العنوان باللغة الانجليزية

Forecasting students Grades based on students' performance using machine learning techniques

العنوان باللغة العربية

التنبؤ بدرجات الطلاب بناء على أداء الطلاب باستخدام تقنيات تعلم الاله

Dr. Mohamed Ali Elhayes
Associate professor and head of the Information Systems Department
Faculty Of Business, Economics & Information Systems
Misr University for Science& Technology

المجلد الثالث - العدد الثامن – مايو ٢٠٢٥

ISSN-Print: 2812-6114 ISSN-Online: 2812-6122

موقع المجلة على بنك المعرفة المصري

https://aiis.journals.ekb.eg/contacts?lang=ar

Abstract

The performance of higher education institutions (HEIs) is extensively assessed by student success rates, particularly when it comes to the quality of educational services. In this paper, the forecasting of the student grades with the different factors influencing academic performance using the advanced machine learning techniques is presented. Predictive models are developed to forecast student performance based on an analysis of a dataset of 10,000 records that includes details about exam results, the amount of time students spend studying, their use of online learning, and other pertinent characteristics. This paper examines the efficacy of many machines learning models, including Artificial Neural Networks, Random Forest, Support Vector Machines (SVM), Naive Bayes, and XGBoost, in forecasting final grades. Based on our findings, XGBoost performs better than other models, averaging 95% accuracy, precision, recall, and F1_Score. This paper summarizes how machine learning techniques can advance educational research, offering practical insights for educators to help identify at-risk students and establish interventions that can positively impact educational outcomes. Having developed a comprehensive strategy for data analysis, education institutes can now take advantage of advanced analysis techniques to better understand the factors impacting their students thereby making more effective data-driven decisions to support students in their academic endeavors. Making these predictive models more reliable and applicable in educational environments will certainly contribute to better student results and the quality of educational services offered by HEIs.

Keywords: XGBoost model, Students performance, higher education institutions, machines learning algorithms.

التنبؤ بدرجات الطلاب بناء على أداء الطلاب باستخدام تقنيات تعلم الاله

المستخلص:

يُقيَّم أداء مؤسسات التعليم العالي (HEIs) على نطاق واسع من خلال معدلات نجاح الطلاب، وخاصةً فيما يتعلق بجودة الخدمات التعليمية. في هذه الدراسة، سنتنبأ أيضًا بدرجة الطالب مع العوامل المختلفة المؤثرة على الأداء الأكاديمي باستخدام تقنيات تعلم الاله المتقدمة. طُوِّرت نماذج تنبؤية للتنبؤ بأداء الطلاب بناءً على تحليل مجموعة بيانات تضم ١٠٠٠٠ سجل تتضمن تفاصيل حول نتائج الامتحانات، والوقت الذي يقضيه الطلاب في الدراسة، واستخدامهم للتعلم عبر الإنترنت، وخصائص أخرى ذات صلة. تبحث هذه الدراسة في فعالية العديد من أساليب التعلم الآلي، بما في ذلك الشبكات العصبية، والغابات العشوائية، وآلات المتجهات الداعمة (SVM)، وخوارزمية بايز، وخوارزميةXGBoost ، في التنبؤ بالدرجات النهائية. بناءً على النتائج التي توصلنا إليها، فإن أداء XGBoost أفضل من النماذج الأخرى، حيث يبلغ متوسط الدقة والدقة والتذكر ودرجة F1 ٩٥% . تُلخّص هذه الدراسة كيف يُمكن لتقنيات تعلم الاله أن تُطوّر البحث التربوي، مُقدّمةً رؤىً عمليةً للمعلمين تُساعدهم في تحديد الطلاب المُعرّضين للخطر، ووضع تدخلاتٍ يُمكن أن تُؤثّر إيجابًا على النتائج التعليمية. بعد تطوير استراتيجية شاملة لتحليل البيانات، يُمكن للمؤسسات التعليمية الآن الاستفادة من تقنيات التحليل المُتقدّمة لفهم العوامل المُؤثّرة على طلابها بشكلٍ أفضل، وبالتالي اتخاذ قراراتٍ أكثر فعاليةً قائمةً على البيانات لدعم الطلاب في مساعيهم الأكاديمية. إنّ جعل هذه النماذج التنبؤية أكثر موثوقيةً وقابليةً للتطبيق في البيئات التعليمية سيُسهم بلا شكّ في تحسين نتائج الطلاب وجودة الخدمات التعليمية التي تُقدّمها مؤسسات التعليم العالي.

الكلمات المفتاحية: XGBoost model، درجة الطالب، أداء الطلاب، مؤسسات التعليم العالي، خوارزميات التعلّم الآلي.

1. Introduction

Graduation rates are not only a measure for how well the educational programs are doing, but they are also an important measure for students and educators across all disciplines. Success rates shed light on what success looks like for students and provide guidance on how to plan their academic journey. Academic progress rates provide students with tangible feedback on their learning, helping to motivate and guide further effort, and for educators are crucial for tracking at-risk students and creating timely interventions [1]. In recent years, the value of accurate student success rate data has become far greater. This data that university decision-makers can use to decide how to increase the academic performance of their institutions. Data-driven solutions also facilitate the creation of impactful initiatives to enhance student achievement, and thus, ultimately fulfil the promise of high-quality education [2]. can extend the highly contextual research and development in the fields of education and routing in order to yield accurate outcomes with this approach, which is the aim of this paper: A case region was conducted where the final grades of the students can be predicted from their activities in the academic term. This paper aims to provide a model where students predict performance using advanced machine learning techniques. "Not only does this model help identify students that may need additional help, but it will also give information about the factors that matter most for achieving academic success". Machine learning plays a useful role in educational data mining because the idea is to mine and analyze a huge amount of data and find detectable patterns that may not be instantly recognized. Insofar as employ these models, can illuminate the tourney of factors at play in the student success journey and enable more effective interventions to help students navigate their academic paths [3]. The methodology used for this paper, specifics of the data collection procedure, and the machine learning algorithms used will all be covered in the parts that follow. Additionally, the Results out will discuss the model's prediction and their consequences for educational policy and practice.

