Original Article

A Review of the Methods of Predicting Students' Performance Using Machine Learning Algorithms

Marzieh Babaie [1]*, Mahdi Shevidi Noushabadi [2]

[1]Department of Rey, Payame Noor University (PNU), Tehran, Iran. [2]Department of Educational Management, Faculty of Management, Center Tehran Branch, Islamic Azad University, Tehran, Iran.

Abstract

Education is a basic right of human where there should be various cases with equal quality and global access. Effective and efficient approaches have to be found to enhance education. Computers and portable devices emerged three decades ago and are used in all stages of daily life today. Artificial intelligence technology has been developed using computer tools to perform some tasks like simulating intelligent methods and problems solving by humans. Artificial intelligence has been considered as one of the most valuable uses in educational requrements. Machine learning is one of these in artificial intelligence, which could prove effective in predicting the pervasive situation in academic results, field selection as well as getting better scores given its different algorithms and methods. The paper intended to review and compare some of these algorithms by examining some papers and studies mainly condcuted in the last year. Decision tree, support vector machine (SVM), k-means clustering, and neural network could be mentioned among these algorithms.

Keywords: Prediction, decision tree, performance, SVM, machine learning

INTRODUCTION

Numerous and voluminous data have been produced and collected in all areas and have been organized and processed to obtain valuable information. In doing so, data mining techniques are presented and used to build dataset analysis model and identify the useful pattern in the data. Educational Data Mining (EDM) is an emerging field of data mining. The purpose of EDM is to collect data on the learners, their learning environment and new approaches to specify useful patterns. Education institutions have always been interested in collecting data about their students. Processing these data can be significant and identify the areas where it exists and the institutions that are in need of progress. This interest in data collection and processing has increased with the emergence of big data analysis, and online learning has increased the environment for collecting and analyzing data from learners and their environment and supporting and gaining insight into students' learning activities seems useful. One of the significant areas of using EDM is to develop models that predict student performance that predict student performance in education. Institutions are based on some basic factors presented as inputs and timely interventions could assist students enhance their performance.

Here, we intend to briefly review and examine the activities in the field of predicting using machine learning and its related algorithms and compare these algorithms to answer the following questions in this regard.

1. Which algorithms can be used in predicting student status?
2. Which algorithm does a better evaluation of the student?
3. Which algorithm has fewer errors?
4. How can these algorithms be compared in graphical view?

Summarizing

Here, we summarize several new papers written on predicting student performance using machine learning algorithms.

The first paper was written by Boran Sekeroglu in 2019 in Turkey, whose subject was predicting student performance and classifying it using machine learning algorithms [1].

Address for correspondence: Marzieh Babaie, Department of Rey, Payame Noor University (PNU), Tehran, Iran

How to cite this article: Babaie, M., Shevidi Noushabadi, M. A Review of the Methods of Predicting Students' Performance Using Machine Learning Algorithms. Arch Pharma Pract 2020;11(S2):69-79.

The problem presented in the paper is prediction and classification of students' performance and examinination of the effects of variables like social and family environment on student performance and improving the quality of teaching in teaching mathematics and Portuguese language.

The proposed solution of the paper is to use machine learning algorithms and evaluate the performance of raw data that is done using a selection or preparation algorithm.

This method has some advantages and disadvantages. Its advantages are 1) Artificial intelligence brings about intelligent techniques for understanding intelligent mechanisms and behavior 2) Artificial intelligence techniques can help educators enhance their teaching strategies. 3) It is believed that these types of programs can help teachers and students to explain uncertainties and ensure establishment of a similar and consistent concept. This helps educators to act in a timely manner and increase student performance and positive feedback. 4) It does not have the disparity of traditional methods like lecturing. 5) It is accompanied by immediate feedback. 6) It improves teacher-parent relationship.

Its disadvantages are the paper focusing on two subjects of mathematics and learning the Portuguese language and to use its algorithms, one has to collect a series of data like parental status, home address, family size, and so on that may not be available or maybe accessible with difficulty. 2) Some features may be similar and make categorization difficult.

The field results of this paper are: various machine learning algorithms are useful and effective for various types of problems and their prediction and classification. The paper uses four various algorithms for two types of problems: prediction and classification. The data fed to the algorithms are selected without any data algorithm and significant results are obtained in them.

In predicting the minimum value of the mean square and the highest value of R^2 and evaluation scores are obtained by SVR. Even the backpropagation (BP) algorithm had the lowest production in prediction rate, compared to other superior algorithms. In experiments, classification was used with 87.78% test ratio.

The second paper was written by Ijaz Khan et al. (2019) in Oman, entitled “Following up students' performance in introductory programming by machine learning.” [2] The problem is modeling a plan that can examine students (in the introductory programming course) about their possible outcomes in the early semester (when graded for 15% of grades).

In the proposed solution in this paper, 11 machine learning algorithms (from 5 categories) with more than one data source using WEKA are applied concluding that the decision tree (J48) shows higher accuracy in terms of correctly identified cases. The actual F-scale measurement and positive diagnosis of this study helps students identify their possible final scores and correct their academic status based on behavior to achieve higher scores.

