Prediction of Students Performance Level Using Integrated Approach of ML Algorithms

https://doi.org/10.3991/ijet.v18i01.35339

Mirza Azam Baig, Sarmad Ahmed Shaikh(✉), Kamlesh Kumar Khatri,
Muneer Ahmed Shaikh, Muhammad Zohaib Khan, Mahira Abdul Rauf
Department of Computer Science, Sindh Madressatul Islam University, Karachi, Pakistan
sarmad@smiu.edu.pk

Abstract—In this paper, the efficacy of machine learning (ML) techniques for predicting the academic success of students is investigated. In issues pertaining to higher education, as well as machine learning, deep learning, and its linkages to educational data, predicting student achievement is essential. The choice of courses and the development of effective future study plans for students can be easier with the help of the capacity to forecast a student's success. In addition to predicting student achievement, it makes it easier for instructors and administrators to keep an eye on children so that they can offer support and integrate trainings for the greatest outcomes. In this study, we define the idea of predicting the student performance in education and its several iterations. We discuss a number of ML approaches, such as the Fuzzy C-Means, the Multi-Layer Perceptron (MPL), the Logistic Regression (LR), and the Random Forest (RF) algorithms, for predicting student achievement in the classroom. The models for forecasting student performance that are now in use and those that have been proposed in this paper are carefully investigated. The paper examines different combinations of the algorithms including FCM – MLP, FCM – LR, and FCM – RF, and provides the detailed results of each combination. These strategies are assessed using quantitative standards including accuracy, detection rate, and false alarm rate.

Keywords—Machine Learning (ML), Fuzzy C-Means (FCM), Multi-Layer Perceptron (MPL), Logistic Regression (LR), and Random Forest (RF) Algorithms

1 Introduction

Today, major emphasis is being made on predicting student performance as a result of the relevance of that type of issue in the development of nations all over the world. This is due to the fact that the educational process leads to the production of a generation capable of leading the country and march towards development in all aspects of life. The efficacy of academic institutions, which are in charge of shaping future generations in accordance with the various stages of people's lives in each nation, is also reflected in the evaluation of students' performance. As a result, one of the most urgent needs that motivates governments to make significant efforts is to advance the educational process. In this regard, the students' performance can be monitored by using technological approaches such as the machine learning (ML) approach which can predict the features of students' performance. ML is a method for finding hidden information by studying numerous data sources in fields including business, social, medical, and education. The more data there is, such as in large databases, the better the forecast [1]. Several ML techniques are in use right now, particularly in the field of industry. Some approaches necessitate the use of supervised learning techniques, whereas others necessitate the use of unsupervised learning techniques. On structured or labeled data, the supervised learning approach works. Beyond that, supervised learning applies what it has learnt from past data to new data (new data). It gives the machine model training data and forecasts whether it's a circle or a square, for example. For subcategory classification and regression, this supervised learning method is used. Unsupervised learning makes use of unstructured or hidden data and necessitates the machine being given a diverse set of inputs. Unsupervised learning evaluates data before classifying it; as a result, we get different data in each group after classification, and each group is distinct from the others [2].

In order to gather the essential information, it is possible to study the knowledge provided by various educational data resources. A new field called Educational Data Mining (EDM) was established as a method of identifying essential information [3]. Sharing statistics and integrating deep learning with exploration create the best learning environment. The importance of EMD has rapidly increased in recent years because of increase in data collected, according to educational data obtained from various e-learning systems and the expansion of traditional educational systems. The abundance of data across numerous domains and their relationships with one another demonstrate the power of EMD. It is focused with the extraction of traits from vast volumes of data provided by the institute to assist the educational process [4]. Academic institutions work to create a student model that can predict each student's characteristics and performance [5]. From those vast datasets, we need to classify the useful information and process it for predicting the desired parameters [6–7].

2 Literature review

The use of the ML approach to forecasting students' performance based on their history and term exam results has proven to be useful in predicting different levels of performance. Using such ML algorithms, it is possible to anticipate which pupils are likely to fail and as a result deliver a solution to students in a timely manner. It can also assist the educational institutes in identifying high-achieving students and assisting in offering scholarships. Ioannis E. Livieris, et al. in [8] predict student success in mathematics. The authors created an Artificial Neural Network (ANN) classifier. They discovered that the modified spectral Perry trained artificial neural network performs superior classification than other classifiers in their testing. S. Kotsiantis, et al. [9] studied the use of ML techniques in remote learning for the prediction of student dropout. This study made a significant contribution since it was a trailblazer in the field of educational data mining. Although ML techniques had been used in other fields, he and his colleagues were the first to use them in an academic setting. Rather of using class performance data, an algorithm was given demographic data and numerous project assignments. Moucary, et al. in [10] proposed a hybrid technique K-Means Clustering and Artificial Neural Network (ANN) for students pursuing higher education while learning a new foreign language as a method of teaching and communication. To begin, a neural network was used to forecast the students' performance, and then the K-Means technique was used to group them into a certain cluster. This clustering assisted instructors in identifying a student's skills throughout the early stages of their academic careers.