2. Related Work

P.Gil etal [4] used a database from a Higher Institution in Portugal, covering 10 academic years and 9652 first-year bachelor's degree students, to predict academic success with data mining techniques. The methodology used was based on CRISP-DM, it included gathering and pre-processing data as well as using machine learning models including Artificial Neural Networks, Random Forests, Decision Trees, and Support Vector Machines (SVM). Feature engineering and sensitivity analysis were the models applied. When the AUC values are 0.77 at the start of the first semester, 0.91 at the completion of the first semester, and 0.94 at the completion of the second semester, SVM yields the best results. In future work, which will involve improving data quality, investigating new data sources along with designing models for each individual school. Limitations included reliance on institutional data, potential biases, and the exclusion of certain features due to data quality issues.

F.Qiu etal [5] conducted a study on the Open University Learning Analytics Dataset (OULAD), which includes information on 22 courses and 32,593 students, to predict the performance of e-learning. The Behavior Classification-based E-learning Performance (BCEP) prediction framework is part of the suggested approach, along with related techniques like feature fusion, data cleaning, and feature selection. Machine learning algorithms like SVM, Naïve Bayes, KNN, and SoftMax are then used for model training. To increase prediction accuracy, the Process-Behavior Classification (PBC) model was additionally Presented. Predictive performance (accuracy, Kappa, F1_Score) of the BCEP framework and PBC model were found to be significantly better than traditional models, where Group 3 exhibited maximum accuracy (95.44%-97.40%), F1_Score (0.9685-0.9818) and Kappa values (0.8865-0.9364), respectively. In the future, researchers can further evaluate and optimize the BCEP framework in various e-learning entities, as well as explore multi-aspect learning outcomes such as e-learning emotions. Limitations included the reliance on a single dataset and the need for further validation in real-world applications.

A.Daza etal [6] investigated the deployment of algorithms for data mining to predict the academic achievement and Performance of students at universities. The dataset used in this study includes attributes like CGPA, gender, socio-economic background, and has been gathered from IEEE Xplore and Science Direct among other sources. The approach adopted the KDD process, which is characterized by a series of iterative and interactive phases. Naïve Bayes, Artificial Neural Networks (MLP) and Decision Trees (J48) were used in research, based on Weka used as a tool for developing models. Naïve Bayes and J48 are found to have high accuracy with precision and recall metrics justifying the models' functionality. Future work may consider other attributes and more advanced algorithms to improve prediction accuracy. While useful, such approaches are limited to the particular datasets that were used to train this model and will need to be validated further in different educational settings.

M.Yağcı [7] examined the prediction of the undergraduate students' final assessments were based on their instructor, department, and midterm grades. The dataset includes the academic records of 1854 students from a Turkish institution during the 2019–2020 autumn semester. We report on performance over ten-fold cross-validation using machine learning techniques in this paper, including Random Forests, Artificial Neural Networks, Support Vector Machines, Logistic Regression, Naïve Bayes, and k-neighbor classifiers. The classification accuracy was maximum around 74.6% for Random Forests and Artificial Neural Networks. This could lead to better predictions in future work, by including more parameters or trying other machine-learning algorithms. Limitations such as dependence on certain academic data and necessity for more tests in various educational environments are discussed.

S.Gaftandzhieva etal [8] used a dataset from the University of Plovdiv, including the final grades of 105 university students, Moodle online activity environment, and attendance at Zoom lectures, to predict academic performance. My methodology included balancing the dataset on the basis of single point crossover, as well as using machine learning algorithms (Random Forest, XGBoost, KNN, SVM) with 70% for training data and 30% for testing data, as well as statistical techniques including logistic regression and chi-square tests, along with five-fold cross-validation. Students' final grades showed a substantial correlation with their online activity and attendance, with Random Forest offering the highest prediction accuracy at 78%. Future work includes increasing the dataset, using deep learning models, and generating an application that obtains real-time data. The study's retrospective approach, tiny dataset size, and moderate prediction accuracy were among its limitations.

H.Adu-Twum [9] conducted a Study on the study of the potential of advanced data analytics technology in making predictions about student success in higher education through a dataset composed of demographic, socioeconomic, academic, and financial variables of students who attended the Polytechnic Institute of port Alegre between 2008 and 2019. Data science exploration data analysis, feature engineering and building of both predictive (Logistic Regression, Random Forest, Decision Tree Classifier and Support Vector Machine (SVM) Gradient Boosting. models encompassed data cleansing, correlation examination, and model assessment in terms of specificity, sensitivity, and F1 scores. Of these, results showed that Gradient Boosting model outperformed others by accurately predicting 94.4% of dropouts. Key predictors consisted of approved circular units in the first semester and up to date tuition fees. Future research should diversify data sources to include behavioral and psychological data and apply advanced approaches such as deep learning and Explainable AI. Limitations included the focus on a single institution and the exclusion of qualitative factors.

F.Kohun etal [10] discussed the effects of ai and ML on HE, based on a dataset constructed from earlier research and modern literature. The research was based on the literature review type of methodologies on AI&ML usage in HE. Analyses included the use of AI&ML in admissions processes, models of personalized learning, predictive analytics, and administration-related tasks. Findings underscored that AI&ML had the potential to help improve teaching, learning, and research, as well as making the administration more efficient, and tackle challenges such as algorithmic biases and data privacy issues. Future work must be no-action AI&ML inclusive (from HE to DEI) and tinker towards integration of AI&ML in HE not forgetting ethics and education. That said, there were limitations, such as early stages of AI&ML adoption and a need for more critical preparedness to engage with the challenges and risks associated with it.

S.Hoca etal [11] applied on a dataset from Eastern Mediterranean University, comprising information on 20,974 students enrolled between 2015 and 2020, to use machine learning to predict student dropout. Before training and testing different machine learning models, such as Support Vector Classifier, K-Nearest Neighbors, Logistic Regression, Naïve Bayes, Artificial Neural Network, Random Forest, Classification and Regression Trees, and Categorical Boosting, the methodology comprised data collection, pre-processing, and feature extraction. Techniques included parameter tuning and feature importance analysis. Results indicated that X-Boosting achieved the highest F1_Score of 82%. Future work should incorporate additional data sources like socioeconomic and behavioral information to enhance prediction accuracy. Limitations included the focus on a single institution and the exclusion of detailed course performance metrics.

H.Gharkan etal [12], A scholar used the Open University Learning Analytics (OULAD), Students' Academic Performance Dataset (xAPI-Edu-Data), and Student Performance Dataset (Student-Math and Student-Por) datasets to predict performance of students in higher education. 3 different machine learning approaches, different type of algorithms like Random Forest, SVM, Gradient Boosting, AdaBoost, LSTM, CNN, hybrid (clustering and classification) was used, mentioned regarding methodology (data gathering, preprocessing of the same). It was found that predictive analytics have a significant effect on enhancing performance and retention of students; the accuracy of the models varied from 97% in XGBoost and 95% in ensemble models. Further research should improve data legitimacy, reduce biases in algorithms and produce user-friendly systems. Some of the limitations were the dependency on particular datasets, the possibility of biases within the data used, and the necessity of creating more generalized models.