Of the advantages of this method, one can cite the following: 1) It identifies useful information, 2) determines the points in need of progress, 3) helps educational institutions to create strategies to enhance the quality of the teaching process, 4) is useful in identifying students at low risk and low academic achievement, 5) this prediction is useful in issuing initial warnings to students with low academic performance, and timely interventions can help students improve their performance. Among its advantages, one can cite 1) as novice programmers have almost no idea about learning programming courses; they usually do not study very different topics during it. 2) They are more likely to make mistakes at school and face difficulties that often lead to their failure to score higher. The paper concludes that tree family algorithms have a high accuracy of J48 to 88% and have reached the most appropriate result.

The study helps students and teachers to improve student performance. The study is used to identify the students in need of special attention, as well as to reduce the failure ratio and take appropriate action for the next semester exam.

In the third paper, which was written by Ita Borman et al. (2019) at Indian universities, the following can be deduced [3]. The problem stated in the paper is that it helps students improve their performance using data mining programs, which take advantage of students' psychological characteristics. This paper uses a multi-class support vector machine (SVM). The proposed approach to this paper is to analyze a study in India using logistical regression that showed non-academic correlation on learners' CGPI. Affect. Psychology is a field concerned with behavior and the mind and records these actions.

The advantages of this paper are:

1. It identifies slow learners in the classroom. 2) It helps enhance performance early in learning. 3) It is used for selection in student employment. 4) It uses psychological factors to predict student progress. Its disadvantages are 1) there is some generalized error that uses the largest margin to reduce it (The margin is the maximum distance between the nearest data points of the class.) 2) It only deals with students' non-mental parameters.

The results of this research are as follows: various extraction methods like neural network, decision tree, KNN, New Bayesian and SVM are used on educational data that covers psychological factors. The accuracy of previous studies is less than 89%. Our proposed model uses SVM for data classification and CGP prediction of learners.

According to the statistics in the relevant table, radial basis function kernel gives more accurate results relative to the linear core, which is approximately 90%.

The last paper, written by Mukash Kumar (2019) in India, explores the issue of student performance analysis using machine learning and data mining [4].
The problem is that many traditional evaluation steps may put a lot of pressure on the student. Keeping the student stressed out does not make the student perform well. The proposed solutions of the paper are estimating the performance of students in the university is divided into two different parts:
A) Data reconstruction and design formation using threshold division architecture (DRLFT) and B) Analysis of the data divided for accurate prediction. The paper and its method have advantages and disadvantages. Its advantages are 1) ensuring that students' performance is evaluated based on the actual pattern, 2) identifying mean square of the error, the difference between the prediction and the actual analysis, 3) compared to SVM, artificial neural network (ANN) performs better in terms of error rate production. There is a lot of effort. Moreover, its disadvantages are 1) As the storage increases, the average value of the error squares increases and 2) The average square error coefficient of the SVM algorithm is always higher than ANN because the number of recordings increases each time. The conclusion from this algorithm can be that a combination of K-Means clustering algorithm with ANN and SVM classification algorithm has been proposed to evaluate students' performance to reduce human effort. Evaluation is based on the mean error of the squares and the estimated effort.

Method

In each of the papers examined, a specific method has been used to review and implement the algorithms, we will review here.

The first paper:

In the first paper, Boran Sekeroglu et al. considered two sets of data; the Student Performance Database (SPD) and the Student Academic Performance Database (SAPD) [1].

Experiments are divided into two categories according to their outputs, including predicting and classification experiments. SPD is used for prediction and SAPD for classification experiment.

Table 1: Math prediction results with all algorithms and the test ratios

Algorithm	40% of Testing Ratio		
	MSE	R^2	EV score
SVR	0.013	0.785	0.797
LSTM	0.014	0.775	0.779
BP	0.019	0.694	0.700
Algorithm	**30% of Testing Ratio**		
	MSE	**R^2**	**EV score**
SVR	0.016	0.757	0.775
LSTM	0.019	0.727	0.742
BP	0.024	0.646	0.680

Table 2: The results of prediction for the Portuguese language course from all tables of algorithms and test ratios

Algorithm	40% of Testing Ratio		
	MSE	R^2	EV score
SVR	0.0054	0.834	0.838
LSTM	0.0055	0.824	0.839
BP	0.0085	0.728	0.733
Algorithm	**30% of Testing Ratio**		
	MSE	**R^2**	**EV score**
SVR	0.0046	0.827	0.828
LSTM	0.0063	0.765	0.765
BP	0.0079	0.708	0.708

Table 3: The results of classification for all algorithms and test ratios

Algorithm	40% of Testing Ratio
	Classification Accuracy
BP	80.91%
SVM	79.38%
GBC	74.04%
Algorithm	**30% of Testing Ratio**
	Classification Accuracy
BP	87.78%
SVM	83.20%
GBC	82.44%

SPD includes student performance as an output for mathematics and Portuguese language courses, which is based on 33 features related to parental status, home address, family size, and so on.

