



# Prediction of student exam performance using data mining classification algorithms

Dalia Khairy<sup>1</sup> · Nouf Alharbi<sup>2</sup> · Mohamed A. Amasha<sup>1</sup>  · Marwa F. Areed<sup>4</sup> · Salem Alkhalaf<sup>3</sup> · Rania A. Abougalala<sup>1</sup>

Received: 25 July 2023 / Accepted: 6 March 2024  
© The Author(s) 2024

## Abstract

Student outcomes are of great importance in higher education institutions. Accreditation bodies focus on them as an indicator to measure the performance and effectiveness of the institution. Forecasting students' academic performance is crucial for every educational establishment seeking to enhance performance and perseverance of its students and reduce the failure rate in the future. The main goal of this study is to predict the performance of undergraduate first-level students in the Computer Department during the years 2016 to 2021 to enhance their performance in future by discovering the best algorithm use to analyze the educational data to identify the students' academic performance. The secondary data was collected by reviewing the Student Affairs Department at the Faculty of Specific Education at Damietta University, in addition to the Statistics Department at the university. The dataset contained 830 instances after excluding 139 instances of missing values, irrelevant rows, and outliers. The dataset was divided into train (577 instances (70%)), test (253 instances (30%)) and involved six features such year, midterm, practical exam, writing exam, final total degree, and grade. This paper use five machine learning (ML) algorithms which was selected according to the literature review and high accuracy in predicting educational data mining: For the purpose of comparison, a number of different machine learning algorithms, such as Random Forest, Decision Tree, Naive Bayes, Neural Network, and K-Nearest Neighbours, were utilized and evaluated with evaluation metrics such as confusion matrix, accuracy, precision, recall, and F-measure. The Random Forest and Decision Tree classifiers emerged as the top-performing algorithms, accurately categorizing 250 instances when predicting students' performance in the statistics course. This was determined based on the findings of the study. Out of a total of 253 instances that were included in the testing set, they only made three incorrect classifications.

**Keywords** Data mining · Prediction of student performance · Machine learning algorithms

---

Extended author information available on the last page of the article

## 1 Introduction

The utilization of data mining techniques in the field of education has garnered significant interest in recent years. Data Mining (DM) involves uncovering new and valuable information or significant findings from large datasets (Witten et al., 2011). It also seeks to extract new trends and patterns from extensive datasets through the utilization of various classification algorithms (Baker et al., 2016). The Egyptian government launched the National Strategy for Higher Education (Egypt's Vision 2030), which includes three main axes and seven principles and stresses the importance of investing in the human element and universities seek help raise students' academic performance and achieve this vision by having an efficient graduate who can face labor market requirements. Decision makers and stakeholders expect students to graduate with high grades and outstanding distinction to achieve academic performance that meets their vision and helps fulfill good economic growth. Although students make a lot of effort in studying at the university, they may succeed or fail due to many economic, social, and psychological factors. Many previous Studies on student failure in computer science have been undertaken, mathematics, and physics courses, as well as on predicting student performance in mathematics at the pre-university level. Five data mining classification algorithms have been chosen to predict students' performance and the likelihood of passing based on their high accuracy in educational data mining. These algorithms include Random Forest, Decision Tree, Naive Bayes, Neural Network, and K-Nearest Neighbors. Various evaluation metrics were utilized for assessing the algorithms, including accuracy, preciseness, recall, matrix of confusion, and F-measure. This study aims to address the rising failure rate in statistics courses and the growing enrollment of students in supplementary exams or summer semesters, which leads to financial waste for parents and the government. The goal is to identify the most effective algorithm for predicting students' performance in statistics courses, in order to prevent future issues and strive for optimal outcomes for students.

## 2 Literature review

DM techniques have proven their superiority in different sectors, such as e-commerce and business, and recently their usage in the field of education has been rapidly growing. This section analyzes the efficacy of educational data mining (EDM) techniques. Various studies are examined. The articles under review are categorized according to the various types of DM algorithms used to forecast the ultimate results, namely: "decision tree (DT), Artificial Neural Network (ANN), Naive Bayes (NB), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest (RF)."

### 2.1 Decision Tree (DT)

Defined as a tree-like graph constructed established upon a series of conditions, Decision Tree (DT) takes a specific feature as input and produces class labels as output (Tomasevic et al., 2020). The student's future performance was forecasted

through Decision Tree analysis, which involved categorizing the student's scores in past quizzes (Adebayo & Chaubey, 2019). "A different research project examined how the behavioral characteristics of students impact an e-learning system, with the classification process being conducted by a DT classifier (Ajibade et al., 2022). Several studies indicate that utilizing the personal, social, economic, and cognitive traits of students can be used to forecast their exam performance (Aman et al., 2019; Zhao et al., 2020a, 2020b). Nevertheless, the utilization of numerous arbitrary features could potentially hinder the classification accuracy. Dataset preprocessing, along with suitable feature selection, plays a crucial role in improving prediction results, as highlighted by Al-Obeidat et al. (2018) and Wong and Senthil (2019). A refined version of the DT algorithm is presented, comprising two primary phases: an entropy-based feature selection stage and the construction of a prediction model (Patil et al., 2018; Santoso, 2020)."

## 2.2 Naive Bayes (NB)

This is a popular technique that uses the Bayesian theorem (Tomasevic et al., 2020). Academic performance is built using demographic data forecasting model. When comparing the accuracy of the NB classifier with KNN, it is evident that the NB classifier achieves a notably high accuracy of 93.6% (Amra & Maghari, 2017). Furthermore, the NB algorithm is utilized for forecasting students' final exam scores based on various relevant features, including assignment scores, lab assessments, previous exams, and course attendance. The assessment demonstrates that the performance of NB surpasses that of SVM in terms of accuracy, with respective scores of 92% and 63.5% (Kaur & Bathla, 2018). Personality traits, like time management, stress control, and concentration, significantly influence the ability to predict one's performance in upcoming exams. Combining cognitive and non-cognitive features enhances the performance of the Bayesian prediction model, as demonstrated by Sultana et al. (2017). Moreover, the Forward Selection approach (Saifudin & Desyani, 2020) and Wrapper (Usman et al., 2020) feature selection algorithms are employed alongside the NB model to enhance the accuracy of the prediction model.

## 2.3 Artificial Neural Network (ANN)

Within the field of DM, a prevalent classification algorithm is known as an Artificial Neural Network (ANN). The input layer, hidden layer, and output layer constitute the three layers of a biological neural network (Amazona & Hernandez, 2019). Lau et al. (2019) employed traditional statistical methods to identify the key factors impacting students' academic performance. Subsequently, the ANN model was developed using 11 input variables, incorporating two hidden layers of neurons, and concluding with one output layer. Analyzed the model's performance using various metrics including the ROC curve, confusion matrix, error performance, regression, and error histogram. Overall, the prediction model achieved a satisfactory accuracy rate of 84.8%. Another study examined the impact of input factors on the ability to predict output classifications. The study showed that the most effective

input variables for predicting students' performance using an Artificial Neural Network model were the students' attendance and study time (Aydoğdu, 2020). Various supervised learning algorithms were compared, along with different attributes of students (Tomasevic et al., 2020).

## 2.4 Support Vector Machine (SVM)

By employing a model-based approach, Support Vector Machine (SVM) separates the dataset into several classifications. It creates a hyperplane between two separate classes and depicts data points in a 2D or 3D space. (Sen et al., 2020). To anticipate student performance as early as possible, an ensemble model integrates the findings of various DM approaches to increase the precision of prediction (Gil et al., 2021). SVM, NB, and DT are combined in this hybrid technique to enhance the prediction outputs. It is demonstrated that the ensemble model has a 98.5% accuracy rate. A deep neural network prediction model combining CNN, LSTM, and SVM models was created by Wu et al. in 2019. It was demonstrated that the hybrid model could more accurately predict the results as F-measure = 95.03%, while SVM achieved an F-measure = 92.48%. Correlation-based filtering was employed by Zafar et al. (2020) to recognize the most prominent features of the prediction process. The F-measure for the features-based SVM model was 90%. To investigate the relationship between students' social interactions and their English test results, a Principal Component Analysis (PCA) is introduced. (Zhao et al., 2020a, 2020b).