3. Research Methodology

This paper employed a systematic methodology to Forecasting students grades using machine learning models and compare their performance to identify the most effective model. The dataset, sourced from Kaggle, comprised 10000 student records with features such as study hours, use of educational technology, and self-reported stress levels, with final grades as the target variable. Data preprocessing, conducted in Jupyter Notebook using Python libraries (pandas, NumPy, scikit-learn), involved handling missing values (imputed with median/mode), encoding categorical variables (one-hot encoding), scaling numerical features (StandardScaler), and selecting key predictors via Recursive Feature Elimination. Five models—Artificial Neural Network, Random Forest, Support Vector Machine, Naive Bayes, and Gradient Boosting (XGBoost). Model performance was evaluated on a test set (20% of data) using metrics accuracy, Recall, Precision, and F1-score for categorical grades.

4. The Proposed Model for Forecasting the Student Grade

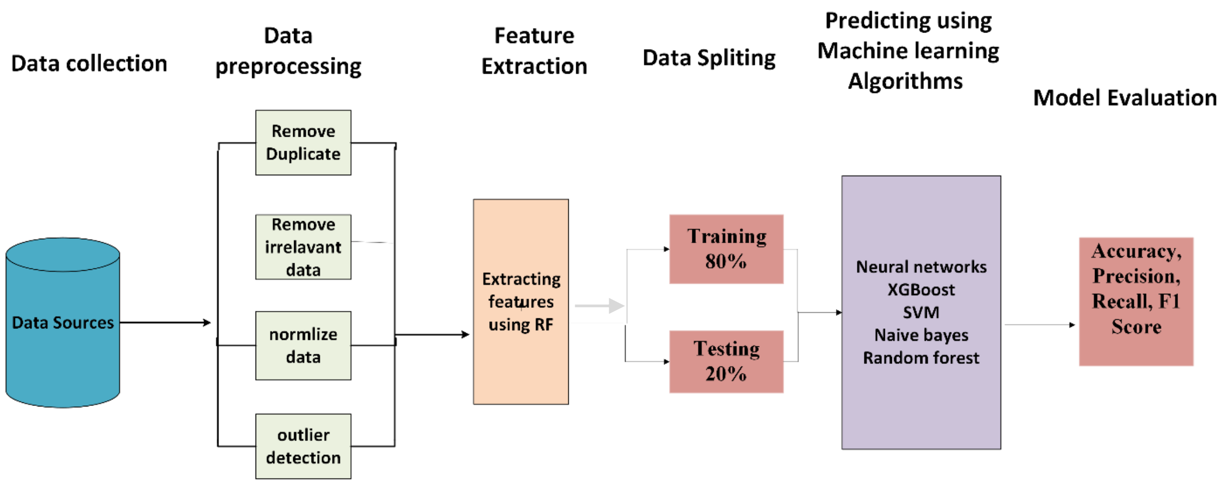

Figure (1): The Proposed Model

4.1 Data Collection

This dataset provides insights into how different study habits, learning styles, and external factors influence student performance. It includes 10,000 records, covering details about students' study hours, online learning participation, exam scores, and other factors impacting academic success.

4.2 Dataset Features

Table (1): illustrates the Features and their description.

Feature	Description
Student_ID	Each student is given a unique identification number.
Age	The student is between the ages of 18 and 30.
Gender	The student's gender (either male, female, or other).
Study_Hours_per_Week	the total amount of time a student spends studying per week, which might range from five to fifty hours
Preferred_Learning_Style	The students preferred primary techniques of learning (visual, auditory, reading/writing, or kinesthetic).
Online_Courses_Completed	The number of online courses the student has completed, ranging from 0 to 20.
Participation_in_Discussions	Whether the student actively participates in academic discussions (Yes or No).

Assignment_Completion_Rate (%)	the proportion of the student's assignments that were finished, which can range from 50% to 100%.
Exam_Score (%)	The final exam score of the student, which might range from 40% to 100%.
Attendance_Rate (%)	The student's attendance percentage, which might range from 50% to 100%.
Use_of_Educational_Tech	refers to the application of technological tools, resources, and systems to enhance learning processes (Yes- No).
Self_Reported_Stress_Level	reflects how stressed a person perceives themselves to be at a given time or in a specific context (High, Medium, Low).
Time_Spent_on_Social_Media (hours/week)	Weekly hours spent on social media (0-30 hours)
Sleep_Hours_per_Night	Average amount of time spent sleeping (4–10 hours)
Final_Grade	Exam results determine the grade (Excellent, Very Good, Good, Fair, Failed).

4.3 Data pre-processing

An essential phase in every data analysis procedure is data pre-processing, which keeps outputs from being illogical or deceptive.

Data cleaning allows to remove errors and inconsistencies from the data and make better decisions [13].

The stage of this phase aimed to address inconsistencies in the data by processing lost and noisy data. More specifically, we removed inaccurate data from the dataset. To achieve this, the following steps were undertaken:

– Removing duplicate data: When examining the data, no duplicated data was found.

– Handle missing: In this paper, when examining the data, no missing values were found.

– Irrelevant data: Irrelevant data, this as unwanted columns that would not be useful for the forecast models. including the ('Student_ID'), are removed.

– Outlier Detection: When examining the data, no outlier’s data was found.

– Encoding Categorical Variables: Applying label encoding techniques to transform categorical data into numerical format

4.4 Feature Selection

Finding a subset of features that can accurately describe the input data while lowering the dimensionality of the feature space and getting rid of extraneous data is the primary goal of feature selection [14]. The student dataset describes many features. When collecting data, emphasis was placed on the student performance, which helped predict the classification of student grade according to a specific set of criteria. Analyzing feature importance is crucial for predicting classification of student grade accurately. In the random forest algorithm, feature importance is evaluated by analyzing each feature's contribution value in every tree and then averaging these values across all trees. This comparison helps determine which features have the most significant impact on predicting classification of student grade. In this paper, Random Forest was employed, where feature selection naturally occurs during the construction of decision trees within the ensemble. Random Forest assesses each feature's significance according to how it affects the forest's overall performance.