SPAD has relatively similar features to SPD, yet has 21 full features and 3 outputs: good, medium and weak.

- #### Prediction experiments

In prediction experiments, BP, SVR and LSTM algorithms are used, which are in 40% and 30% ratios in the samples and the evaluation is done according to 3 criteria as the mean square of error (MSE), R^2 score and explanatory variable (EV) that are the main indices of success of the predicted results.

In the first group of prediction experiments, 40% test ratio (60% training for all subjects) is used for the math course and the optimal scores for MSE, R^2, and EV scores are obtained by SVR and followed by LSTM.

The lowest prediction rate by BP is for all evaluation criteria in this experimental group.

For 30% ratio of the second math test, similar results were obtained like 40% of the results, which are the test ratio and the highest prediction rate by SVR and the lowest rate by BP.

In the second group of prediction experiments, the students' performance is analyzed and predicted in the Portuguese language course. Firstly, 40% of the test ratio in the math course, minimum MSE and maximum R^2 scores are considered, where the highest performance was obtained by SVR followed by LSTM.

However, considering EV criteria, one can understand that both SVR and LSTM produce similar results, but the maximum speed has been obtained by LSTM with the lowest prediction rate obtained by BP like the first group of prediction experiments.

The second experiemnt of the second group in prediction includes 30% of the ratio of experiments in the Portuguese language course, where the highest prediction and the lowest error rate are obtained by SVR, followed by LSTM and BP, respectively. Tables 1 and 2 show the math and Portuguese courses results, which, respectively, are the highest and lowest prediction results for the ratio. The test shows 30% and 40% in the results obtained.

- ## Classification experiment

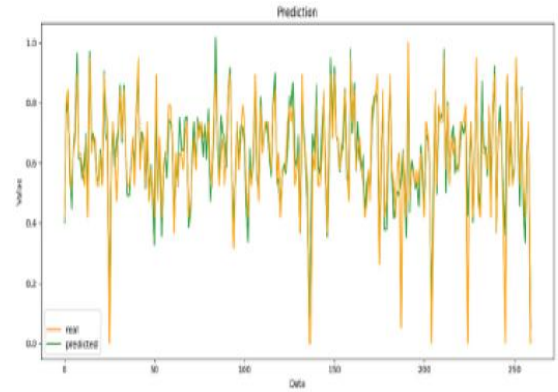
In these experiments, BP, SVM, and GBC algorithms were used for 40% and 30% test ratios with SPAD, and the evaluation according to the Accuaracy result as far as possible is obtained from the following equation.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

Here, TP and TN are true positive and true negative, respectively, and FP and FN false positive and false negative values and the results are accordingly.

Cosidering 40% of the sample test ratio, BP, SVM and GBC have 80.91%, 79.38% and 74.04% of the classification rate for test and training data, respectively.

For 30% test ratio, higher results are obtained for BP, SVM, where GBC was 87.78, 83.20 and 82.44, respectively. Table 3 shows all the results of the classification experiment.



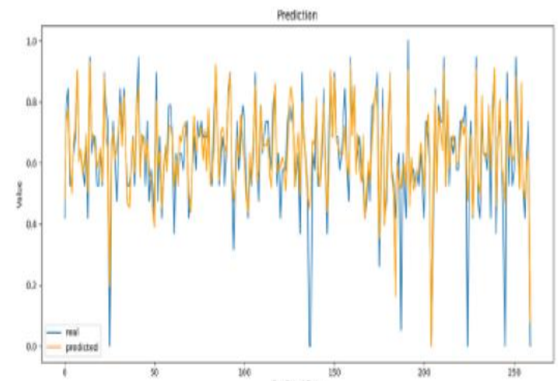


Figure 1: A sample of prediction charts a) the highest prediction rate by SVR and b) the lowest prediction rate by BP for 40% test ratio in the Portuguese language course

Second paper

In this paper, Ijaz Khan et al. have focused on students' ability in the introductory programming lesson. Here, various machine learning algorithms are used.

According to them, machine learning (ML) is a branch of artificial intelligence that may bring about learning from past experience (previous data) to enhance future performance automatically regardless of human help. More formally, ML is "A computer program that states for learning from experience measures E according to some classes with task T and P performance. If its performance in tasks T as measured by P, it enahnces with E experience." Prediction models not only examine students during their studies, but also provide real-time counseling to correct their problems. The main difference between a computer and a human is in the ability to think and learn. Computers cannot learn from experience. They run human-made learning algorithms to learn from it, and the predictions are made based on that.

No effective ML algorithms have been found to be effective for some types of learning tasks. They are especially useful when a person may not have sufficient and useful knowledge and it is difficult to develop effective knowledge engineering algorithms.

ML algorithms receive input data to train and generate hypotheses. This training makes it possible for them to predict the future. This section will briefly eplore some of ML algorithms widely used.

A) Naive Bayes algorithm:
NB is a classification method based on the Bayesian theorem. This classification algorithm is simple and efficient. This effectiveness depends on the independence of class conditions, assuming that the effect of a value of a property on class A is given by the values of other independent features and (b) assuming that none of the hidden features can affect the forecasting process.