## 2.5 K-Nearest Neighbor (KNN)

Analogous to this approach is the K-Nearest Neighbor (KNN) method, which organizes data according to common attributes. Within this approach, the value "K" represents the quantity of nearest neighbors chosen to classify an ambiguous object. Objects with comparable characteristics are expected to be grouped into the same category (Sen et al., 2020). "Five data mining techniques are being analyzed to develop the most effective prediction model for students' test scores: Naive Bayes, Decision Trees, K-Nearest Neighbors, Artificial Neural Networks, and Support Vector Machines. The KNN model achieved a superior level of accuracy at 100%, outperforming all other classification models (Vital et al., 2021). In order to decrease the time it takes to process the model while maintaining prediction accuracy, a rapid KNN method was proposed. When comparing the suggested model to the traditional KNN model, there was an enhancement in accuracy, achieving a rate of 96.6%. Moreover, the proposed model decreases processing time by 90% according to Ahmed et al. (2020)."

## 2.6 Random Forest (RF)

Multiple DTs make up the ensemble ML method known as Random Forest (RF). The class that receives the most votes at the end is chosen as the anticipated class (Bruce, 2019). Using lecture views, resource access, and test results, RF is utilized to predict

performance. The suggested model demonstrated that LMS interactions and grades of students may be utilized to predict performance with 84% accuracy (Wakelam et al., 2020). For segmenting students into pass/fail segments, three data mining algorithms were examined. In comparison to KNN and NB, RF outperformed with a 95.45% accuracy rate (Lenin & Chandrasekaran, 2019). The studies demonstrate that not all attributes are included in the prediction process but using an unimportant feature may have a detrimental impact on the outcome of the prediction (Nuankaew & Thongkam, 2020).

All explored literature and previous related work clearly establishes a picture that the DM techniques can be effectively utilized to predict academic performance. To this extent, the proposed approach aims to predict the performance of undergraduate first-level students in the Computer Department at Damietta University during the years 2016 to 2021 in order to improve their academic performance in future. This paper use five different machine learning (ML) algorithms which was selected according to superiority and accuracy in predicting educational data mining tasks: RF, DT, NB, ANN and KNN.

### 3 Method

#### 3.1 Dataset descriptions

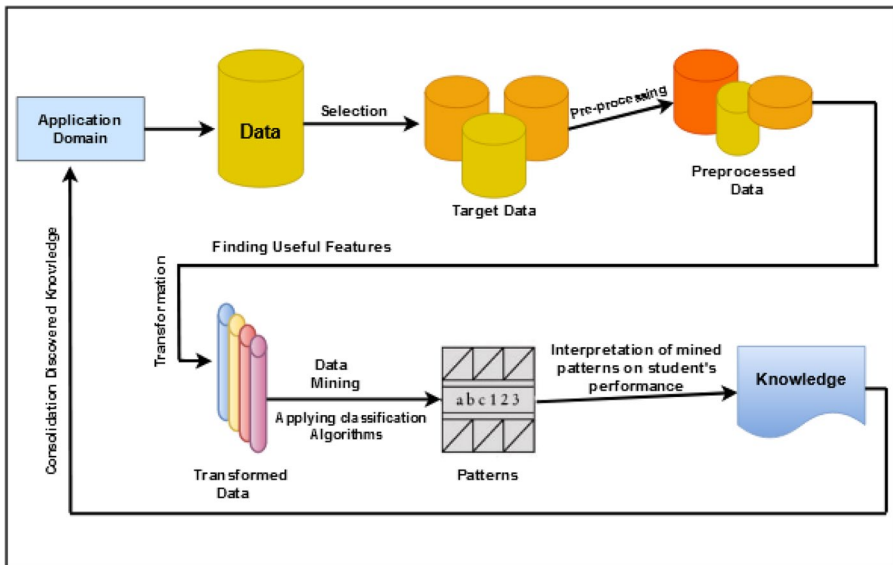
Students' results over the years were analyzed as indicated in Table 2, for the academic years 2015/2016 to 2020/2021, 196 students (23.61%) had a pass class during the first semester of the first year, while 325 students (39.16%) had a failing class. The data for grades at the ordinary level are shown in Fig. 6. In terms of grades, 325 students (39.16%) received a F from the 2015–2016 academic year to the 2020–2021 academic year, 196 students (23.61%) received a D, 104 students (12.53%) received a C, 89 students (10.72%) received a B, and 116 students (13.98%) received an A.

Figure 1 illustrates the block diagram used in this study to predict student statistics correction exam performance from input data to output information. To ensure that only the relevant information was carried out and combined to create a dataset, data collected from the field was collected and cleaned before being preprocessed to eliminate outliers. The data was then translated into a format that ML algorithms may use after the useful features have been found. The models' outputs were then gathered and assessed for knowledge discovery. As a result, useful information was produced.

The information that was learned about the performance of students was merged through the processes of data scrubbing and preliminary processing, selecting data and integrating, transforming the data, data mining, an assessment of the knowledge that was learned about students' performance, and the collection of the output data.

#### 3.2 Data cleaning and preprocessing

"Data cleaning is the process of removing or changing information that is incorrect, incomplete, irrelevant, duplicated, or incorrectly formatted in order to get the data ready for use. This is done to get the data ready for use. the data for



**Fig. 1** A diagram for the prediction of student statistics correction exam performance

the analysis. The preprocessing of data is yet another essential step in the development of deep machine learning algorithms. The results are improved, and the amount of noise is decreased. The following section provides a description of the processes that are involved in the preparation of data purification using MATLAB (<https://www.mathworks.com>). A CSV or Excel file is used to present the sample data, which consists of many columns, many rows, and the presence of some values that are missing. Cleaning the data is the most important step in the process of developing a data culture and using it to generate accurate forecasts. Correction of grammatical and syntactical errors, standardization of data sets, correction of errors such as empty fields, identification of duplicate data points, and scaling of features are all associated with this process. When preparing data, it is easier to work with if you are aware of what kinds of things to look for. In order to clean the data, different procedures are utilized depending on the type of data that is being used; however, the steps that are included in the preparation of the data are always the same."

### 3.2.1 Remove duplicate observations

The process of collecting data most frequently results in data duplication. This mostly occurs when merging information from many sources, including information obtained from clients or other departments. It is necessary to get rid of every duplicated instance of data and eliminate pointless observations from the dataset.

### 3.2.2 Filter unwanted outliers

Unusual values make up outlier values in a dataset. These contradict assumptions and are quite dissimilar from other data points, which can skew the study. The method of deletion is randomly done and depends on the data to be examined, and its performance is often improved by removing undesired outliers: firstly, removing an outlier if the data is known to be inaccurate or if we are confident that the data values fall within a certain range and safely discarding results that are outside of that range; secondly, losing a suspect outlier if there are many data points in the sample.

### 3.2.3 Fix structural errors

Structural errors can be weird naming conventions, typos, case sensitivity, and so on. Inconsistencies cause categories to be mislabeled. This is best demonstrated by providing “N/A” and “Not Applicable” together. Both appear in various categories; however, they should be considered a part of the same category for analysis.

### 3.2.4 Fix missing data

Many algorithms do not accept missing values. so, observations with missing values can be removed or filled by values based on other observations.

### 3.2.5 Feature scaling

This is important because datasets have different types of characteristics or variables, for example, location, age, salary, and years of experience (Onawumi et al., 2023; Matsson & De Geer, 2023). If there are similar values, data can be standardized. Otherwise, aim for normalization. However, this is not a requirement, and any method can be choosed. There are two ways to do the feature scaling (Fig. 2).