A data mining-based prediction model for student performance with a few characteristics called student behavioral features is introduced in [11–12]. Three distinct classifiers were used to evaluate the model including Nave Bayesian (NB), ANN, and Decision Tree (DT). Ensemble approaches such as Random Forest (RF), Bagging, and Boosting were utilized to increase the classifier's performance. When behavioral elements were eliminated, the model's accuracy increased by up to 22.1%. After employing ensemble approaches, the accuracy climbed to 25.8%. The authors in [12] recommended prediction model structure that provides characteristics that result in less variation in information acquisition. Because ML algorithms break down problems and solve them one at a time, it's possible that when one condition fails, other crucial attributes are lost. This has an impact on the model's overall performance. To do so, the model must allow end-to-end functionality. Shaymaa E. Sorour et al. in [13] used DT and RF to create an interpretable model. They used a technique called remark data-mining to forecast grades and extracted rules. In reference [14], the authors used Neuro-Fuzzy to predict and classify student academic success based on CGPA. Quadriet et al. in [15] used J48 to forecast the dropout rate of students based on their CGPA. The elements impacting the pupils' performance were found by Christian et al. in [16] using the NB method. The authors looked at the schooling, particular, induction, and educational data of students. Muslihah Wook et al. in [17] anticipated academic achievement of pupils by using the classification approaches such as ANN and DT. Furthermore, Shaleenaet et al. in [18] used DT classifiers for identifying and predicting dropout students ratio as well as a discussion of the topic of class imbalance. Another study in [19] used the deep neural network (DNN) and ANN algorithms to improve the DL prediction model and implementation technique. V. Vijayalakshmi et al. in [20] adopted ML and DL techniques in the domain of education system on student performance to test the performance of their suggested strategy.

Most of the reported techniques have achieved the results of predicting the students' performance with complex structures and the results still need to be improved. In this paper, by examining both supervised and unsupervised learning methods, followed by the Fuzzy C-Means (FCM) clustering approach, and classification techniques i.e., Multi-Layer Perceptron (MLP), Logistic Regression (LR), and Random Forest (RF) algorithms, we predict the student's performance based on education parameters. Different combinations of the considered algorithms are investigated. Most of the combinations justified improved accuracy, precision, recall, f1-score, classifier, prediction, and clusters in a rapid, precise, and accurate manner.

3 Proposed methodology

This section describes the proposed method which begins with an algorithm, subsequently loads data from a database, and finally, prepares the information. The information is pre-processed, normalized, and standardized. The standardized approach works by first normalizing data to convert floating frequency values to the same (0.1543648795), then doing standardized analysis and calculating the result. Then we use the Fuzzy C-Means clustering technique and the classification algorithms including Multi-Layer Perceptron, Logistic Regression, and Random Forest supervised classification algorithms to design and create the target model. The sample size for the test is 25%, whereas the sample size for the implementation training is 75%. Figure 1 illustrates a flowchart that can assist in understanding the research activity and outlining the job breakdown structure concisely. In this illustration, we initially retrieve data from the database to prepare data, then actually begin pre-processing the data in the data pre-processing section using standardized data cleaning techniques, then work on clustering and classification implementing 75% processed training data and 25% test data for model validation.

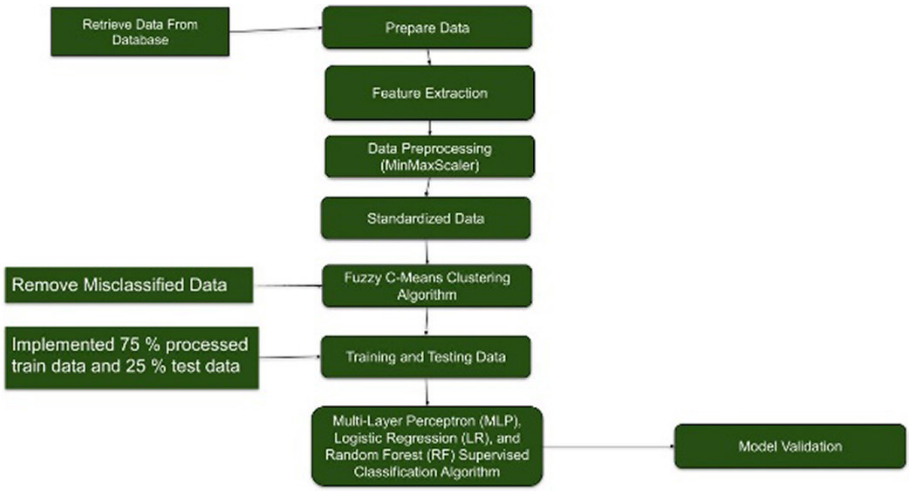

Fig. 1. Proposed method for students performance prediction

The proposed method consists of the following two subsections: 1) Pre-processing, and 2) Classification.

3.1 Preprocessing

In preprocessing phase, to obtain meaningful data, we clean the data and use it for clustering. As explained below, this is done in two stages i.e., data collection, and clustering by using the Fuzzy C-means clustering technique; and then further process the data in the classification phase.

Data collection. The dataset of the international high education students performance, has been collected from the Kaggle website, which had previously been utilized by the researchers in [20]. The dataset contains a total of 480 tuples and 17 properties, each of which represents a student's performance metric. The dataset contains 305 male and 175 female students sample records from international, with one target class (which denotes each class's status separately) and a total of 128 Low Class, 211 Medium Class, and 142 High Class. The actual international high education students performance dataset feature description is shown in Table 1 [21].

Table 1. Actual dataset for international high education students performance

Gender	Nationality	Place of Birth	Stage ID	Grade ID	Section ID	Topic
M	KW	KuwaIT	lowerlevel	G-04	A	IT
M	KW	KuwaIT	lowerlevel	G-04	A	IT
M	KW	KuwaIT	lowerlevel	G-04	A	IT
M	KW	KuwaIT	lowerlevel	G-04	A	IT
M	KW	KuwaIT	lowerlevel	G-04	A	IT
Semester	**Relation**	**Raised Hands**	**Visited Resources**	**Announcements View**	**Discussion**	**Parent Answering Survey**
F	Father	15	16	2	20	Yes
F	Father	20	20	3	25	Yes
F	Father	10	7	0	30	No
F	Father	30	25	5	35	No
F	Father	40	50	12	50	No

Parent School Satisfaction	Student Absence Days	Class
Good	Under-7	M
Good	Under-7	M
Bad	Above-7	L
Bad	Above-7	L
Bad	Above-7	M

Table 2 displays the preprocessed dataset that has been cleaned and standardized, and prepared for the clustering and classification phases to predict the students performance. Moreover, Figure 2 shows the mixed data chart of the corresponding cleaned data before clustering.

