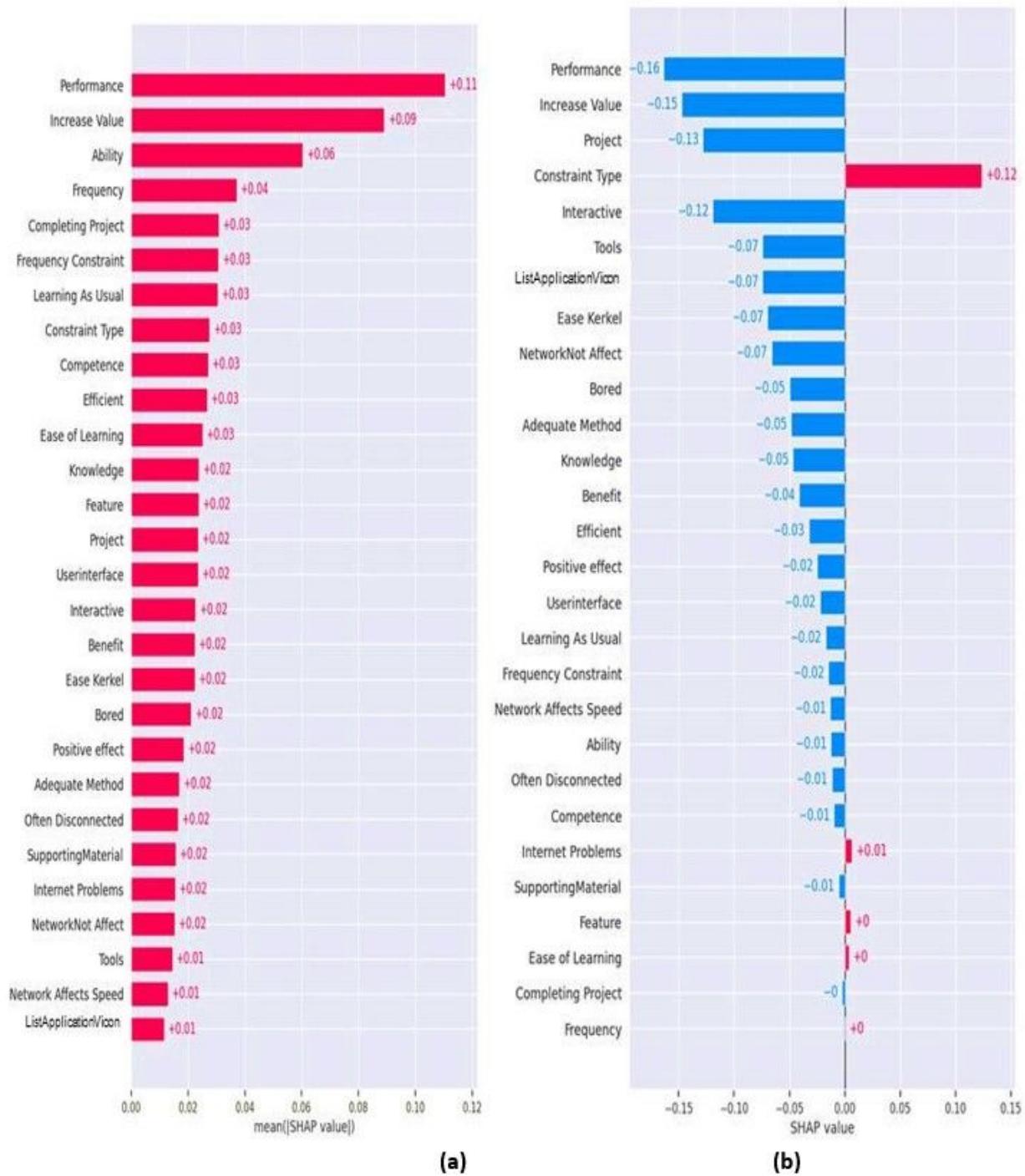


Figure 6

(a): Global interpretability: feature importance for the RF algorithm to predict students' academic performance (as visualized by summary_plot method with plot type bar in the Python library), (b): Local interpretability for the RF algorithm to predict students' academic performance (as visualized by the plot method with plot type bar plot in the Python library).