B) Multilayer perceptron (MLP):
MLP is one of the most common uses of NNAs. This is especially true when the relationship between input and output features is ambiguous, called classification approximation.

C) SVM:
SVM extracts its features from variables and manages them in linear composition by achieving a prediction (or classification or regression) decision.

D) Decision tree:
Decision tree uses a propagation approach to build the tree and achieve maximum prediction accuracy. Decision tree uses different mathematical algorithms like the Gini index, information enhancement, and Chi-square statistics.
Decision tree starts with a root node, and ends with branching from the middle nodes (called leaf nodes) to the last nodes (called end nodes). Some popular decision tree algorithms are ID3, C4.5, C5 and shopping cart.

Third paper:
In the third paper, Ita Burman et al. uses a SVM algorithm to predict student performance.

The steps of the proposed method are as follows:
A. SVM
SVM is a supervised learning technique to classify data. It uses hyperplane to divide the dataset into as many distance classes as possible, which is called margins and creates parallel lines to create partitions. Margin is the maximum distance between the closest data points of the classes. The largest margin is selected to reduce the generalization error.

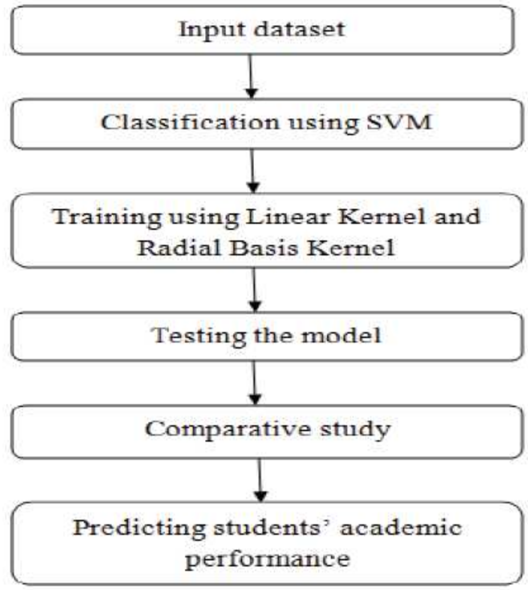


Figure 2: Suggested model

SVM algorithm acts as follows:
Decomposable case is one of the data that can be quite suitable for linear analysis. Here, there are unlimited borders that may select the optimal hyperplane with the maximum border distance. The stated function is as follows:

$f(y) = x.y + z$ (2)

SVM divides the data points as follows:

$f(y) > 0$, iff $y \in X$, and (3)
$f(y) <= 0$, iff $y \in Z$ (4)

The distance between observation and hyperplane by x ||| x.y + z | / | | is given and the margin is as 2 / || x ||.
2. The indecomposable case is where the data points overlap to classify these data points. SVM performs data reconstruction using the conversion function as (Φ). It works from the scalar points of the data point with a higher dimension, which allows for sufficient linear separation.

- ## Analysis

The most significant and common task of machine learning is classification that could be done with various data mining techniques. The paper classifies student data based on psychological components into three categories including high, medium, and low. Our problem is a multi-classifier problem. This leads to the use of linear kernel and radial base functions.

The dataset includes the students' files collected using a questionnaire according to psychological parameters, involving personality, motivation, social psychological effect, learning strategies, learning approach and socioeconomic status. The dataset consists of a thousand files based on 29 non-mental structures of students. We divided the dataset as 70% for training the model data and 30% for testing the rest of the dataset.

The linear kernel can be defined as the product of the interior point [a, b] and a desired constant c, which can be mathematically as follows:

$$k(a, b) = a^T b + c \qquad (5)$$

radial basis function (RBF) can be defined as follows:

$$K(x,x') = \exp[(-\|x-x'\|^2) * y] \qquad (6)$$

Here, x and x' are the two input feature vectors, $\|x-x'\|^2$ is squares Euclidean distance and y is calculated as follows: 1.2 $\varnothing^2$

The value of RBF kernel can be used as a similar criterion that varies between 0 and 1, which decreases with distance.

Results and Discussions

The proposed model for predicting student performance is evaluated using sensitivity, specificity, and accuracy.

The results of the training dataset were using different cores as shown in Figures 2 and 3 and the results of data testing are discussed in Tables 1 and 2.