Table 2. Preprocessed dataset for international high education students performance

Gender	Nationality	Place of Birth	Stage ID	Grade ID	Section ID	Topic
0.276542	0.456038	0.416909	0.675114	0.664516	0.959812	0.820159
0.595424	0.976172	0.329478	0.635595	0.804943	0.677397	0.132544
0.181578	0.429730	0.063652	0.445259	0.907755	0.918585	0.755372
0.830483	0.088741	0.159844	0.527569	0.263223	0.741383	0.871822
0.210892	0.701391	0.766264	0.172061	0.103681	0.024694	0.419428
Semester	**Relation**	**Raised Hands**	**Visited Resources**	**Announcements View**	**Discussion**	**Parent Answering Survey**
0.609305	0.522702	0.782175	0.892479	0.164947	0.646486	0.145420
0.169182	0.336317	0.587580	0.624576	0.543965	0.676087	0.104618
0.045972	0.292576	0.704262	0.028200	0.060382	0.371900	0.112741
0.152801	0.884067	0.992723	0.117836	0.414137	0.987809	0.355952
0.949947	0.700166	0.426451	0.145548	0.839285	0.519597	0.009649

Parent School Satisfaction	Student Absence Days	Class
0.667879	0.930874	0.996918
0.115494	0.212968	0.871445
0.355699	0.400581	0.200401
0.783681	0.098311	0.266751
0.763605	0.616193	0.617324

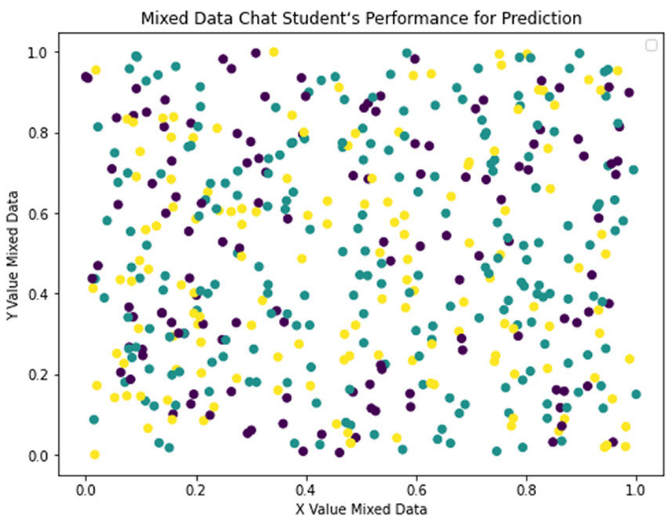

Fig. 2. The mixed data chart before applying the clustering technique

Fuzzy C-means clustering technique. Clustering is the most common type of unsupervised learning approach, having a wide range of applications and implementations in a variety of industries. Clustering is the process of splitting and processing data on behalf of an information machine, resulting in a collection of data known as clustering, with each cluster given a unique identification number (ID). For data clustering, the Fuzzy C-means (FCM) clustering approach is the most often employed. FCM is a soft cluster method in which individually facts argument is given a probability or possibility score that indicates whether it belongs to that cluster. It accomplishes this by introducing new uncorrelated variables in a sequence that maximizes variance. This logic is instantly applied to the data matrix, resulting in a membership matrix that shows the degree of connection between the samples and each cluster.

Taking into consideration the space amongst the centroid and the target value, this method assigns participation to each data point corresponding to each centroid. The data belongs to the cluster more if it is near to the centroid. The actual number for each data point should be one [22]. After each cycle, the following formula is used to change participation and cluster centroid.

$$\mu ij = 1 / \sum_{k=1}^{c} (d_{ij}/d_{ik})^{(2/m-1)}, \tag{1}$$

$$\mathrm{v}j = \frac{\left(\sum_{i=1}^{n} (\mu ij)^m x_i\right)}{\left(\sum_{i=1}^{n} (\mu ij)^m\right)}, \forall j = 1, 2, 3, \ldots c, \tag{2}$$

where, '*n*', '*vj*', '*m*', '*c*', '*μij*', and '*dij*' denote the number of data points, cluster center, the fluffiness directory, the number of cluster centers, the connection between the *ith* data and the *jth* cluster center, and the Euclidean distance between the *ith* data and the *jth* cluster center, respectively.

The core impartial of the fuzzy c-means approach is to decrease the Euclidean distance given below

$$J(U,V) = \sum_{i=1}^{n} \sum_{j=1}^{c} (\mu ij)^m \left\| x_i - v_j \right\|^2. \tag{3}$$

The Euclidean distance amongst both the *ith* data point and the *jth* cluster center is signified by '$\|xi - vj\|$'. There are a few mathematical functions that used for FCM algorithm such as the Euclidean, Manhattan, and Hamming distances. This method is constructed on categorization. Some of these approaches are as follows:

$$\text{Euclidean} \quad \sqrt{\sum_{i=1}^{k} (xi - yi)2} \tag{4}$$

$$\text{Manhattan} \quad \sum_{\mathrm{i}=1}^{\mathrm{k}} |\mathrm{xi} - \mathrm{yi}| \tag{5}$$

$$\text{Minkowski} \quad \left(\sum_{\mathrm{i}=1}^{\mathrm{k}} \left(|\mathrm{xi} - \mathrm{yi}| \right) \mathrm{q} \right) 1/\mathrm{q} \tag{6}$$

FCM methodology steps:

Let X = {x1, x2, x3..., xn} represent a set of data points and V = {v1, v2, v3..., vc} represent a set of centers, then

1. Randomly select 'c' cluster centroids
2. O use the following formula to calculate the fuzzy participation 'μij':

$$\mu ij = 1 / \sum_{k=1}^{c} (d_{ij}/d_{ik})^{(2/m-1)} \tag{7}$$

3. Compute the fuzzy centroids '*vj*' as follows:

$$\mathrm{v}j = \left(\sum_{i=1}^{n} (\mu ij)^m x_i \right) / \left(\sum_{i=1}^{n} (\mu ij)^m \right), \forall j = 1, 2, 3, \ldots c \tag{8}$$

4. Repeat steps 2 and 3 until the smallest 'j' value is attained or ||U(k + 1) – U(k)|| < β is achieved, whereas the repetition step is signified by the letter '*k*.', the completion criteria between [0, 1] is 'β', the fuzzy relationship matrix is well-defined as '*U* = (*ij*) *n***c*', and the impartial function is represented by the letter '*J*.'