4.5 Data split

In data analysis, separating data into training and test sets is a standard procedure that evaluates model performance. The paper data was separated into:

– **Training Data (80%):**

- Your model is trained using this section.
- The model gains knowledge of the data's relations, patterns, and architecture.

– **Test Data (20%):**

- This is held out during training and is only used after the model has been trained to evaluate its performance.
- An objective assessment of the model's performance on new, untested data is given by the test set.

4.6 Evaluations Metrics

Imbalanced Performance evaluation metrics including accuracy, recall, precision, and F-measure are used for evaluating classification performance. were determined by applying the guidelines in the definitions and formulas.

[15] of the following:

– **Accuracy (AC):** is the proportion of accurate predictions among all predictions [16].

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Number\ of\ Instances}$$

– **Recalling:** the number of samples assigned to a class divided by the actual number of samples in that class (equivalent to TP rate) [17].

$$Recall \quad = \frac{TruePositives}{TruePositives + falseNegatives}$$

– **Precision:** is the proportion that the model accurately predicted [18].

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

F_ Scores: This represents a categorization calculation sample that takes both recall rate and precision rate into account [19].

$$\text{F\_Scores} = 2 * \frac{(precision)(recall)}{precision + recall}$$

5. Experimental Results

The results of each algorithm Random Forest (RF), Artificial Neural Network (NN), Naïve Bayes, XGBoost, Support Vector Machine (SVM) and Decision Tree (DT) will be illustrated in the following sections.

5.1 Artificial Neural Network / Multi-Layer Perceptron (MLP)

Artificial Neural Network (NN) algorithm is applied in the experiment to measure Precision, Recall, F1_Score, and accuracy. Table (2) illustrates the result of Artificial Neural Network algorithm to products Students' grade.

Table (2): Artificial Neural Network results of student's grade

Matrices	Precision	Recall	F1_Score	Support
Excellent	0.895	0.938	0.916	535
Fair	0.918	0.922	0.920	486
Good	0.874	0.871	0.873	488
Very Good	0.923	0.874	0.897	491

The algorithm performs consistently well, with all F1_Scores at or above 0.87. The "Good" class is the weakest in terms of metrics and might benefit from more training examples or better feature engineering. The "Excellent" class has the highest recall, while "Fair" is the most balanced. If improvements are needed, the focus could be on reducing confusion between "Good" and other categories.

The following Table (3) illustrates the result of the Artificial Neural Network algorithm using only average results and measuring accuracy.

Table (3): Artificial Neural Network average results of student's grade

Matrices	Precision	Recall	F1_Score	Accuracy
Average	0.902	0.901	0.902	0.90

All metrics are around 0.90, indicating that the model performs very well overall in terms of correctly identifying and classifying the different classes. In the following figure (2) the confusion matrix of the applied Artificial Neural Network algorithm is shown.

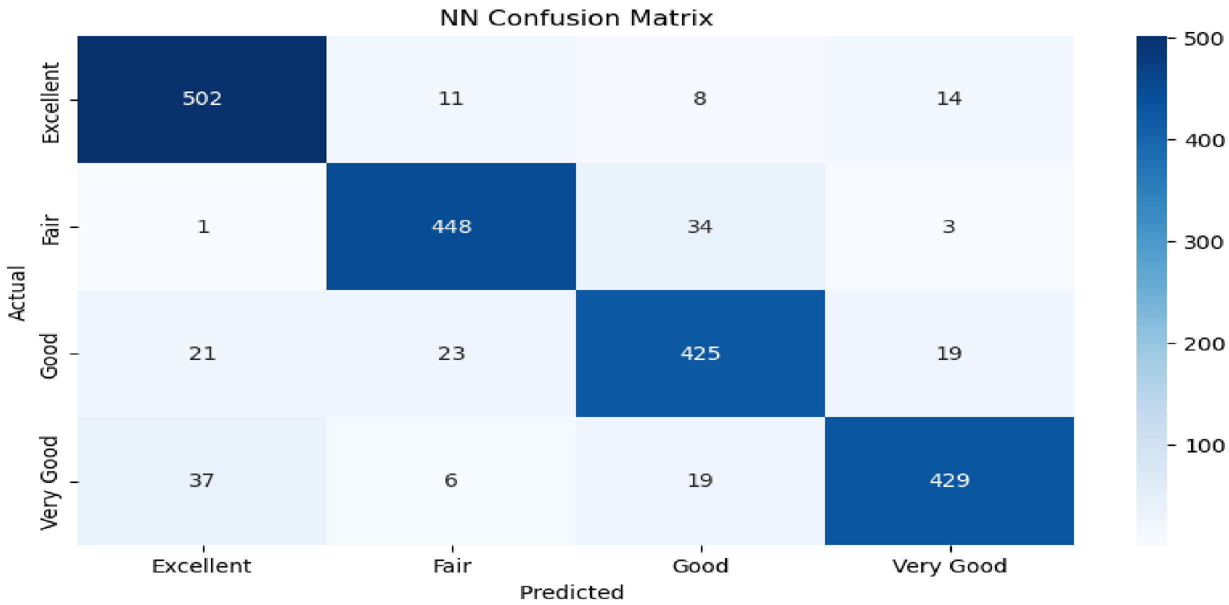

Figure (2): confusion matrix of Artificial Neural Network

The diagonal entries (502 for Excellent, 448 for Fair, 425 for Good, and 429 for Very Good) indicate the number of correct predictions for each category. The high values suggest strong performance in accurately classifying instances within these categories. The model performs well overall, particularly with the Excellent and Fair categories, which have high correct classification rates. However, there are significant misclassifications, especially for the good category, which confuses with

Fair and Very Good. The misclassifications, particularly in the good category, indicate that further optimization may be needed. Enhancing feature selection or model tuning could help improve the model's ability to distinguish between closely related categories.

5.2 Random Forest (RF)

Random Forest (RF), algorithm is used in this experiment to measure Precision, Recall, F1_Score, and accuracy. Table (4) illustrates the result of Random Forest algorithm to products Students' grade.

Table (4): Random Forest results of student's grade

Matrices	Precision	Recall	F1_Score	Support
Excellent	0.934	0.931	0.933	535
Fair	0.960	0.972	0.966	495
Good	0.931	0.935	0.933	493
Very Good	0.953	0.941	0.947	477

All classes have precision, recall, and F1_Scores above 0.93, which indicates exceptionally strong performance. The model is not only accurate but also balanced, with precision and recall closely aligned for each class.