- Standardization

Often referred to as Z-score normalization, it is sometimes a method of rescaling values like normalization while adhering to the properties of the standard normal distribution. Standardization is very important as it enables reliable data transmission between different systems. Standardization facilitates data exchange and communication between computers. In addition, standardization makes data easier to process, analyze, and store in databases. This method allows businesses to use data to make better decisions. Organizations can compare and analyze their data more easily when standardized, gaining insight into how to run their business better. The method transforms features from the range of -1 to +1:

$$x_{transformed} = \frac{x - mean(x)}{stander\ deviation(x)} \quad (1)$$

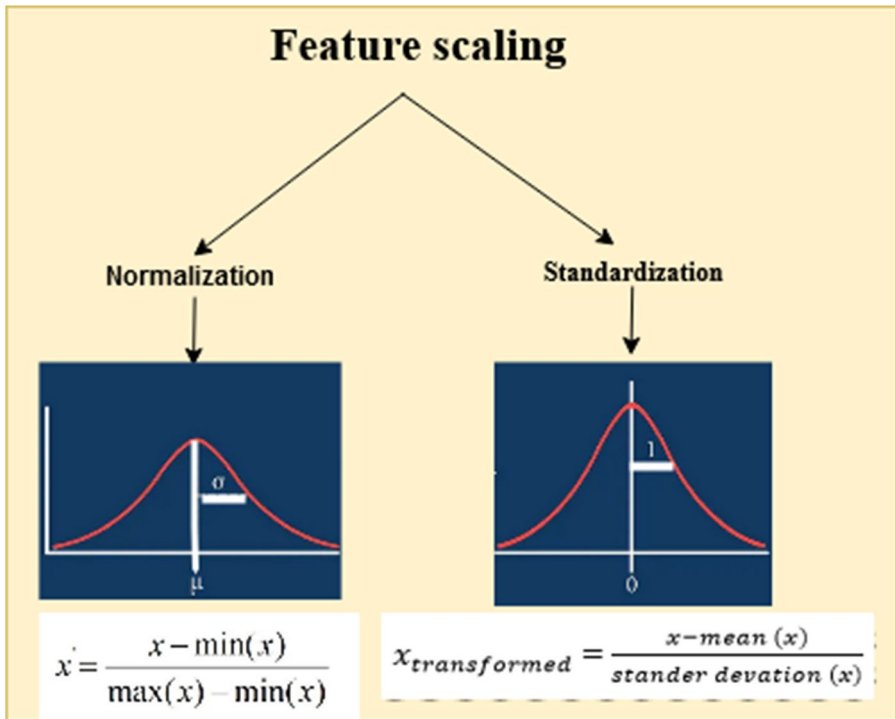


Fig. 2 Normalization and standardization

- Normalization

Normalization is one of the most used techniques for data preparation. This enables to convert the values of data set's numerical columns to a standard scale. Normalization is a way of organizing data in a database. This is a scaling method that reduces duplication by scaling and shifting numbers between 0 and 1. If there are no outliers because they cannot be handled, use normalization to remove unwanted features from the dataset. Understanding the normalization formula will help to decide if this is the best way to process dataset. It transforms features from the range of 0 to +1:

$$x_{transformed} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2)$$

The  $z$  score informs us of the value's standard deviation from the mean. Depending on whether the result is above or below the mean, it is either positive or negative:

$$z_{score} = \frac{\text{value}(x) - \text{mean}(x)}{\text{standard deviation}(sd)} \quad (3)$$

After understanding, cleaning, and preparing the data, for performance prediction, it's crucial to determine the predictor (input) variables and the target variables (output). before applying ML algorithms to training datasets.



## 4 Feature selection

One of the goals of the feature selection methods that are utilized in machine learning (Fig. 3) is to identify the most suitable collection of characteristics that will allow for the development of optimized models of the phenomenon that is being investigated. According to Ghosh et al., 2023, the following categories can be utilized to facilitate the broad classification of feature selection strategies in machine learning: When it comes to selecting features for a labeled dataset, it is possible to employ techniques for supervised feature selection that consider the target variable. On the other hand, techniques for unsupervised feature selection can be utilized for datasets that do not have labels when the target variable is not included.

It is crucial for machine learning engineers to understand which feature selection technique will best suit their model. The selection of the right statistical measure for feature selection becomes simpler as our knowledge of the data kinds of variables increases.

Firstly, it's necessary to determine the type of input and output variables in order to choose the appropriate feature selection algorithm (Fig. 4). The two main types of variables in ML are:

- Numerical variables are those that have continuous values, such floats and integers.
- Categorical variables are those that have categorical values, including nominal, ordinal, and Boolean variables.

Table 1 summarizes the cases with appropriate measures for feature selection (Rizvi, 2018).

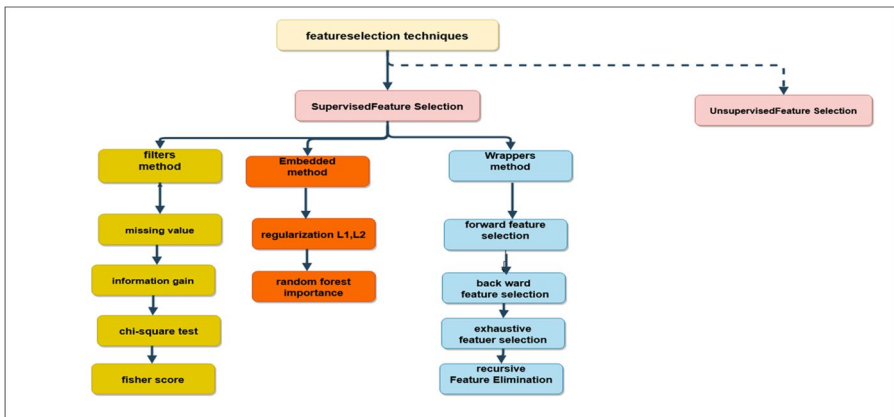
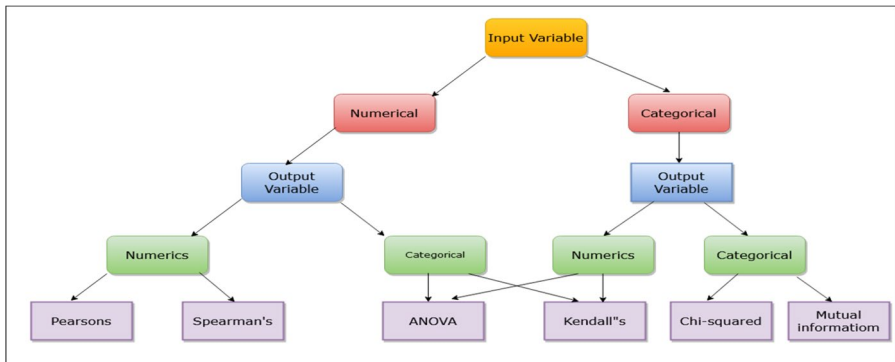


Fig. 3 Feature selection techniques in machine learning. (<https://www.javatpoint.com/feature-selection-techniques-in-machine-learning>)



**Fig. 4** Feature selection measures algorithm in machine learning.( <https://www.javatpoint.com/feature-selection-techniques-in-machine-learning>)

## 5 ML algorithms training

Since a successful prediction model should not have any missing values in the dataset, the current study used techniques like deleting the entire row containing the missing values to address the issue of missing values. Because some of the students' fields had no values in them, the approach was used. In order to prevent biased models and incorrect predictions or classifications, it was imperative to address them. Out of the 969 occurrences in this study, 139 were eliminated because some of the students had delayed their coursework or exams. Thus, 830 occurrences were needed to create a dataset.

The used dataset is split into 70% for machine learning algorithm training, 15% for trained ML algorithm validation, and 15% for trained ML algorithm testing. "Following a review of the study for lowering dropout rates using ML techniques, the dataset's division was taken into consideration (Mduma et al., 2019). The performance of pupils has been predicted using a variety of EDM techniques, including SVM, K-NN, ANN, and DT (Roy & Garg, 2017; Shahiri & Husain, 2015). K-NN, SVC, RF, DT, and MLP are some of the machine learning algorithms that will be utilized in this study to predict the performance of students in statistical courses. These algorithms are widely utilized in EDM and have proven to be effective. The data for this study will be obtained from the Computer Teacher Preparation Department at Damietta University through the faculty of Specific Education."