The student's performance prediction and clusters dataset is clearly well-defined as a 17-dimensional dataset, with three features based on student's performance prediction values and one feature target property cluster number (see Table 3). The FCM Clustering centroid value briefly defined is given below, whereas Figure 3 shows the three clusters after the unstructured dataset has been converted to structured data. This graph depicts two clusters: red and green, with centroid values specified by the yellow star (*). Figure 4 shows FCM determined sum of squared error line chart.

FCM Clustering Centroid Value.

```
array([[0.49492933, 0.4917658 , 0.51983247,
0.47320961, 0.49158729, 0.52289268, 0.5230012 ,
0.47840191, 0.50558121, 0.51672337, 0.49937354,
0.51753685, 0.49081544, 0.48903864, 0.51368392,
0.48831137, 0.51521064],
[0.4949292 , 0.49176579, 0.51983274, 0.47320943,
 0.49158728, 0.52289266, 0.52300126, 0.47840171,
0.50558125, 0.51672342, 0.49937341, 0.51753673,
0.49081553, 0.48903883, 0.51368394,
0.48831103, 0.51521079],
[0.49492982, 0.49176602, 0.51983238, 0.4732094 ,
 0.49158729, 0.52289266, 0.5230013 , 0.47840182,
0.50558108, 0.51672348, 0.49937376, 0.51753657,
0.49081558, 0.48903829, 0.513684,
0.48831106, 0.51521069]])
```

Table 3. FCM three clusters preprocessed for international high education students performance dataset

Gender	Nationality	Place of Birth	Stage ID	Grade ID
0.276542	0.456038	0.416909	0.675114	0.664516
0.595424	0.976172	0.329478	0.635595	0.804943
0.181578	0.429730	0.063652	0.445259	0.907755
0.830483	0.088741	0.159844	0.527569	0.263223
0.210892	0.701391	0.766264	0.172061	0.103681
Section ID	**Topic**	**Semester**	**Relation**	**Raised Hands**
0.959812	0.820159	0.609305	0.522702	0.782175
0.677397	0.132544	0.169182	0.336317	0.587580
0.918585	0.755372	0.045972	0.292576	0.704262
0.741383	0.871822	0.152801	0.884067	0.992723
0.024694	0.419428	0.949947	0.700166	0.426451
Visited Resources	**Announcements View**	**Discussion**	**Parent Answering Survey**	**Parent School Satisfaction**
0.892479	0.164947	0.646486	0.145420	0.667879
0.624576	0.543965	0.676087	0.104618	0.115494
0.028200	0.060382	0.371900	0.112741	0.355699
0.117836	0.414137	0.987809	0.355952	0.783681
0.145548	0.839285	0.519597	0.009649	0.763605

Student Absence Days	Class	Clusters
0.930874	0.996918	2
0.212968	0.871445	2
0.400581	0.200401	2
0.098311	0.266751	2
0.616193	0.617324	0

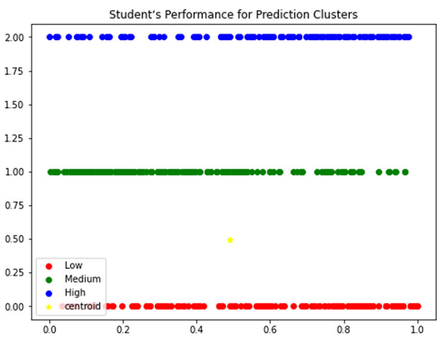

Fig. 3. The FCM three clusters

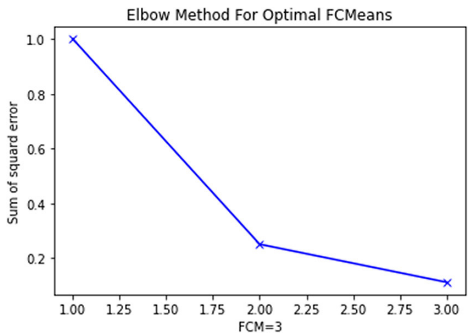

Fig. 4. The FCM sum of squared error line chart

A concern value is produced based on the degree of divergence from particular dimensions and thereby, the based on student's performance prediction result is classified as valid or suspect, or predictions.

3.2 Classification

The classification is a supervised learning strategy that determines the category of observations using training data. The process of learning from a dataset of observations and then categorising the observations into one of several classes or groupings is known as classification. In our dataset, we test the following classification methods to see which one performs better.

Multi-layer perceptron algorithm. A multi-layer perceptron (MLP) is a sophisticated optimization technique. It's finished up of countless perceptron. The algorithm consists of three layers: an input layer that receives data, an output layer that takes a decision or guesses about the input, and an endless number of hidden layers that act as MLP's true processing arrangement. MLP may approximate any continuous function with a varied number of hidden layers. This technique effectively estimates each and every linear method. It classifies collections that are not conditionally independent. Participants achieve this by creating machine learning and predicting models for challenging facts using a more elastic and complicated infrastructure, and this strategy is widely used to deal with supervised learning challenges [23]. This method, like the Sigmoid, Linear, Cost Linear, and Non-linear Regression, is built around classification, as given below.

$$\text{Sigmoid} \quad \mathrm{S}(z) = \frac{1}{(1 + \mathrm{e}^{-z})} \tag{9}$$

$$\text{Linear Logistic Regression} \quad y = e \wedge (b0 + b1 * x) / (1 + e \wedge (b0 + b1 * x)) \tag{10}$$

$$\begin{array}{ll} \text{Cost Linear Logistic} & (Cost\,(h\theta(x), y)) = -log\,(h\theta(x)), if\ y = 1\, and \\ \text{Regression} & (Cost\,(h\theta(x), y)) = -log\,(1 - h\theta(x)), if\ y = 0 \end{array} \tag{11}$$

$$\text{Nonlinear Logistic Regression} \quad \mathrm{Y} = \mathrm{f}(\mathrm{X}, \beta) + \varepsilon \tag{12}$$

The main steps of MLP algorithm are given below.