Table (5) illustrates the result of the Random Forest algorithm using only average results and measuring accuracy.

Table (5): Random Forest average results of student's grade

Matrices	Precision	Recall	F1_Score	Accuracy
Average	0.945	0.945	0.945	0.945

An average score of 0.945 (94.5%) across all major metrics indicates that the model is performing exceptionally well. This level of performance suggests very few classification errors and a strong generalization ability on the test data.

In the following figure (3) the confusion matrix of the applied Random Forest algorithm is shown.

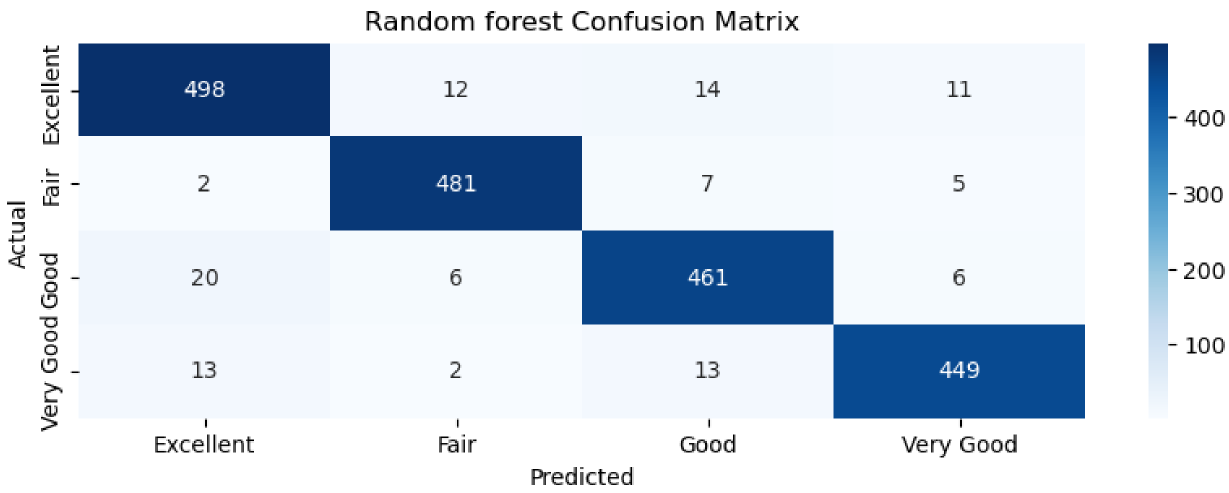

Figure (3): confusion matrix of Random Forest

The diagonal entries (498 for Excellent, 481 for Fair, 461 for Good, and 449 for Very Good) indicate the number of correct predictions for each category. The high values on the diagonal suggest that the model has effectively classified many instances correctly in each category. While the model performs well overall, the misclassifications in closely related categories (especially Fair, Good, and Very Good) suggest that additional feature engineering or model tuning might be necessary to enhance differentiation among these classes.

5.3 Support Vector Machines (SVM)

Support Vector Machines (SVM), algorithms are used in this experiment to measure Precision, Recall, F1_Score, and accuracy. Table (6) illustrates the result of Support Vector Machines algorithm to products Students' grade.

Table (6): Support Vector Machines results of student's grade

Matrices	Precision	Recall	F1_Score	Support
Excellent	0.909	0.916	0.912	535
Fair	0.914	0.944	0.929	486
Good	0.866	0.850	0.858	488
Very Good	0.883	0.864	0.873	491

The model performs very well across all classes, with F1_Scores above 0.85 for each. Minor performance drops in the Good and Very Good categories suggest potential overlaps or ambiguity in those class definitions, but overall, the classifier is reliable and effective.

Table (7) illustrates the result of the Support Vector Machines algorithm using only average results and measuring accuracy.

Table (7): Support Vector Machines average results of student's grade

Matrices	Precision	Recall	F1_Score	Accuracy
Average	0.893	0.894	0.893	0.894

With high values on all metrics, these metrics indicate that the model performs effectively, showing efficient classification with a good balance between accurately recognizing positives and reducing erroneous classifications.

In the figure (4) the confusion matrix of the applied Support Vector Machines algorithm is shown.

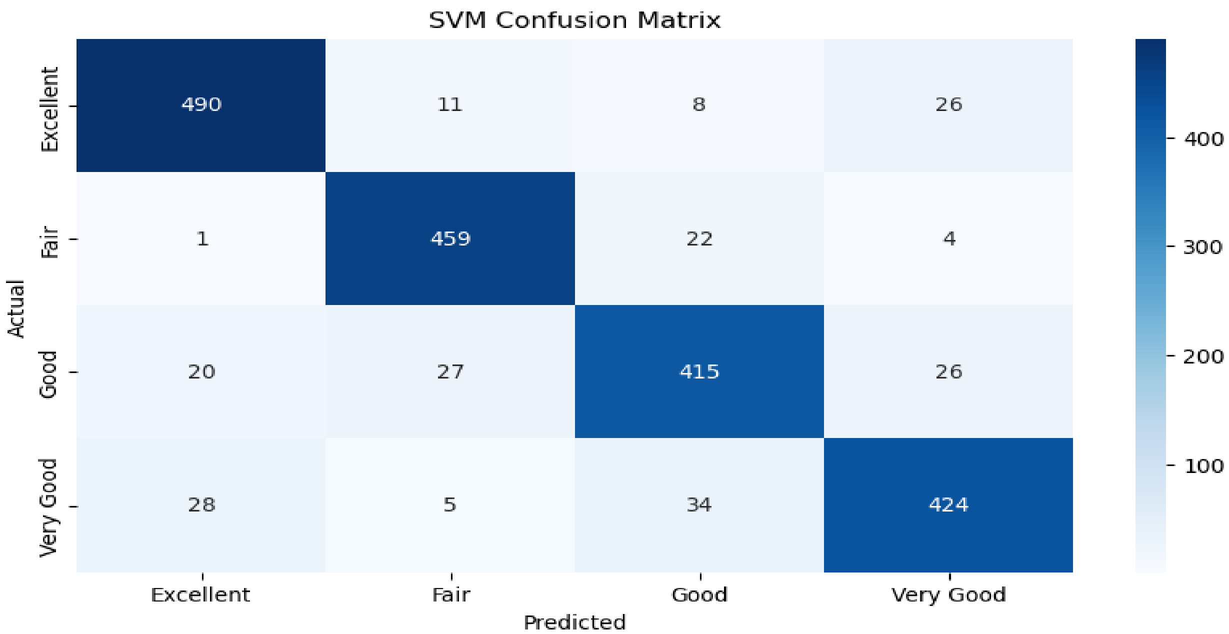

Figure (4): confusion matrix of Support Vector Machines

The diagonal values (490 for Excellent, 459 for Fair, 415 for Good, and 424 for Very Good) represent the number of correct predictions for each category. High values on the diagonal indicate strong performance in accurately classifying each category. The model shows strong performance with the majority of predictions correctly classified, particularly for the Excellent and Fair categories. However, there are notable misclassifications, especially in distinguishing between categories like Good and Fair. The model may benefit from further tuning or additional features to improve differentiation between the closely related categories, particularly Fair, Good, and Very Good.