**Table 1** The cases with appropriate measures for feature selection

Input Variable	Output Variable	Feature Selection technique
Numerical	Numerical	<ul style="list-style-type: none"> <li>• Pearson's correlation coefficient (for linear correlation)</li> <li>• Spearman's rank coefficient (for non-linear correlation)</li> </ul>
Numerical	Categorical	<ul style="list-style-type: none"> <li>• ANOVA correlation coefficient (for linear correlation)</li> <li>• Kendall's rank coefficient (for non-linear correlation)</li> </ul>
Categorical	Numerical	<ul style="list-style-type: none"> <li>• Kendall's rank coefficient (linear)</li> <li>• ANOVA correlation coefficient (non-linear correlation)</li> </ul>
Categorical	Categorical	<ul style="list-style-type: none"> <li>• Chi-squared test (contingency tables)</li> <li>• Mutual information</li> </ul>

## 6 Evaluation of trained ML algorithms

"When it comes to machine learning, one of the most important problems to solve is figuring out how to calculate the future value of educational data. This problem involves determining how accurately they anticipate the desired outcome. As a result of the application of machine learning classification algorithms, significant findings and forecasts have been generated (Tohka & Van Gils, 2021). It is possible to evaluate the execution of classifiers using a variety of different methods." When it comes to modifying and assigning novel cases to classes in actual usage, all these techniques are associated with the number of times the classifier was either "true" or "false." Regardless, different approaches offer a variety of perspectives on what we mean when we say "true" or "false," and not all errors are of the same significance. Consequently, we have a wide range of different implementation strategies to choose from (Tohka & Van Gils, 2021). As was mentioned earlier, this section provides the metrics that are used to estimate the implementation of the classification technique:

- **Multiclass Confusion Matrix** The measures designed for binary classification do not fully apply in the situation of multiclass classification. The dimension of the multiclass confusion matrix is  $N \times N$ , where  $N$  is the total number of distinct class labels (e.g., 11 for NPS) that there are. Thus, in this instance, the characterization of TP, TN, FP, and FN cases is not applicable. On the basis of the characterization, it is possible to do an analysis that focuses on a certain class instead. Using this method, a collection of metrics can be defined for every class. Then, it is possible to produce measurements for the complete confusion matrix based on the appropriate mix of these metrics. As follows, gives a summary of the metrics—accuracy, recall, precision, and F1-score, in particular—defined for a multiclass confusion matrix (Markoulidakis et al., 2021).
- **Precision (P)** is estimated by the following equation to be the number of precise classes produced by the classification measure. (Hossin & Sulaiman, 2015):

$$PPV(C_i) = \frac{TP(C_i)}{TP(C_i) + FP(C_i)} \quad (4)$$

- **Recall (R)** is the measure is the number of accurate positives divided by the sum of the number of correct positives and the number of incorrect negatives and is used to find all acceptable cases in a dataset. (Hussain et al., 2022):

$$TPR(C_i) = \frac{TP(C_i)}{TP(C_i) + FN(C_i)} \quad (5)$$

- **F-measure** is the harmonic mean of precision and recall and can be determined by using the following equation (Hossin & Sulaiman, 2015):

$$F1(C1) = 2 * \frac{TPR(C_i) * PPV(C_i)}{TPR(C_i) + PPV(C_i)} \quad (6)$$

- **Accuracy (A)** is the most used evaluation measure in practice for either binary or multi-class classification problems. Accuracy determines the quality of the produced solution

estimated based on the percentage of true predictions over total examples (Muntean & Militaru, 2023). Accuracy is the number of true forecasts made as a ratio of all predictions made and is determined by the following equation (Hussain et al., 2022):

$$\text{Acc}(\text{A}_{\text{reduced}}) = \frac{\sum_{i=1}^N \text{TPR}(C_i)}{\sum_{i=1}^N \sum_{j=1}^N C_{ij}} \quad (7)$$

- **Classification report** is used to estimate the rate of forecasts of a classification algorithm; it includes these measurements: precision, recall, and F1-score (Muntean & Militaru, 2023).

## 7 Results

### 7.1 Dataset

The survey on the results of students in the Computer Department during the years 2016 to 2021 revealed a high percentage of failure in the statistics course for first-level students. The highest failure rate was in 2016 when it reached 48.06%, and the failure rate for other years ranged between 22 and 48%. Figure 5 shows the percentage of students failing the statistics course during the years 2016 to 2021.

The categorical frequency distribution of the remark's variable was used to examine the dataset. For the academic years 2015/2016 to 2020/2021, during the first semester of the first year, it was noted that 196 students (23.61%) had a pass class, while the failing class had 325 students (39.16%) as shown in Table 2. Figure 6 illustrates the statistics of ordinary level grades. From the 2015–2016

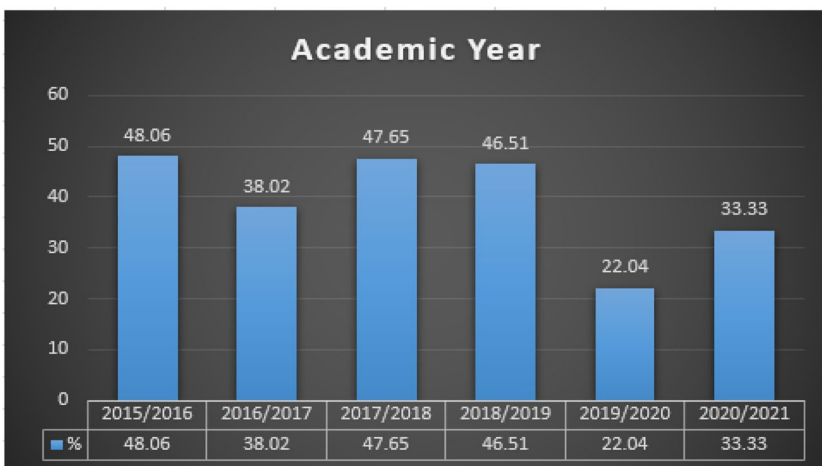


Fig. 5 The % failing the statistics course during the years 2016 to 2020

**Table 2** A comparison of the numbers of successful students

Acad. year	Total	Pass	%	Fail	%	A	%	B	%	C	%	D	%	F	%
2016	129	29	22.3	62	48.06	8	6.2	16	12.4	14	10.85	29	22.3	62	48.06
2017	192	61	31.8	73	38.02	12	6.25	21	10.94	25	13.02	61	31.8	73	38.02
2018	170	36	21.2	81	47.65	19	11.77	11	6.47	23	13.53	36	21.2	81	47.65
2019	129	20	15.5	60	46.51	17	13.18	13	10.08	19	14.73	20	15.5	60	46.51
2020	186	43	23.1	41	22.04	56	30.11	26	13.98	20	10.75	43	23.1	41	22.04
2021	24	7	29.2	8	33.33	4	16.67	2	8.3	3	12.5	7	29.2	8	33.33
Total	830	196	23.61	325	39.16	116	13.98	89	10.72	104	12.53	196	23.61	325	39.16

academic year to the 2020–2021 academic year, there were 325 students (39.16%) with a F grade, 196 students (23.61%) with a D grade, 104 students (12.53%) with a C grade, 89 students (10.72%) with a B grade, and 116 students (13.98%) with an A grade.

Figure 6 shows a comparison of the numbers of successful students versus those who failed in the statistics course during the years 2016 to 2021. Although the number of successful students is greater than the number of failed students, the number of failed students represents a high percentage. Figure 7 also shows that the highest failure rate was in 2016 (48.06%).