1. The MLP, like the perceptron, advances facts by calculating the partial derivatives of the input records and the parameters that appear between the input and hidden layer. This sampling distribution generates a value in the hidden layer. However, unlike an activation function, we do not increment this value.
2. The MLP employs activation functions in each of their estimated layers. There are various processing mechanisms to evaluate, such as rectified linear units (ReLU), the sigmoid, and the mechanisms. Any of these methods is used to transfer the measured output to the visible layer.

3. After the expected outcome at the invisible layer has been generated through the activation function, extract the partial derivatives with the required values and transmit them to another layer inside the MLP.
4. Repeat the above procedures two – three times till the final output is acquired.
5. The estimates will be used at the output to acquire results for either a feed-forward technique corresponding to the activation methods chosen for only the MLP (in the scenario of training data), or a selection obtained from the results (in the situation of testing data).
6. End.

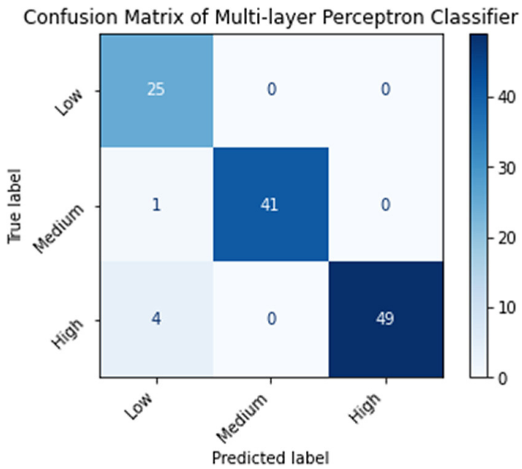

Fig. 5. Confusion matrix of MLP

The MLP labels the old data values and predicts the value of the new data. We try to make predictions fit the labels during preparation. Figure 5 shows the result of that MLP confusion matrix. As of this paper, the confusion matrix has the meaning as [[A B][C D]], whereas

- A is the frequency of correctly predicted negative instances,
- B is the quantity of inaccurate guesses that a positive occurrence has.
- C is the number of inaccurate predictions that a negative occurrence has, and
- D is the frequency of right predictions of a positive case.

We used the matrix of MLP Algorithm for assessing the model accuracy and model loss concepts. This will increase the approximated prediction approach's accuracy while also ensuring that the patterns of student's performance prediction are fulfilled on a frequent basis. The achieved model accuracy is 0.9583 and further discussed in Section 5 of Results.

Logistic regression algorithm. The logistic regression (LR) is a ML approach used for classification procedures. It is a probability hypothesis-based predictive analytic technique. A LR model is functionally equivalent to a linear regression model. However, LR employs an additional complicated cost-function, known as the sigmoid-function or often as the logistic-method, rather than a linear function. The LR hypothesis limits the differential equation to values amongst 0 and 1. As an outcome, linear-functions flop to characterize it since they might have values more than one or even less than zero, which is not possible under the LR assumption.

The LR algorithm's major steps are given below and the detail about the LR and its mathematical functions can be found in [24]:

1. Initialize all parameters *(B_0, B_1, etc.)*.
2. Compute (predict) dependent variable *($h_\theta(x)$)*.
3. Compute cost function *(Cost ($h_\theta(x),y$))* or any Logistic Regression function.
4. Compute gradient for the cost function.
5. Update all parameters.
6. Repeat steps 2 to 5.
7. End.

The LR model's true and predicted labels are shown in confusion matrix in Figure 6. We apply the LR algorithm on our above dataset and label the old data values. It predicts the value of data – where we tried to make our predictions fit the labels during preparation with the use of the LR algorithm. Figure 6 shows the result of that matrix. This will increase the approximated prediction approach's accuracy while ensuring that the patterns of students' performance prediction are fulfilled on a frequent basis. Furthermore, the achieved model accuracy is 0.9583, identical to MLP, and further discussed in Section 5 of Results.

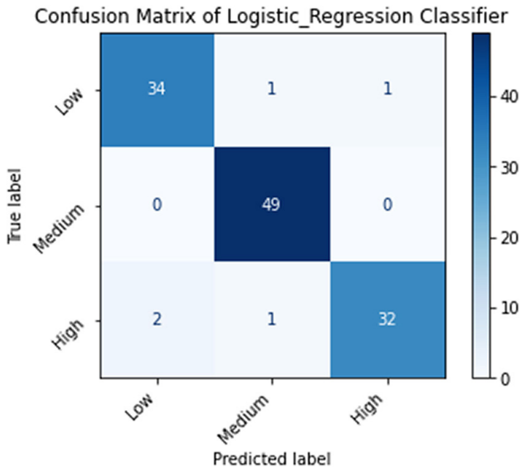

Fig. 6. Confusion matrix of LR algorithm

Random forest algorithm. A random forest (RF) is a ML technique that is use to resolve the classifier issues. It makes use of different classifiers, a sort of difficult resolving system, that makes use of classifying approaches. That's the technique of merging numerous categories to solve complicated issues and enhance the effectiveness of the system. The RF approach, which is built on classification tree predictions, decides the effectiveness. It makes assumptions by approximating or calculating the results of numerous trees. The quality of the output increases even as number of nodes increases [25–26].