5.4 XGBoost

XGBoost, algorithms are used in this experiment to measure Precision, Recall, F1_Score, and accuracy. Table (8) illustrates the result of XGBoost algorithm to products Students' grade.

Table (8): XGBoost results of student's grade.

Matrices	Precision	Recall	F1_Score	Support
Excellent	0.942	0.950	0.946	535
Fair	0.954	0.971	0.962	486
Good	0.936	0.936	0.936	488
Very Good	0.971	0.945	0.958	491

All classes have very high precision, recall, and F1_Scores, typically above 0.93, indicating strong and consistent model performance across all categories. The model demonstrates strong and well-balanced classification performance across all categories, with particularly high recall for Fair and high precision for Very Good.

Table (9) illustrates the result of the XGBoost algorithm using only average results and measuring accuracy.

Table (9): XGBoost average results of student's grade

Matrices	Precision	Recall	F1_Score	Accuracy
Average	0.951	0.951	0.951	0.95

These metrics reflect excellent overall model performance, with near-perfect balance between precision and recall. The consistency across all four values suggests that the model is not only accurate but also fair and robust in its predictions across different classes. There is a minimal trade-off between precision and recall, which is a strong indicator of a well-tuned classifier.

In figure (5) the confusion matrix of the applied XGBoost algorithm is shown.

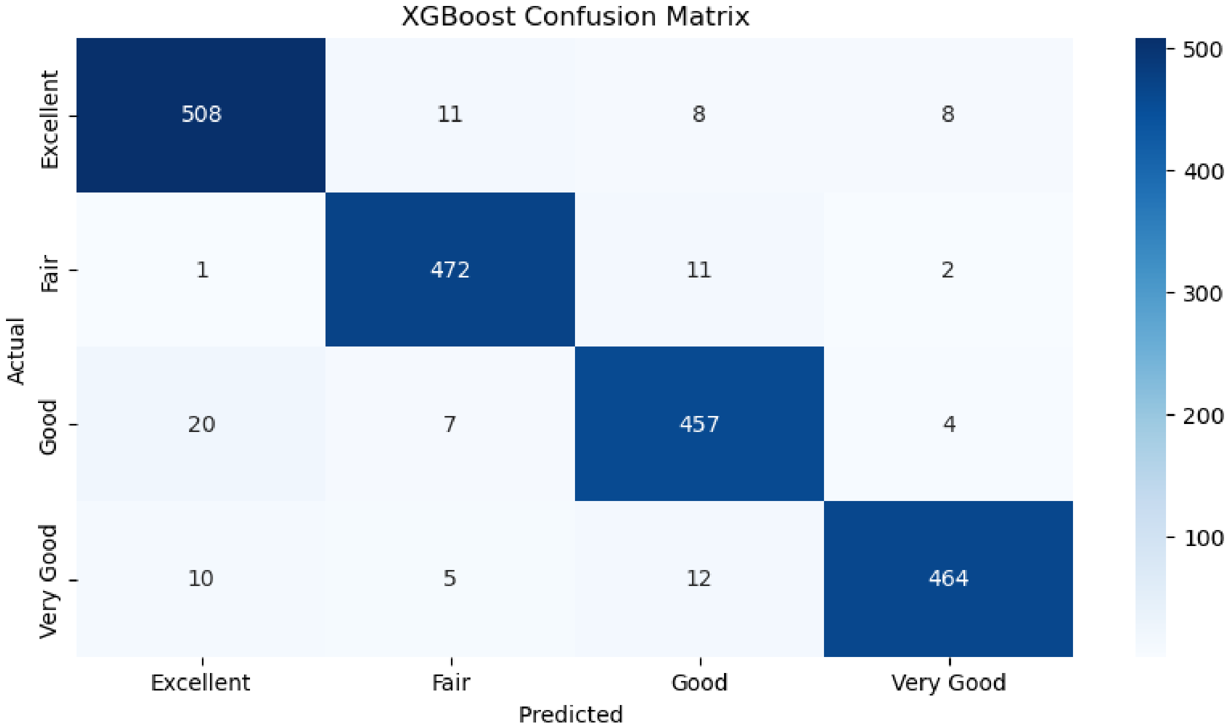

Figure (5): confusion matrix of XGBoost.

The diagonal entries (508 for Excellent, 472 for Fair, 457 for Good, and 464 for Very Good) indicate the correct predictions for each category. High values on the diagonal suggest that the model performs well in classifying instances accurately within these categories. The misclassifications, particularly in the good category, indicate that further tuning of the model or feature engineering may be needed to enhance the model's ability to differentiate between closely related classes.

5.5 Naïve Bayes

naïve bayes, algorithms are used in this experiment to measure Precision, Recall, F1_Score, and accuracy. Table (10) illustrates the result of naïve bayes algorithm to products Students' grade.

Table (10): naïve bayes results of student's grade.

Matrices	Precision	Recall	F1_Score	Support
Excellent	0.894	0.931	0.912	535
Fair	0.902	0.972	0.936	495
Good	0.904	0.876	0.890	493
Very Good	0.956	0.866	0.909	477

The model demonstrates strong performance across all categories, with particular strengths in precision and recall for the "Fair" category. However, the "Good" and

"Very Good" categories present opportunities for improvement, especially in increasing recall without compromising precision. Further analysis may involve examining the misclassifications in these categories to refine the model's predictive capabilities.

Table (11) illustrates the result of the naïve bayes algorithm using only average results and measuring accuracy.

Table (11): naïve bayes average results of student's grade

Matrices	Precision	Recall	F1_Score	Accuracy
Average	0.914	0.911	0.912	0.912

The metrics suggest that the model performs very well across the board, with high precision, recall, F1_Score, and accuracy. The metrics are closely aligned, indicating that improvements in one area do not come at the expense of others.