## 7.2 Evaluation of the best selected trained ML algorithms

The dataset contains 830 instances after excluding 139 instances of missing values, irrelevant rows, and outliers after data collection. It was divided into train (577 instances (70%)) and test (253 instances (30%)). Besides, the dataset involved six features as depicted in Table 3 such as year, midterm, practical exam, writing exam, final total degree, and grade. Table 3 explains the features, their descriptions, and possible values.

The grades in the dataset are broken down into five distinct categories: A, B, C, D, and F. Figure 8 offers a visual representation of the number of students who fall into each category over the course of the academic years (2016–2021).

This section makes use of machine learning algorithms to categorize the grades that students received in the statistics course during the academic years 2016–2021. Table 3 and Fig. 8 both contain a presentation of the different categorizations of class grades. There were a total of 830 instances included in the dataset, which were split into a train set (70%) and a test set (30%). A comparison of machine learning algorithms (RF, DT, NB, NN, and KNN) was

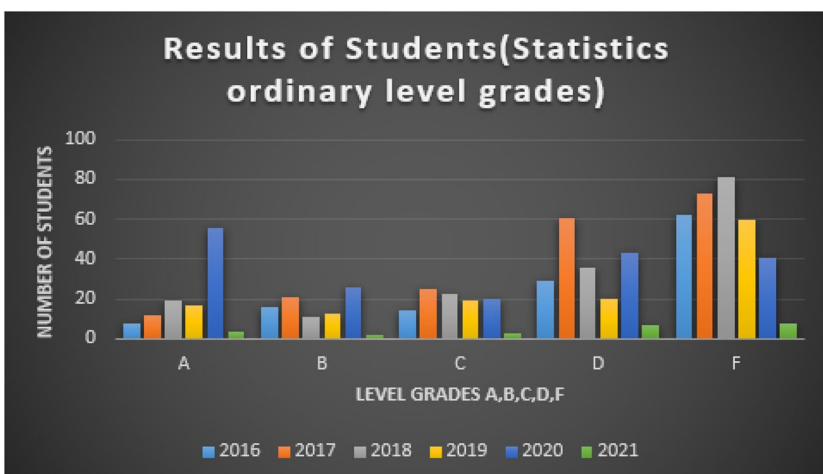


Fig. 6 Results of students (statistics ordinary level grades A, B, C, D, F)

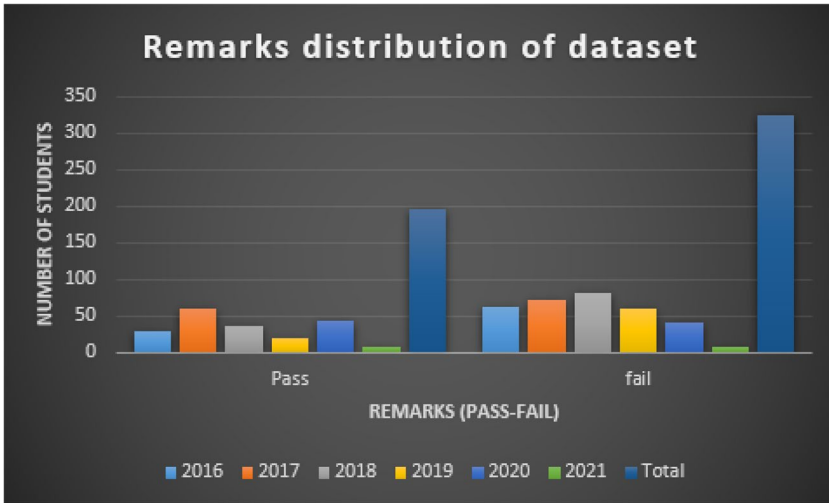


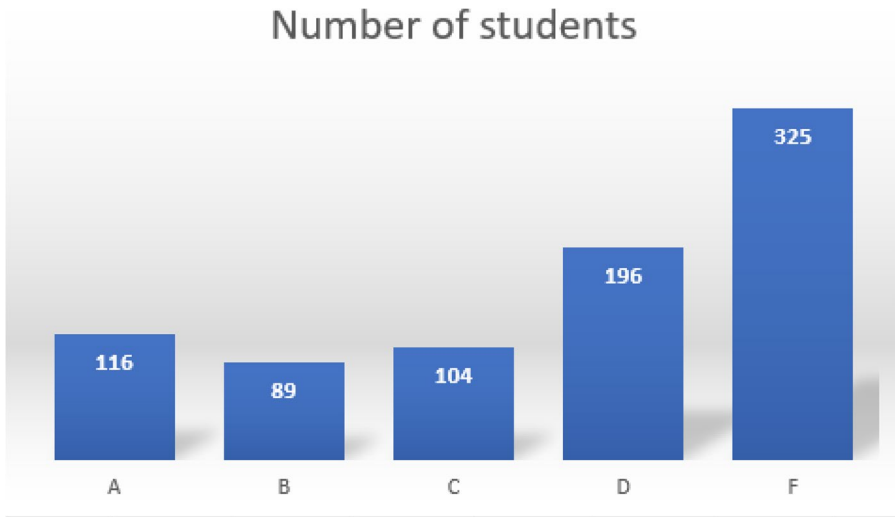
Fig. 7 Remarks distribution in dataset

carried out using the evaluation metrics. To determine the accuracy, precision, recall, and F-measure for each classifier, the values of the confusion matrix were used to calculate the TP, TN, FP, and FN values. Figure 9 illustrates these values.

Table 4 presents a list of the evaluation metrics, such as the number of each correctly and incorrectly classified instance, the accuracy of each classifier, precision, recall, and F-measure.

Table 3 List of the features of the dataset

No	Feature	Description	Value
1	Year	Academic year	From 2016 to 2021
2	Midterm	The degree of the exams through the academic year	Quantitative data (0–20)
3	Practical exam	The degree of the practical exams at the end of the term and a preface to the final exam	Quantitative data (0–20)
4	Writing exam	The degree of the final exam as a writing exam that was held at the end of studying the course	Quantitative data (0–60)
5	Final total degree	The sum of the degrees of midterm, practical exam, and writing exam	Quantitative data (0–100)
6	Grade	Grades obtained from the final total degree A: degree range [85–100] B: degree range [75–84] C: degree range [65–74] D: degree range [50–64] F: degree range [0–49]	Qualitative data (A, B, C, D, E, or F)



**Fig. 8** Number of students in each category of grade (A, B, C, D, F) through academic years (2016–2021)



**Fig. 9** Confusion matrix of classifiers applied to the dataset



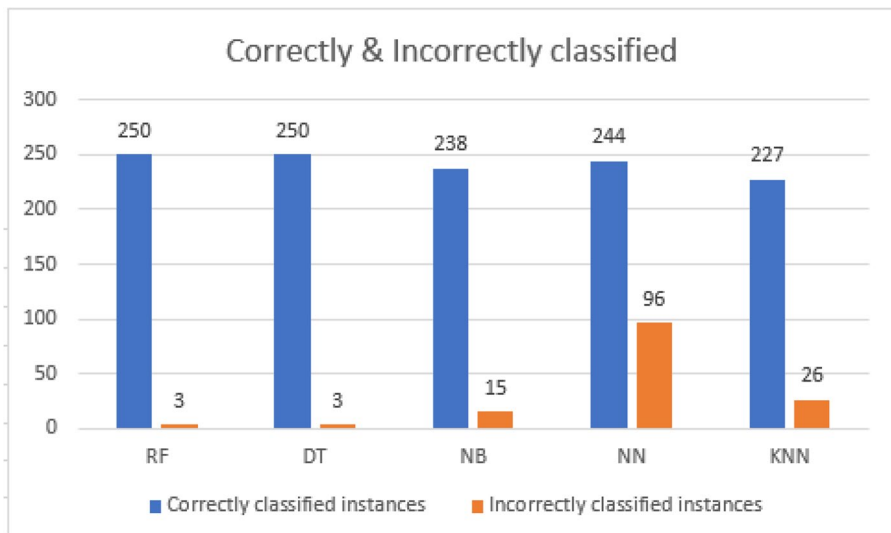
**Table 4** List of features that formed the dataset

No	ML algorithms	TP	FP	TN	FN	Precision	Recall	F-measure	Accuracy
1	RF	250	3	130	2	0.99	0.99	0.99	98.70%
2	DT	250	3	130	2	0.99	0.99	0.99	98.70%
3	NB	238	15	240	15	0.94	0.94	0.94	94%
4	NN	244	9	260	10	0.96	0.96	0.96	96.40%
5	KNN	227	26	215	25	0.9	0.9	0.9	89.60%

Figure 10 displays the the ratio of cases that were correctly categorized to those that were wrongly labelled for each classifier (RF, DT, NB, NN, and KNN). Based on the prediction results in the statistics course, the RF and DT classifiers demonstrated superior performance by accurately classifying 250 out of 253 instances. They only misclassified 3 instances, indicating a high level of accuracy. Similarly, the NB classifier accurately classified 238 instances and misclassified 15 instances. Similarly, the NN classifier accurately classified 244 instances and misclassified 9 instances, while the KNN algorithm correctly classified 227 instances and misclassified 26 instances.