The RF algorithm is based on the following steps:

1. Initiate by randomly picking observations through facts.
2. The program will instead create a tree structure for every instance. The outcomes for every tree structure will then be produced.
3. Throughout this stage, every generated came as a result would be selected or decided on.
4. Lastly, select the most preferred forecasting outcome as unique of the most predicted results.

Moreover, the RF algorithm uses specific mathematical functions. This approach, as well as Mean Squared Error (MSE), Gini (Coefficient, Index, or Ratio), and Entropy [27], may be used by the RF algorithm such as given below.

$$\text{Mean Squared Error (MSE)} \quad \frac{1}{N} = \sqrt{\sum_{i=1}^{N}(xi - yi)^2} \tag{13}$$

$$\text{Gini Coefficient} \quad Gini = 1\sum_{i=1}^{C}(p_i)^2 \tag{14}$$

$$\text{Entropy} \quad \text{Entropy} = \sum_{i=1}^{c} -p_i * \log_2(p_i) \tag{15}$$

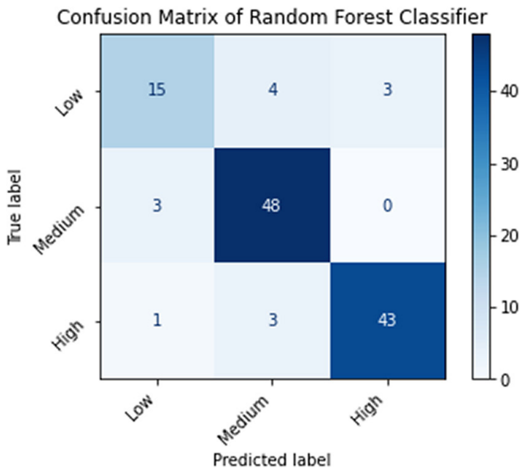

Fig. 7. Confusion matrix of RF algorithm

We apply the RF algorithm on our dataset and label the old data values. It predicts the value of data – where we tried to make our predictions fit the labels during preparation with the use of the RF approach. Figure 7 shows the confusion matrix. Moreover, the achieved model accuracy is 0.8833 and further discussed in Section 5 of Results.

4 Experimental setup

Python is a high-level scripting language that's mostly used for general-purpose programming and machine learning methods, as well as web development and database management. We have used the Anaconda Navigator –>Jupiter Notebook GUI framework from the Python tool. All the desired algorithms i.e., FCM, MLP, LR, and RF, have been simulated in Python programming language. The dataset is mostly used to assess student performance. This dataset has 480 tuples (rows) and 17 dimensional characteristics (columns). We simulate the programs of different combinations of the algorithms. The first program simulated FCM and MLP schemes, the second program simulated FCM and LR schemes, and the third program simulated FCM and RF schemes. The configuration of the computer used is as follows:

- Second Generation Intel (R) Core (TM) i5-2520M CPU @ 2.50 GHz.
- RAM 4.00 GB.
- The system is a 64-bit operating system.
- Windows 10 (Home).
- 500 GB hard disk.

5 Results and discussions

ML is a scientific approach in which CPUs learn how to resolve complications without being explicitly programmed. DL is now persuasive the ML competition, owing to better-quality processes, computation power, and immense datasets. Nonetheless, conventional ML algorithms hold a prominent place in the sector. This research employs a novel technique designed on the integration of FCM. The performance of the classifier was then compared against other supervised ML algorithms to choose the best classifier. In this study, ML supervised approaches have been considered which include MLP, LR, and RF (classification techniques), as well as FCM (clustering technique). Furthermore, we combine the algorithms to get the best accurate combination for predicting the student's performance. We have formed the different combinations which include FCM – MLP, FCM – LR, and FCM – RF, and simulated the programs. The achieved results are compared with the conventional algorithms such as K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Deep Neural Network (DNN). Tables 4 and 5 show the acquired accuracy and other performance characteristics. The observed results are also shown in Figures 8 and 9. It is apparent from the results that the combinations FCM – MLP, and FCM – LR are reached at their maximum outcome in terms of accuracy i.e., 95.833%. However, the combination FCM – RF is ranked second having accuracy of 88.33%.

Table 4. Model accuracy for the combination of algorithms for students performance prediction

Algorithms	Accuracy (%)	Algorithms	Accuracy (%)
K-Nearest Neighbor (KNN) [20]	69.00	Support Vector Machine (SVM) [20]	75.00
Random Forest (RF) [20]	79.00	Deep Neural Network (DNN) [20]	84.00
FCM – MLP **Proposed Method**.	95.833333	FCM – LR **Proposed Method**.	95.833333
FCM – RF **Proposed Method**.	88.333333		

Table 5. Parameter score for the combination of algorithms for students performance prediction

S/No.	Parameter Score (%)	FCM – MLP	FCM – LR	FCM – RF
1	Accuracy	0.95833333	0.95833333	0.88333333
2	Precision	0.95833333	0.95833333	0.88333333
3	Recall	0.95833333	0.95833333	0.88333333
4	Sensitivity	0.99555555	0.97142857	0.78947368
5	Specificity	0.97619047	0.99333333	0.94117647
6	F1-Score	0.95833333	0.95833333	0.88333333

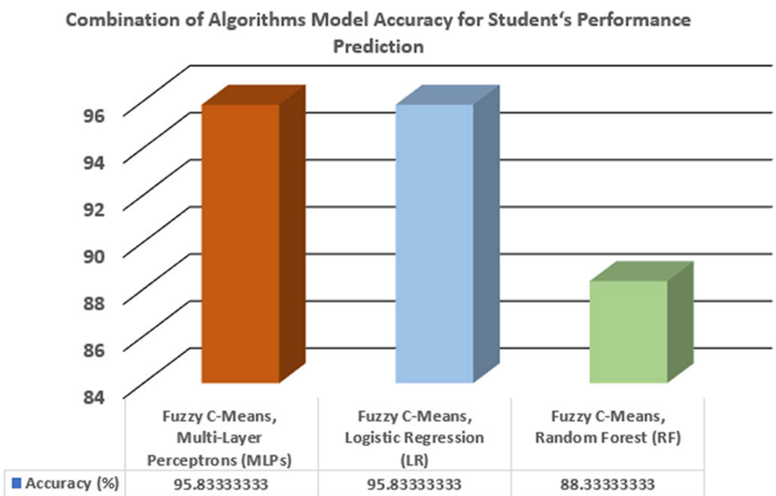

Fig. 8. Model accuracy for various combined algorithms for students' performance prediction