In the figure (6) the confusion matrix of the applied naïve bayes algorithm is shown.

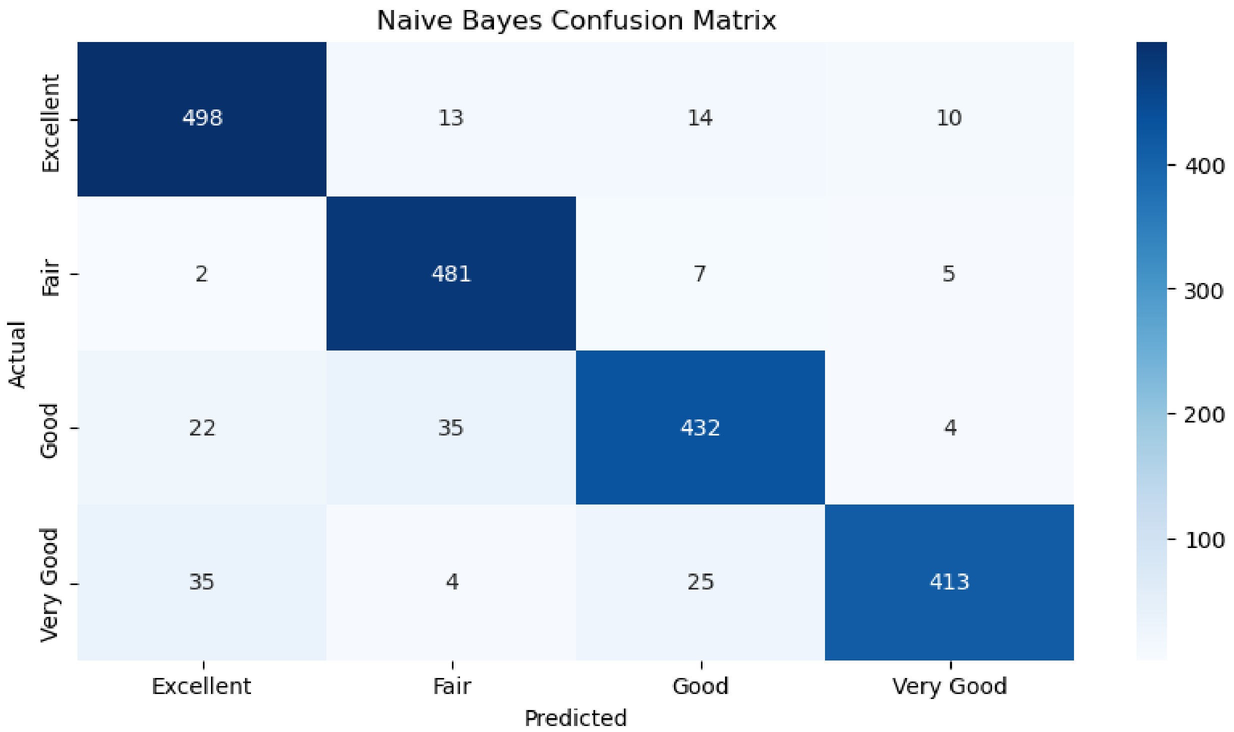

Figure (6): confusion matrix of naïve bayes.

The diagonal entries (498 for Excellent, 481 for Fair, 432 for Good, and 413 for Very Good) indicate the correct predictions for each category. High values on the diagonal suggest that the model performs well in classifying instances accurately within these categories. The misclassifications, particularly in the good category, indicate that further tuning of the model or feature engineering may be needed to enhance the model's ability to differentiate between closely related classes.

5.6 Comparison of Machine Learning Algorithms

We trained five different machine learning algorithms This as Artificial Neural Networks, XGBoost, SVM, Naive Bayes and Random Forest to see how well each model can predict student grades. Each of the algorithms were evaluated on important performance parameters This as Precision, Recall, F1_Score, and Accuracy. Table (12) summarizes the meaning values of these metrics for each algorithm and can therefore be used to give a clear comparison for predictive efficiency. Evaluation of certain models can also be made on this comparison giving more insights into their feasibilities in terms of their uses or applicability in Educational Data Mining and interventions to boost student success.

The following table (12) expresses the differences between results of machine learning algorithms.

Table (12): expresses the differences between applied algorithms.

The applied Algorithms	Precision	Recall	F1_Score	Accuracy
Artificial Neural Network	0.902	0.901	0.902	0.90
Random Forest	0.945	0.945	0.945	0.945
Support Vector Machines	0.893	0.894	0.893	0.894
XGBoost	0.951	0.951	0.951	0.95
Naive Bayes	0.914	0.911	0.912	0.912

The following figure (7) illustrates the differences between the applied Algorithms.

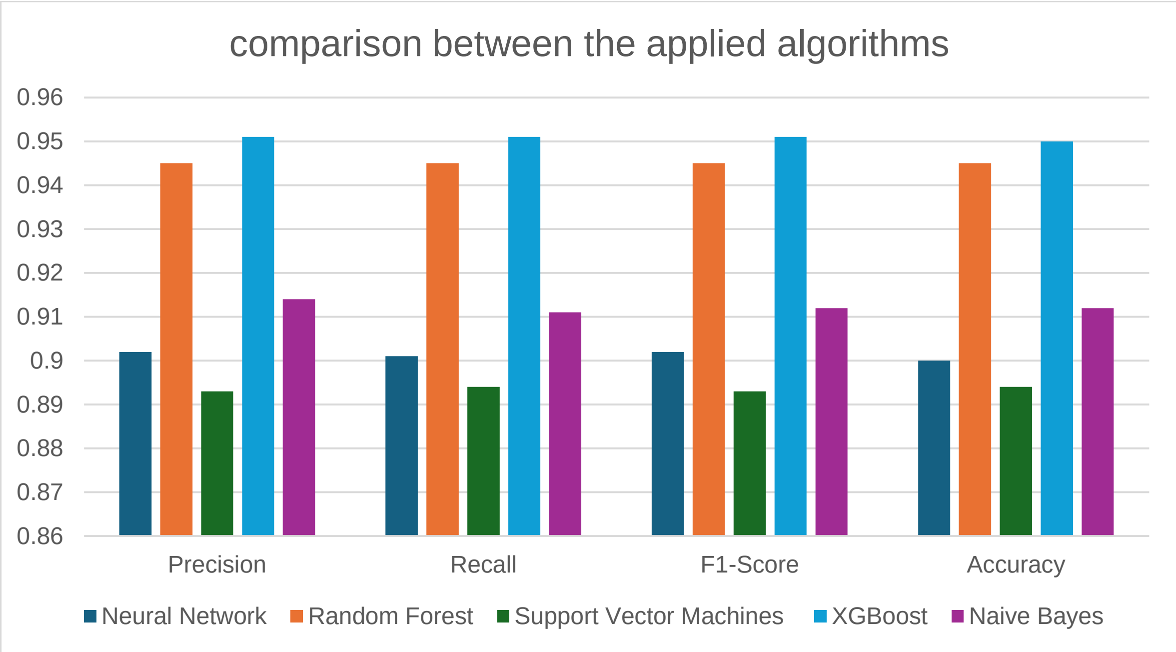

Figure (7): A comparison between algorithms.