Figure 11 displays the precision of each classifier as a result. The accuracy achieved during testing with the RF classifier was on par with that of the DT classifier, both reaching 98.7%. Likewise, the NB classifier obtained a 94% accuracy rate. Similarly, NN attained a success rate of 96.4%, while KNN reached 89.6%.

Figure 12 displays additional evaluation metrics, including precision, recall, and F-measure. Regarding the RF classifier and DT, the precision, recall, and F-measure metrics reached a value of 0.99. Similarly, the NB classifier obtained a score of 0.94. In addition, the NN classifier attained a score of 0.96, while the KNN achieved a score of 0.90.

**Fig. 10** Correctly vs. incorrectly classified instances for each classifier

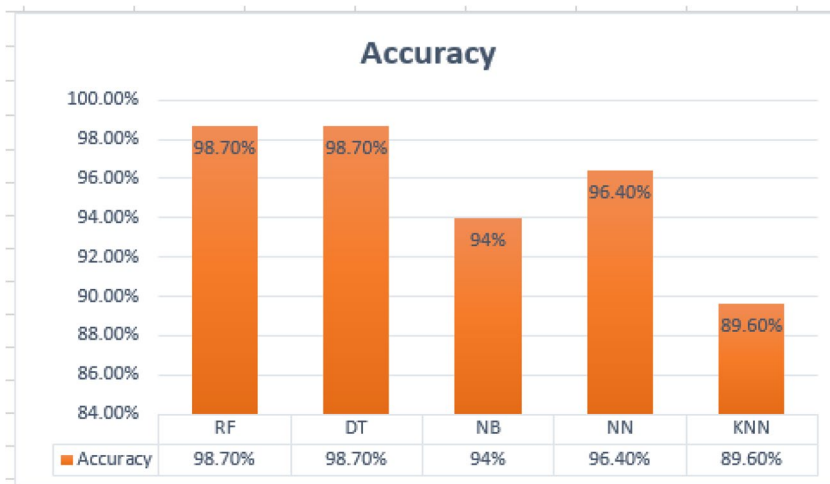


Fig. 11 Comparison accuracies of each classifier

## 8 Discussion

### 8.1 Identification of the requirements for prediction performance

The examples provided in the current study were adequate for generating a dataset. Upon comparing our study with others, we observe that some studies have utilized 210 examples to investigate student performance prediction (Asif

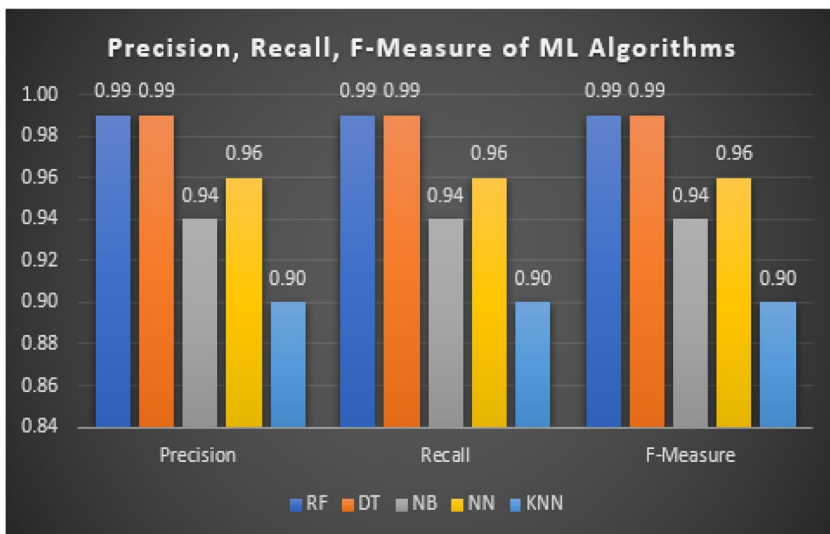


Fig. 12 Comparison of precision, recall, and F-measure metrics among ML classifiers

based on the F-measure demonstrate the predictive ability of the trained ML system on the validation dataset for two classes. Given that accuracy is the criterion for assessing the top-performing ML algorithms, Random Forest (RF), Neural Networks (NN), and Decision Trees (DT) demonstrated superior performance in F-measure validation tests compared to Naive Bayes (NB) and K-Nearest Neighbors (KNN). The F-measure achieved by the RF algorithm was 99%, outperforming KNN which scored 90%. The F-measure achieved by the DT algorithm was 99%, followed by 94% for the NP algorithm, and 96% for the NN algorithm.

### 8.3 Evaluation of trained ML algorithms

The results of the validation and testing to assess the accuracy of the most well-trained machine learning algorithms were displayed in Fig. 11. Both the RF algorithm and the DT algorithm demonstrated similar levels of accuracy during testing and validation. The RF algorithm achieved a 98.70% accuracy during testing, while the DT algorithm's accuracy closely matched that of the validation set. Moreover, the accuracy of the neural network decreased by 2% in the testing set, dropping from 98.70% in the validation results to 96.40% in the testing findings. The best-chosen trained ML algorithms showed no overfitting or underfitting when comparing accuracies between validation and testing. Additionally, throughout the ML algorithms' training, the F-measure in testing was contrasted with the validation outcomes (Fig. 12). In the validation findings, the F-measure for the RF method was 99%. Additionally, the DT algorithm's F-measure decreased from 99 to 96% in the validation findings. The F-measure for the KNN algorithm dropped from 96% in the validation results to 90% in the testing findings, a 6% decline.

"This evidence shows that the RF and DT algorithms successfully predicted the performance of the statistics course from new data, achieving a maximum accuracy of 98.70%. The neural network (NN) demonstrated the second-highest level of accuracy at 96.40%. Moreover, the RF and DT algorithm successfully predicted the performance of the statistics course on new data, achieving the highest F-measure of 99%, with NN following at 96% and NB at 94%. The results were compared to those of a previous study that examined accuracy ratings for the 5-level grading system. The RF algorithm achieved a 71.14% accuracy rate, while the binary level grading method reached 91.39% (Ünal, 2020). Within the same study, the DT algorithm's accuracy rose from 73.42% with the use of 5-level grading to 89.11% with the application of binary level grading. Given that accuracy tends to increase as classification becomes more specific, the findings of the present study showed strong accuracy in binary classification regarding whether students will pass or fail a statistics course. Based on this discussion, it is evident that the Random Forest (RF) and Decision Tree (DT) algorithms performed the best in predicting statistics performance of the course in the present research, accomplishing a precision of 98.70% and an F-measure of 99%. Therefore, the RF prediction model proved to be the most effective for predicting the performance of management degree students in statistics courses in the present study."

et al., 2017). Another research conducted by Saa in (2016) employed classification methods to forecast performance based on a dataset of 270 instances. "The researchers utilized 279 instances from the academic years 2007 to 2010 to predict students' math performance by considering factors such as oral, test, and final grades in the first and second semesters (Vihavainen et al., 2013). Compared to this study, 830 instances were analyzed, and six predictor factors were considered for their influence on the output variable."