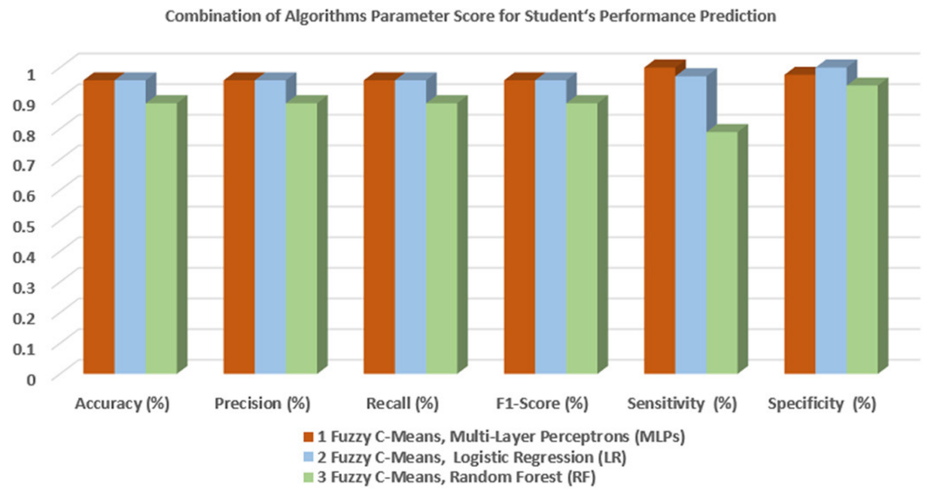

Fig. 9. Parameter score for various combinations of algorithms for students' performance prediction

6 Conclusion

The effectiveness of ML algorithms for predicting student performance has been examined in this study. The considered dataset of high education students in this study was evaluated using a range of algorithms including the fuzzy C-means (FCM), Multi-Layer Perceptron (MLP), Logistic Regression (LR), and Random Forest (RF). Initially, the dataset was preprocessed and clusters were formed using FCM technique. Later, the preprocessed data was classified by using a number of classification algorithms such as MLP, LR, and RF. The FCM technique has been particularly combined with each classification algorithm to improve the precision. As a result of combining each classification algorithm with the FCM, the achieved accuracy is 95.833%, 95.833%, and 88.333%, for the combinations FCM – MLP, FCM – RF, and FCM – LR, respectively. It is concluded that the combination of FCM with MLP, and FCM with LR yields the most accurate results.

7 References

[1] W. Zhu, L. Xie, J. Han, and X. Guo, "The application of deep learning in cancer prognosis prediction," Cancers, National Library of Medicine vol. 12, no. 3, p. 603, 2020. https://doi.org/10.3390/cancers12030603

[2] Machine Learning Mastery, https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/, accessed on 2021-12-23.

[3] B. Bakhshinategh, O. R. Zaiane, S. El-Atia, and D. Ipperciel, "Educational data mining applications and tasks: a survey of the last 10 years," Education and Information Technologies (Educ Inf Technol), vol. 23, no. 1, pp. 537–553, 2018. https://doi.org/10.1007/s10639-017-9616-z

[4] P. Chen, Y. Lu, V. W. Zheng, X. Chen, and B. Yang, "KnowEdu: a system to construct knowledge graph for education," IEEE Access, vol. 6, pp. 31553–31563, 2018. https://doi.org/10.1109/ACCESS.2018.2839607
[5] Q. E. Aissaoui, Y. E. A. E. Madani, L. Oughdir, A. Dakkak, and Y. E. Allioui, "A multiple linear regression-based approach to predict student performance," in Proceedings of the International Conference on Advanced Intelligent Systems for Sustainable Development, Springer, Cham, Switzerland, January 2020. https://doi.org/10.1007/978-3-030-36653-7_2
[6] Datavedas, Classification-Problems, https://www.datavedas.com/wp-content/uploads/2018/05/3.1.1.2-CLASSIFICATION-PROBLEMS-1.png/, accessed on 2021-01-01.
[7] Researchgate, Document Clustering, https://www.researchgate.net/profile/Laith-Aaligah/publication/322455242/figure/fig1/AS:582970224644096@1516002340891/An-example-of-the-document-clustering.png/, accessed on 2022-01-02.
[8] I. D. Livieris, et al., "Predicting students' performance using artificial neural networks," 8th PanHellenic Conference on International Participation Information and Communication Technologies, International Journal of Computer Science and Mobile computing pp. 321–328, 2012.
[9] S. Kotsiantis, et al., "Preventing student dropout in distance learning systems using machine learning techniques," Applied Artificial Intelligence, vol.18, no.5, pp. 411–426, 2003. https://doi.org/10.1080/08839510490442058
[10] C. E. Moucary, M. Khair, and W. Zakhem, "Improving student's performance using data clustering and neural networks in foreign-language based higher education," The Research Bulletin of Jordan ACM, vol. 2, no.3 , pp. 27–34, 2011.
[11] Yi, Hongsuk, H. Jung, and S. Bae, "Deep Neural Networks for traffic flow prediction," in IEEE – International Conference on Big Data and Smart Computing (BigComp), pp. 328–331, 2017. https://doi.org/10.1109/BIGCOMP.2017.7881687
[12] E. A. Amrieh, T. Hamtini, and I. Aljarah, "Mining educational data to predict student's academic performance using ensemble methods," International Journal of Database Theory and Application, vol. 9, no. 8, pp. 119–136, 2016. https://doi.org/10.14257/ijdta.2016.9.8.13
[13] S. E. Shaymaa and M. Tsunenori, "Building an interpretable model of predicting student performance using comment data mining," IEEE 5th International Congress on Advanced Applied Informatics (IIAI-AAI), 2016.
[14] N. Buniyamin, M. B. Usamah, and A. Pauziah Mohd, "Educational data mining for prediction and classification of engineering student's achievement." IEEE 7th International Conference on Engineering Education (ICEED), 2015. https://doi.org/10.1109/ICEED.2015.7451491
[15] M. N. Quadri and N. V. Kalyankar. "Drop out feature of student data for academic performance using decision tree techniques." Global Journal of Computer Science and Technology, 2010.
[16] C. Tjioe Marvin and A. Mewati, "Exploration of classification using NBTree for predicting students' performance," in IEEE 2014 International Conference on Data and Software Engineering (ICODSE), 2014.
[17] M. Wook, Y. H. Yahaya, N. Wahab, M. R. M. Isa, N. F. Awang, and H. Y. Seong, "Predicting NDUM student's academic performance using data mining techniques," 2009 Second International Conference on Computer and Electrical Engineering, 2009, pp. 357–361, https://doi.org/10.1109/ICCEE.2009.168
[18] K. P. Shaleena and S. Paul, "Data mining techniques for predicting student performance," 2015 IEEE International Conference on Engineering and Technology (ICETECH), 2015, pp. 1–3, https://doi.org/10.1109/ICETECH.2015.7275025
[19] Bendangnuksung and P. Prabu,"Students'performance prediction using deep neural network", International Journal of Applied Engineering Research, vol.13, no.2, pp. 1171–1176, 2018.