These results illustrate that XGBoost has the best results among the others applied algorithms.

6. Conclusion

Predicting student grades using machine learning models This as Artificial Neural Network, XGBoost, Support Vector Machines (SVM), Naive Bayes, and Random Forest was the purpose of this paper. The paper has been demonstrated that each algorithm performs differently in terms of accuracy, precision, recall, and F1_Score. XGBoost outperformed the other algorithms examined overall, achieving an average precision, recall, F1_Score, and accuracy of 0.951 respectively.

It is safe to say that XGBoost is doing a marvelous job of predicting student grades for different categories. Random Forest was not far behind with an average score of 0.945, another very strong contender and a model that can handle diverse data features very well and robustly. Artificial Neural Networks scored somewhat lower on the metric compared to others with an average score of 0.902. Naive Bayes and SVM performed quite well with mean scores of 0.912 and 0.894. Even though these algorithms seem to work well, maybe their parameters should be adjusted, or we can add new features to improve their performance. Essentially, the metrics give an overall idea of how well each algorithm performed in correctly labelling students based on study-hours, online learning participation, exam scores and attendance rates — and therefore, potentially predicting if any student would pass/ fail. The major importance that machine learning approaches can have in the field of educational data mining and the development of interventions that can be targeted to promote student achievement is demonstrated by high precision and recall in the majority of the algorithms. Lastly, the study's findings demonstrate how well machine learning algorithms predict students' academic performance. Since these techniques are considerably complex, educational institutions can use them to find out a lot regarding the success rates of students and implement essential strategies to raise academic performance and help students who belong to the lists of at-risk students.

7. Future Work

– Future research should focus on optimizing the machine learning algorithms that applied in predicting student grades.

– Incorporating additional data sources can significantly enhance the predictive models.

– Exploring different features and conducting cross-institutional studies can provide valuable insights into student performance.

References

[1] M. Beseiso, "Enhancing Student Success Prediction: A Comparative Analysis of Machine Learning Technique," *TechTrends,* pp. 1-13, 2025.

[2] J. E. Alderson, N. A. Lewis, P. L. Hill and N. A. Turiano, "Sense of purpose in life predicts university performance and attrition," *Student Success,* vol. 16, no. 1, pp. 48-60, 2025.

[3] H. Luan and C.-C. Tsai, "A review of using machine learning approaches for precision education," *Educational Technology & Society,* vol. 24, no. 1, pp. 250-266, 2021.

[4] P. D. Gil, S. da Cruz Martins, S. Moro and J. M. Costa, "A data-driven approach to predict first-year students' academic success in higher education institutions," *Education and Information Technologies,* pp. 2165-2190, 2021.

[5] F. Qiu, G. Zhang, X. Sheng and P.-k. Chen, "Predicting students' performance in e-learning using learning process and behaviour data," *Scientific Reports,* p. 15, 2021.

[6] A. Daza, C. Guerra, N. Cervera and E. Burgos, "Predicting academic performance through data mining: a systematic literature," *TEM Journal,* vol. 11, no. 2, p. 11, 2022.

[7] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learning Environments,* vol. 9, no. 1, p. 11, 2022.

[8] S. Gaftandzhieva, A. Talukder, P. Theodorou and R. Doneva, "Exploring online activities to predict the final grade of student," *Mathematics,* vol. 10, no. 20, p. 20, 2022.

[9] H. T. Adu-Twum, E. A. Sarfo and T. Walugembe, "Role of Advanced Data Analytics in Higher Education: Using Machine Learning Models to Predict Student Success," *Journal of Data Science and Artificial Intelligence,* vol. 3, no. 1, p. 8, 2024.

[10] F. Kohun, B. Rassameethes, K. Pushavat and D. Lesjak, "Artificial intelligence and machine learning: usage and impacts on future higher education.," *Issues in Information Systems,* p. 19, 2024.

[11] S. Hoca and N. Dimililer, "A Machine Learning Framework for Student Retention Policy Development: A Case Study," *Applied Sciences,* vol. 15, no. 6, p. 30, 2025.

[12] H. K. Gharkan, M. J. Radif and A. H. Alsaeedi, "Analysis of AI-Empower Predictive Models for Predicting Student Performance in Higher Education," *Journal of Al-*

Qadisiyah for Computer Science and Mathematics, vol. 17, no. 1, p. 19, 2025.

[13] Natarajan, Jiuyong and Koronios, Andy, "Data mining techniques for data cleaning," in *Engineering Asset Lifecycle Management: Proceedings of the 4th World Congress on Engineering Asset Management (WCEAM 2009), 28-30 September 2009*, 2010.

[14] Maldonado, Riff and Neveu, "review of recent approaches on wrapper feature selection for intrusion detection," *Expert Systems with Applications,* vol. 198, 2022.

[15] W. Hersh, "Information retrieval," in *Biomedical Informatics*, vol. 18, Springer, 2021, pp. 755-794.

[16] G. Zeng, "On the confusion matrix in credit scoring and its analytical properties," *Communications in Statistics - Theory and Methods,* p. 15, 2019.

[17] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing and Management,* vol. 45, p. 11, 2009.

[18] D. Gray, D. Bowes, N. Davey, Y. Sun and B. Christianson, "Further Thoughts on Precision," in *15th Annual Conference on Evaluation & Assessment in Software Engineering (EASE 2011)*, 2011.

[19] C. Ferri, J. Hernández-Orallo and M. R, "An experimental comparison of performance measures for classification," *Pattern Recognition Letters,* p. 12, 2009.