A. Sensitivity

This statistical measure was measure by applying the positive values (like our study, correct identification of students in the higher, middle, and lower classes to the given parameters).

$$Sensitivity = \frac{TP}{TP+FN} \times 100 \qquad (7)$$

B) Specificity

It measures negative or incorrect values.

$$Specificity = \frac{TN}{TN+FP} \times 100$$

C) Accuracy

Bayesian t is a statistic that measures accuracy: the difference between the observed and the actual values.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100$$

Here, TP shows the number of true positives, FP false positives, TN true negatives, and FN false negatives.

```
> svm_Linear
Support Vector Machines with Linear Kernel

701 samples
 29 predictor
  3 classes: 'H', 'L', 'M'

Pre-processing: centered (29), scaled (29)
Resampling: Cross-Validated (10 fold, repeated 3 times)
Summary of sample sizes: 631, 630, 630, 631, 632, 630, ...
Resampling results:

  Accuracy   Kappa
  0.6310801  0.29116

Tuning parameter 'C' was held constant at a value of 1
```

Figure 3: The results of training datasets using linear core

```
> svm_Radial
Support Vector Machines with Radial Basis Function Kernel

701 samples
 29 predictor
  3 classes: 'H', 'L', 'M'

Pre-processing: centered (29), scaled (29)
Resampling: Cross-Validated (10 fold, repeated 3 times)
Summary of sample sizes: 630, 630, 631, 632, 633, 631, ...
Resampling results across tuning parameters:

  C       Accuracy   Kappa
    0.25  0.6362190  0.2779799
    0.50  0.6723807  0.3532726
    1.00  0.7236603  0.4583860
    2.00  0.7946017  0.5994590
    4.00  0.8493637  0.7082804
    8.00  0.8782999  0.7657433
   16.00  0.8849666  0.7789071
   32.00  0.8972937  0.8025902
   64.00  0.8992262  0.8064044
  128.00  0.8992262  0.8064044

Tuning parameter 'sigma' was held constant at a value of 0.02206044
Accuracy was used to select the optimal model using the largest value.
The final values used for the model were sigma = 0.02206044 and C = 64.
```

Figure 4: The results of the training dataset using radial basis core

Table 4: Complexity matrix

iction	Linear Kernel			Radial Basis Function Kernel		
	H	L	A	H	L	A
H	100	0	52	140	0	14
L	0	6	0	0	6	0
A	53	0	88	13	0	126

Table 5: General Statistics

Statistics	Linear Kernel			Radial Basis		
Accuracy	0.6488			0.9097		
95% CI	0.5918, 0.7029			0.8713, 0.9396		
p-value	1.158e-06			2.2e-16		
	High	Low	Average	High	Low	Average
Sensitivity	0.6536	1.0000	0.6286	0.9150	1.0000	0.9000
Specificity	0.6438	1.0000	0.6667	0.9041	1.0000	0.9182

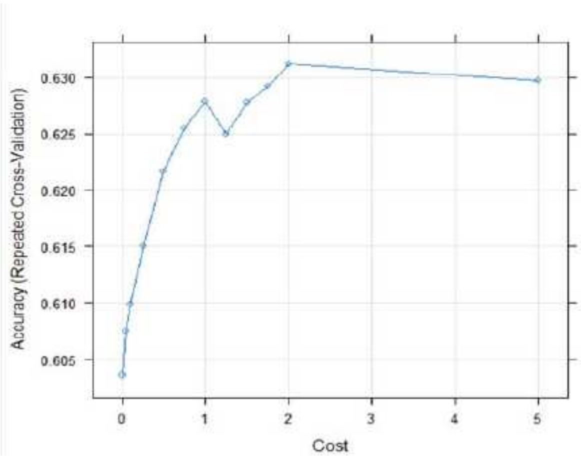


Figure 5: Using SVM from the linear kernel

Setting the model is done using a grid with various values from 0 to 5, with the linear kernel providing the best values by setting the parameter C = 2, as shown in Figure 4.

When the model is run using the Radial Grid Kernel core, the changes are calculated and the best Sigma results are 0.1 and C are 2, as is seen in the figure below.

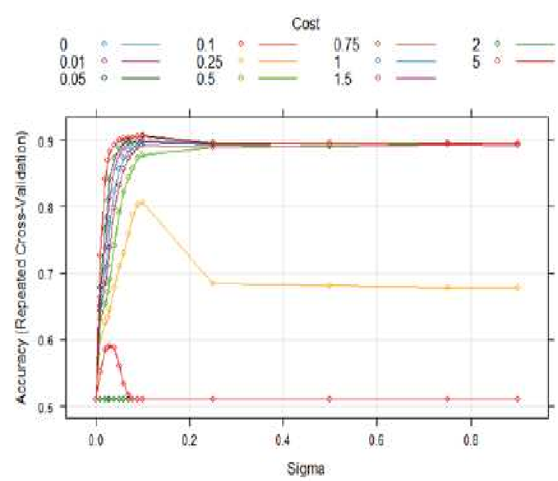


Figure 6: Using SVM from radial grid kernel

Fourth paper:

In the fourth paper, Mokash Kumar et al. have used the dataset used to run the site http://invertory.data.gov/dataset. The dataset includes various school / college records and various topics. However, the problem is that the data existing in the dataset is not structured to be used in MATLAB. The data and analysis of students where machine learning and data mining approach are used are as follows.

Figure 7: Unstructured dataset

Data in the dataset are in need of being reconstructed and reconstructing is done using the following algorithmic structure.

Algorithm 1: Reconstruct_Data function (un_st)

1. // un_st is the unstructured data.

2. sub_codes = find (un_st = = sub_code); // the data is getting arranged on the basis of the subject codes.