## 8.2 Instruction and verification of algorithms for ML

Within this research, the accuracy of the DT model was recorded at 98.7%. The verification results were compared to a research study that achieved a DT accuracy of 91.5%. This study examined the relationships between predictor variables, and machine learning algorithms were trained using all features (Ma & Zhou, 2018). Moreover, the RF and NP demonstrated accuracy rates of 72.4% and 88.3%, respectively, compared to 98% and 94% in the present study. Comparatively, the KNN algorithm achieved an accuracy of 89.60% in this study, slightly lower than the 92.6% reported by Ma and Zhou (2018) (Fig. 13).

Moreover, the RF and NP accuracy rates in Ma and Zhou's (2018) study were 72.4% and 88.3%, while in the present research, they were 98% and 94%. In the current study, the KNN algorithm achieved an accuracy of 89.60%, which is slightly lower than the 92.6% reported in Ingale (2021). These results show that due to their interconnections, the algorithm accurately forecasts how well students perform in the course when taught alongside every aspect. Moreover, in a research study, the F-measure was employed as an assessment metric to predict the academic success of the students (Sokkhey et al., 2020). The current work utilized the F-measure metric to validate the ML algorithms trained efficiently. The instances that were correctly classified compared to those that were inaccurately categorized

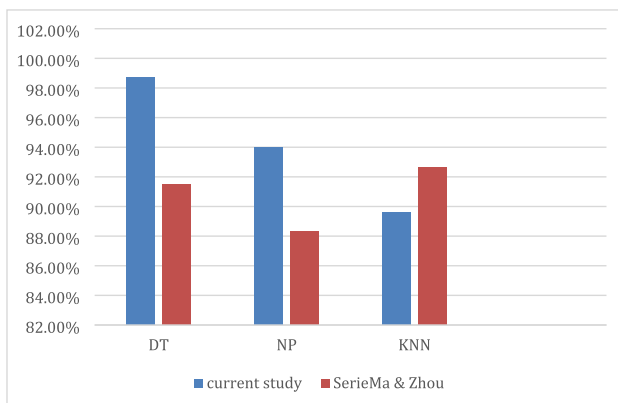


Fig. 13 Comparison between Algorithms for ML

## 9 Conclusion, limitations, and future work

The ability to forecast students' academic achievement using educational data is one of higher education's promising developments. The statistical metrics and DM algorithms presented in this research can be used to assess academic success. "These techniques utilize machine learning algorithms to assess student academic performance and decide if additional promotion is warranted. This study aims to predict the academic outcomes of freshman students majoring in Computer Science from 2016 to 2021. After removing 139 instances with missing values, irrelevant rows, and outliers, the dataset now consists of 830 instances. The dataset was split into a training set consisting of 577 instances (70%) and a test set consisting of 253 instances (30%). Besides, the dataset involved six features, that is, year, midterm, practical exam, writing exam, final total degree, and grade."

This paper involved five ML algorithms selected according to the literature review and high accuracy in predicting educational data mining. The ML algorithms are RF, DT, NB, NN, and KNN. Consequently, evaluation metrics were applied to compare ML algorithms, that is, confusion matrix, accuracy, precision, recall, and F-measure.

Based on the findings in this paper, the RF and DT classifiers demonstrated superior performance by accurately classifying 250 out of 253 instances when predicting students' performance in the statistics course, with only 3 instances being misclassified. Similarly, the NB classifier accurately classified 238 instances and misclassified 15 instances. Moreover, the NN classifier accurately classified 244 instances and misclassified 96 instances, while the KNN algorithm correctly classified 227 instances and misclassified 26 instances.

Furthermore, the RF and DT classifiers achieved an accuracy of 98.7% during testing. Likewise, NN attained a 96.4% accuracy. Similarly, NB attained a 94% accuracy rate, whereas KNN achieved 89.6%. When considering additional evaluation metrics like precision, recall, and F-measure for the RF and DT classifiers, the metrics reached 0.99. Similarly, the NN classifier obtained a score of 0.96. In addition, the NB classifier attained a score of 0.94, while the KNN achieved a score of 0.90. Hence, the paper has effectively attained a high level of accuracy in forecasting the academic performance of statistics students using ML algorithms.

This paper encountered some limitations in obtaining the dataset due to time, security, and privacy issues. Hence, this led to the elimination of collecting more features, which could play an essential role in enhancing the student's academic performance prediction. Besides, this paper uses a limited number of classification algorithms. In the future, we intend to collect big data of students' academic performance to predict the academic performance of several educational courses. Consequently, this can help students get appropriate jobs after improving their academic profiles. Furthermore, we hope to combine classification techniques with clustering algorithms and association rule mining to achieve improved results in educational data mining.

**Author contributions** This manuscript was read and approved by every one of the authors.

**Funding** Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB). No funding was applied to this study.

**Data availability** The application as well, the data that were utilised and/or analysed throughout the course of the current study can be obtained from the corresponding author upon making a reasonable request.

## Declarations

**Competing interests** This manuscript has not been submitted to any other journal or other publishing venue, nor is it currently being reviewed by any other publication. Regarding the subject matter that is discussed in the manuscript, the authors do not have any affiliation with any organisation that has a direct or indirect financial interest in the matter.

**Conflicts of interests** Non-financial conflicts of interest include.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adebayo, A. O., & Chaubey, M. S. (2019). Data mining classification techniques on the analysis of student's performance. *GSI*,7(4), 45–52.
- Ahmed, S. T., Al-Hamdani, R., & Croock, M. S. (2020). Enhancement of student performance prediction using modified K-nearest neighbor. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*,18(4), 1777–1783.
- Ajibade, S. S. M., Dayupay, J., Ngo-Hoang, D. L., Oyebode, O. J., & Sasan, J. M. (2022). Utilization of ensemble techniques for prediction of the academic performance of students. *Journal of Optoelectronics Laser*,41(6), 48–54.
- Al-Obeidat, F., Tubaishat, A., Dillon, A., & Shah, B. (2018). Analyzing students' performance using multi-criteria classification. *Cluster Computing*,21, 623–632.
- Aman, F., Rauf, A., Ali, R., Iqbal, F., & Khattak, A. M. (2019). A predictive model for predicting students academic performance. In *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)* (pp. 1–4). IEEE.
- Amazona, M. V., & Hernandez, A. A. (2019). Modelling student performance using data mining techniques: Inputs for academic program development. In *Proceedings of the 2019 5th International Conference on Computing and Data Engineering* (pp. 36–40).
- Amra, I. A. A., & Maghari, A. Y. (2017). Students performance prediction using KNN and Naïve Bayesian. In *2017 8th International Conference on Information Technology (ICIT)* (pp. 909–913). IEEE.
- Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*,113, 177–194.
- Aydoğdu, Ş. (2020). Predicting student final performance using artificial neural networks in online learning environments. *Education and Information Technologies*,25(3), 1913–1927.
- Baker, R. S., Martin, T., & Rossi, L. M. (2016). Educational data mining and learning analytics. *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications*, 379–396.
- Bruce, A. (2019). The prediction of student performance through the use of machine learning, MSc Software Development Dept. of Computer and Information Sciences University of Strathclyde.