[20] V. Vijayalakshmi and K. Venkatachalapathy, "Comparison of predicting student's performance using machine learning algorithms" I.J. Intelligent Systems and Applications, 12, 34–45, 2019. https://doi.org/10.5815/ijisa.2019.12.04
[21] A Students' Academic Performance Dataset, https://www.kaggle.com/datasets/aljarah/xAPI-Edu-Data, accessed on 2022-04-22.
[22] Data Clustering Algorithms, fuzzy-c-means-clustering-algorithm, https://sites.google.com/site/dataclusteringalgorithms/fuzzy-c-means-clustering-algorithm, accessed on 2022-04-30.
[23] Multi-Layer Perceptrons (MLPs) Algorithm, https://wiki.pathmind.com/multilayer-perceptron, accessed on 2022-05-5.
[24] Towards Data Science, Logistic Regression, https://towardsdatascience.com/introduction-to-logistic-regression-66248243c148/, accessed on 2022-05-12.
[25] Random-Forest-in-Machine-Learning, https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/, accessed on 2022-04-18.
[26] Tutorials-Point-Random-Forest, https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_random_forest.htm, accessed on 2022-05-28.
[27] Random-Forest-Algorithm, https://medium.com/capital-one-tech/random-forest-algorithm-for-machine-learning-c4b2c8cc9feb, accessed on 2022-05-30

8 Authors

Mirza Azam Baig pursuing MS degree in Computer Science from Sindh Madressatul Islam University, Karachi, Pakistan. He received BS Software Engineering degree in 2017 from Baqai Medical University, Karachi, Pakistan. From 2019 to the present, he is working as a Software Engineer at Bahria Enterprise Systems & Technologies Pvt Ltd, Pakistan. His research interests include in Artificial Intelligence, the Internet of Things, and Deep Learning.

Dr. Sarmad Ahmed Shaikh was born in Pakistan. He received the Ph.D. degree in Information Technology from the University of Klagenfurt, Klagenfurt, Austria, in 2018. He is currently an Assistant Professor with the Faculty of Information Technology, Sindh Madressatul Islam University (SMIU), Karachi, Pakistan. His research interests include RF/microwave engineering, radio localization, antenna arrays designing, and wireless sensor networking (IoT) & artificial intelligence.

Dr. Kamlesh Kumar Khatri was born in Pakistan. He obtained Bachelors Degree in Computer Science from Shah Abdul Latif University, Khairpur, Pakistan, in 2006, the Master's Degree in Computer Science from SALU Khairpur, Pakistan, in 2010, and the PhD Degree in Computer Science & Technology from UESTC China in 2016. He is currently serving at Sindh Madressatul Islam University, Karachi, Pakistan, as an Assistant Professor in the Department of Software Engineering. His research interests include Image Processing, Computer Vision, Machine Learning, Knowledge Discovery and Data Mining, and Content Based Image Retrieval System.

Muneer Ahmed Shaikh was born in Pakistan. He received the Bachelors Degree in Computer Systems Engineering from Mehran University of Engineering & Technology Jamshoro, Pakistan, in 1994, the M.E. degree in Computer Systems Engineering from NED University, Karachi, Pakistan, in 2010, and the Ph.D. degree (Continued) in Computer Science from SMI University Karachi Pakistan. He is currently an Assistant

Professor in Computer Science, College Education Department, Government of Sindh, Karachi, Pakistan. His research interests include RF/microwave engineering, radio source localization, antenna arrays designing, and wireless sensor networking (IoT) & artificial intelligence.

Muhammad Zohaib Khan obtained MS degree in Computer Science from Sindh Madressatul Islam University, Karachi, Pakistan, in 2022. His research interests include Data Science, Artificial Intelligence, Machine Learning, Deep Learning, Medical Science & Internet of Things (IoT).

Mahira Abdul Rauf obtained MS degree in Computer Science from Sindh Madressatul Islam University, Karachi, Pakistan, in 2022. Her research interests include Machine Learning, Artificial Intelligence, Cloud Computing, the Internet of Things, and blockchain.

Article submitted 2022-09-14. Resubmitted 2022-11-15. Final acceptance 2022-11-15. Final version published as submitted by the authors.