3. data_st = [ ]; // It is the structured set of the data.

4. data_st[count, sub_data] = sub_code.find(data (un_st)); // Finding the details of the subject out of the given series of data

5. count = count+1;

6. end for

7. end function

Algorithm 1 organizes the data in the dataset into a structured format according to the subject code data. The same above algorithm structure can be used to sort data according to school / college name.

Algorithm 2: reconstruct_data_college function (un_st)

1. coll_code = find(un_st == coll_code);// coll_code is the school code

2. col_count = 1;

3. For each col in coll_code

4. data_st [col_count. coll_code.st] = col_data; // Filling the school information into the school structure

5. col_count = col_count + 1;

6. end for

7. end function

Algorithm 2 uses data from the dataset in a structured format according to the school / college code. By performing the two algorithms above in the dataset, the data are placed in form of structural data, which is really needed for the purpose of our implementation.

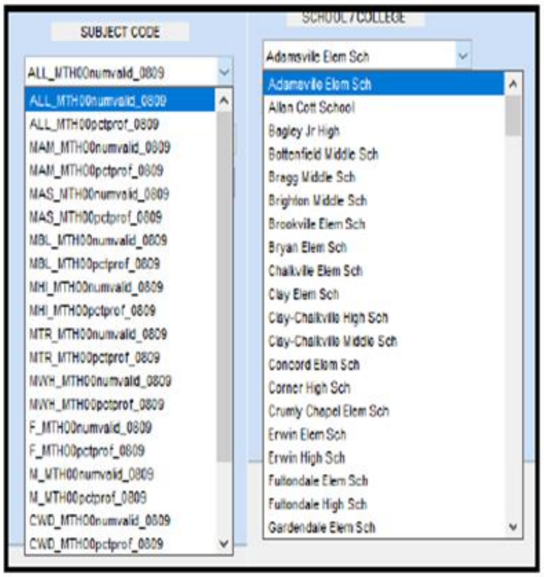


Figure 8: Schools / colleges and separate data

For the subject code of the association besides the school name, K_means clustering algorithm is applied according to the threshold division.

Algorithm 3: Apply Threshold_Segmentation_Kmeans (st.data):

1. For each class label in label properties

2. X (1) = class.label.name;// Collecting the school/college and class related data

3. X (2) = class.label.value

4. end for

5. end Function

After running algorithm 3 in the above structured dataset, the following output values are obtained.

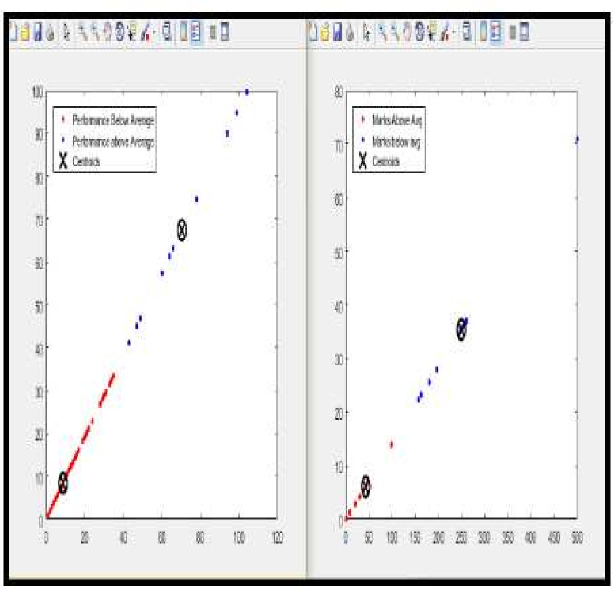


Figure 9: A) Subject code K_mean, B) K_mean school code

Figure 9 shows K_mean clustering algorithm following running in the dataset. Figure 13 A) presents the clustered data based on the subject code and Figure 3 B) the grouped data based on the school code used in the dataset.

Analysis of divided data for accurate prediction:

K_means clustering algorithm not only presents the grouped elements together, but marks the center of their parts. Data have been classified into two various sections, "below-average performance of the students" and "above-average student performance of the students." Cluster elements are confirmed by ANN and SVM. The validation structure is as follows:

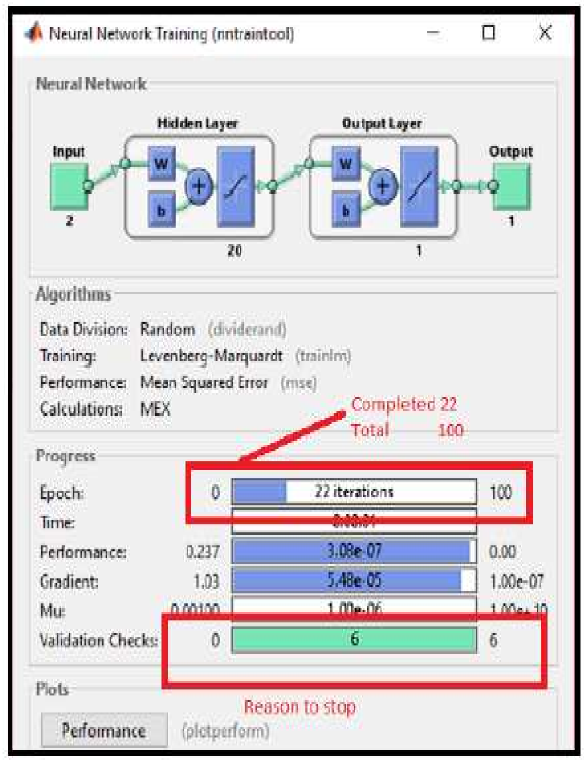


Figure 10: ANN architecture training

ANN classification algorithm trains on the data learned from the dataset based on the divided values of K-Means. This specific classification algorithm reveals some validation factors like mean squared error (MSE) as Figure 10 shows. Hence, our evaluation criteria for this dataset are according to MSE for ANN.

In Figure 11, the same type of training dataset is supported as input to implement SVM algorithm, and the following findings are found.

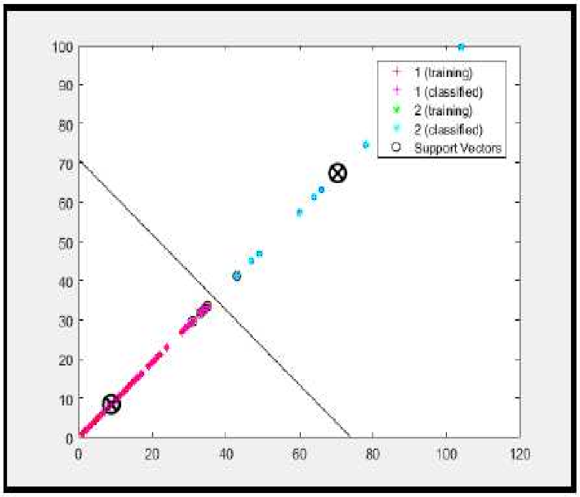


Figure 11: SVM training and classification

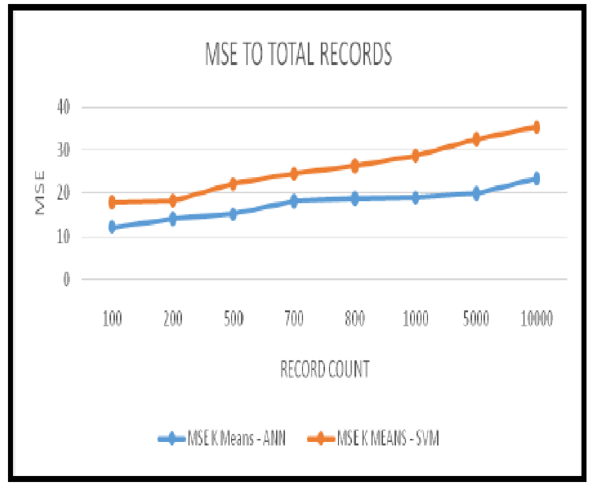


Figure 12: MSE against the number of records

The structure of training and classification of ANN and SVM ensures that the students' performance is evaluated based on the actual pattern.

- ## Analysis of the result

The results of running the above are evaluated according to mean squared error and the effectiveness of the effort required for evaluation. Overall, 10,000 files are recorded and the following results are evaluated.

Mean squared error: MSEs are the difference between the actual prediction and the analysis. Table 1 shows the values of 10,000 records.

Table 6: MSE result using k_means clustering for ANN and SVM algorithms

Record Count	Mean Square Error K Means - ANN	Mean Square Error K Means – SVM
100	12.24	17.96
200	14.14	18.35
500	15.23	22.21
700	18.21	24.53
800	18.78	26.45
1000	19.02	28.69
5000	20.03	32.45
10000	23.5	35.36

Table 6 indicates that as the storage values increase, MSE value increases. Thus, the above implementation shows that with the increase in the number of records, the error rate increase as well. Compared to SVM, ANN has better efficiency in terms of creating error. The overall difference in creating error between ANN and SVM is about 20%. Figure 6 shows the graphical index of the data in Table 4.

Figure 12 clearly shows MSE coefficient of SVM algorithm is always higher than ANN as the number of records increases each time. The difference in error rate is more than 10,000 or maximum records number.

Effort evaluation: The second parameter of judgment is the effort needed to evaluate the student's result. Below is the formula for calculating the estimated effort:

Total Effort = ((1-overallmse) * 100) - (overallmse * group_element_count);

Estimating the total effort in the correct evaluation and the effective effort in grouping the elements to helps evaluate the students. Table 3 shows the total effort values used for both algorithms.

Table 7: The needed estimation using k-means clustering for ANN and SVM algorithms

Record Count	Effort Estimation K Means - ANN	Effort Estimation K Means - SVM
100	32	47
200	38	48
500	39	49.36
700	39.5	49.87
800	40	53.56
1000	41	55
5000	41.3	58
10000	42	69

Table 7 has indicated that with increase in the value of the record count, the value of the estimation effort increases. Thus, the above implementation shows that with increase in the number of records, effort estimation will be high. Compared to SVM, ANN increases in terms of the estimated value of the effort. The overall difference in producing the value of the effort estimate between ANN and SVM is about

27%. Figure 7 shows a graphical profile of the data shown in Table 3.