- Ghosh, P., Kiran, S., Mahalakshmi, J., & Basha, S. A. H. (2023). *Understanding machine learning*. AG PUBLISHING HOUSE (AGPH Books).
- Gil, P. D., da Cruz Martins, S., Moro, S., & Costa, J. M. (2021). A data-driven approach to predict first-year students' academic success in higher education institutions. *Education and Information Technologies*, 26(2), 2165–2190.
- Hossain, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2), 1.
- Hussain, S. A., Al Bassam, N., Zayegh, A., & Al Ghawi, S. (2022). Prediction and evaluation of healthy and unhealthy status of COVID-19 patients using wearable device prototype data. *MethodsX*, 9, 101618.
- Ingale, N. V. (2021). Survey on prediction system for student academic performance using educational data mining. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(13), 363–369.
- Kaur, H., & Bathla, E. G. (2018). Student performance prediction using educational data mining techniques. *International Journal on Future Revolution in Computer Science & Communication Engineering*, 4(12), 93–97.
- Lau, E. T., Sun, L., & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Applied Sciences*, 1, 1–10.
- Lenin, T., & Chandrasekaran, N. (2019). Students' performance prediction modelling using classification technique in R. *International Journal of Recent Technology and Engineering*, 8(2), 5197–5201.
- Ma, X., & Zhou, Z. (2018, January). Student pass rates prediction using optimized support vector machine and decision tree. In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 209–215). IEEE.
- Markoulidakis, I., Kopsiaftis, G., Rallis, I., & Georgoulas, I. (2021). Multi-class confusion matrix reduction method and its application on net promoter score classification problem. In *The 14th Pervasive Technologies Related to Assistive Environments Conference* (pp. 412–419).
- Matsson, A., & De Geer, C. (2023). Personalized software in heavy-duty vehicles-exploring the feasibility of self-adapting smart cruise control using machine learning, master's thesis in complex adaptive systems and systems, control and mechatronics, department of electrical engineering systems and control, chalmers university of technology.
- Mduma, N., Kalegele, K., & Machuve, D. (2019). Machine learning approach for reducing students dropout rates. *International Journal of Advanced Computer Research*, 9(42), 156–169.
- Muntean, M., & Militaru, F. D. (2023). Metrics for evaluating classification algorithms. In *Education, Research and Business Technologies: Proceedings of 21st International Conference on Informatics in Economy (IE 2022)* (pp. 307–317). Springer Nature Singapore.
- Nuankaew, W., & Thongkam, J. (2020). Improving student academic performance prediction models using feature selection. In *2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (pp. 392–395). IEEE.
- Onawumi, A. S., Akinrinade, N. A., Abisoye, A. S., Olalere, S. O., Ojojo, F. E., & Sanyaolu, O. O. (2023). Mismatch between anthropometric measurements of occupational drivers in southwest nigeria and vehicle seat design parameters. *Valley International Journal Digital Library*, 918–931.
- Patil, R., Salunke, S., Kalbhor, M., & Lomte, R. (2018). Prediction system for student performance using data mining classification. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)* (pp. 1–4). IEEE.
- Rizvi, H., Sanchez-Vega, F., La, K., Chatila, W., Jonsson, P., Halpenny, D., ... & Hellmann, M. D. (2018). Molecular determinants of response to anti-programmed cell death (PD)-1 and anti-programmed death-ligand 1 (PD-L1) blockade in patients with non-small-cell lung cancer profiled with targeted next-generation sequencing. *Journal of clinical oncology*, 36(7), 633.
- Roy, S., & Garg, A. (2017). Analyzing performance of students by using data mining techniques a literature survey. In *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)* (pp. 130–133). IEEE.
- Saa, A. A. (2016). Educational data mining & students' performance prediction. *International Journal of Advanced Computer Science and Applications*, 7(5), 212–220.
- Saifudin, A., & Desyani, T. (2020, March). Forward selection technique to choose the best features in prediction of student academic performance based on Naïve Bayes. In *Journal of Physics: Conference Series* (Vol. 1477, No. 3, p. 032007). IOP Publishing.
- Santoso, H. B. (2020). Fuzzy decision tree to predict student success in their studies. *International Journal of Quantitative Research and Modeling*, 1(3), 135–144.

- Sen, P. C., Hajra, M., & Ghosh, M. (2020). Supervised classification algorithms in machine learning: A survey and review. In *Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018* (pp. 99–111). Springer Singapore.
- Shahiri, A. M., & Husain, W. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 72, 414–422.
- Sokkhey, P., Navy, S., Tong, L., & Okazaki, T. (2020). Multi-models of educational data mining for predicting student performance in mathematics: A case study on high schools in Cambodia. *IEIE Transactions on Smart Processing and Computing*, 9(3), 217–229.
- Sultana, S., Khan, S., & Abbas, M. A. (2017). Predicting performance of electrical engineering students using cognitive and non-cognitive features for identification of potential dropouts. *International Journal of Electrical Engineering Education*, 54(2), 105–118.
- Tohka, J., & Van Gils, M. (2021). Evaluation of machine learning algorithms for health and wellness applications: A tutorial. *Computers in Biology and Medicine*, 132, 104324.
- Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & Education*, 143, 103676.
- Ünal, F. (2020). Data mining for student performance prediction in education. *Data Mining-Methods, Applications and Systems*, 28, 423–432.
- Usman, M. M., Owolabi, O., & Ajibola, A. A. (2020). Feature selection: It importance in performance prediction. *IJESCI*, 10(5), 25625–25632.
- Vihavainen, A., Luukkainen, M., & Kurhila, J. (2013, July). Using students' programming behavior to predict success in an introductory mathematics course. In International Conference on *Educational Data Mining 2013*. Memphis, USA, (pp. 300–303).
- Vital, T. P., Sangeeta, K., & Kumar, K. K. (2021). Student classification based on cognitive abilities and predicting learning performances using machine learning models. *International Journal of Computing and Digital Systems*, 10(1), 63–75.
- Wakelam, E., Jefferies, A., Davey, N., & Sun, Y. (2020). The potential for student performance prediction in small cohorts with minimal available attributes. *British Journal of Educational Technology*, 51(2), 347–370.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining practical machine learning tools and techniques* (3rd ed.). Morgan Kaufmann.
- Wong, M. L., & Senthil, S. (2019). Applying attribute selection algorithms in academic performance prediction. *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018* (pp. 694–701). Springer International Publishing.
- Wu, N., Zhang, L., Gao, Y., Zhang, M., Sun, X., & Feng, J. (2019). CLMS-Net: dropout prediction in MOOCs with deep learning. In *Proceedings of the ACM Turing Celebration Conference-China* (pp. 1–6).
- Zaffar, M., Hashmani, M. A., Savita, K. S., Rizvi, S. S. H., & Rehman, M. (2020). Role of FCBF feature selection in educational data mining. *Mehran University Research Journal of Engineering & Technology*, 39(4), 772–778.
- Zhao, L., Chen, K., Song, J., Zhu, X., Sun, J., Caulfield, B., & Mac Namee, B. (2020a). Academic performance prediction based on multisource, multifeature behavioral data. *IEEE Access*, 9, 5453–5465.
- Zhao, Y., Ren, W., & Li, Z. (2020b). Prediction of english scores of college students based on multi-source data fusion and social behavior analysis. *Revue d'Intelligence Artificielle*, 34(4), 465–470. <https://doi.org/10.18280/ria.340411>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



## Authors and Affiliations

**Dalia Khairy<sup>1</sup> · Nouf Alharbi<sup>2</sup> · Mohamed A. Amasha<sup>1</sup>  · Marwa F. Areed<sup>4</sup> · Salem Alkhalaf<sup>3</sup> · Rania A. Abougalala<sup>1</sup>**

✉ Mohamed A. Amasha  
mw\_amasha@yahoo.com

Dalia Khairy  
shamaamora2014@gmail.com

Nouf Alharbi  
nmoharbi@taibahu.edu.sa

Marwa F. Areed  
Marwa\_Areed@du.edu.eg

Salem Alkhalaf  
s.alkhalaf@qu.edu.sa

Rania A. Abougalala  
Ronyabogalala@hotmail.com

<sup>1</sup> Faculty of Specific Education, Department of Computer Teacher Preparation, Damietta University, Damietta, Egypt

<sup>2</sup> College of Computer Science and Engineering, Taibah University, 42353 Madinah, Saudi Arabia

<sup>3</sup> Faculty of Engineering, Department of Computer Science, Damietta University, Damietta, Egypt

<sup>4</sup> Faculty of Science and Arts, Computer Science Department, Qassim University, Alrass, Saudi Arabia