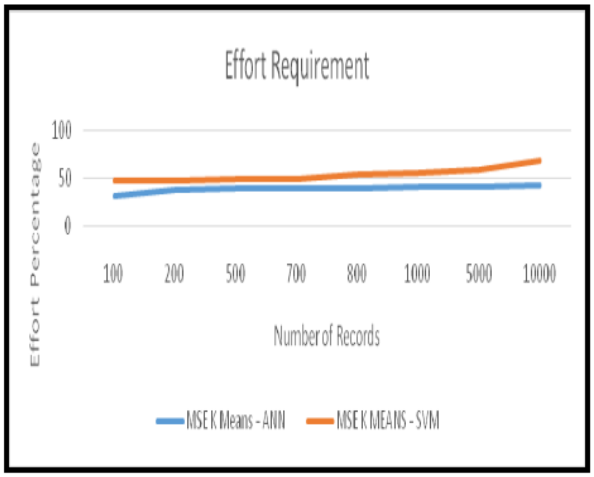


Figure 13: The graphical profile index of Table 3

Figure 13 clearly shows that the estimated value of SVM algorithm is always higher than that of ANN as the number of records increases each time. The difference in error rate is more than 10,000 or maximum record number.

4. Answer to some of the questions raised at the beginning of the paper:
At the beginning of the paper, we posed some questions about learning algorithms, some of which we review and summarize, so that we can finally summarize them and get useful results from them.

In the first question, we said which machine learning algorithms can help us in this regard. As was seen in the previous chapters, there were many algorithms like decision tree, SVM, ANNs, back propagation, and so on that helped us predict the performance of university students or students in different courses like mathematics, Portuguese language, introductory programming, and so on, and even the authors could use them to predict university students' future results.
The next question was which algorithms performed better, which is dealt with here according to the results of the above papers.

According to a paper by Boran Sekeroglu, four various algorithms are used for two types of problems, prediction and classification [1]. The data fed to the algorithms are selected without any data algorithm and significant results are obtained in them.

In predicting minimum MSE value and the highest value of R^2 and EV scores are obtained by SVR. Even BP had the lowest production in the prediction rate, and the other top-class algorithms in the classification experiments were used at 87.78% test ratio.

His studies indicate that any training data can be predicted or classified by machine learning algorithms, and the results obtained may be improved by considering different types of data selection and machine learning algorithms.

According to a second paper by Ijaz Khan et al., 11 classification algorithms were used on a data source using WEKA.

He concluded that tree family algorithms of the decision tree with high accuracy with J48 reached 88% and reached the most appropriate result.

This study helps students and teachers to enhance student performance. Moreover, the study is used to identify the students in need of special attention, as well as to reduce the failure ratio and take proper action for the next semester exam.

The third paper, which has been reviewed by Ita Burman et al., has dealt with students' non-mental parameters that affect their study and academic growth.

Psychological analysis of students' behavior in learning helps enhance their academic performance. Different extraction approaches like ANN, decision tree, KNN, New Bayesian, and SVM have been used on training data that cover psychological factors. As discussed in this paper, the accuracy of previous studies is less than 89%. The model proposed in the paper uses an SVM to classify data and predict CGP of the learners.

According to the statistics in Table 2, radial basis function kernel provides accurate results compared to the linear kernel, which is approximately 90%.

The last paper, by Mukesh Kumar et al., has proposed a combination of k-means clustering algorithm with an ANN and SVM classification algorithm to evaluate students' performance to reduce human effort. Evaluation is based on the mean error of the squares and the estimated effort. Evaluation is based on MSE and the estimated effort. The results of the implementation show that the performance of ANN is better than that of SVM. SME is 5-20% better, whereas the estimated effort is about 15-27% better.

The other questions raised in the paper were answered well and comparison of the algorithms above and the efficiency and error of each were dealt with, with graphic profile of many of them given in the third part, method.

Conclusion

The paper reviewed some of the studies and research done over the past year regarding using machine learning algorithms in predicting the performance of various learners. Later on, we identified the implementation and running them along with their advantages and disadvantages.

Efforts and studies of this kind in the education system can be effective in enhancing the performance of learners and reduce time and money waste in the education system.

However, some parameters and variables used in these algorithms are effective in the result of prediction, and if more signifiant and more parameters are utilized and this is done with correct and accurate psychology, better and more accurate prediction will be obtained with fewer errors.

Future studies

According to each of the studies examined, we can recommned various future studies like in ANNs, one can change the total number of neurons or change the satisfactory parameters.

In line with second paper written by Ijaz Khan et al., one can present an integrated prediction model in Java language that follows student performance during the semester and informs them in case of the risk of failure. Another line of research can be the studies with a larger dataset in the future as the selected dataset and experiments were small.

In the study by Boran Sekeroglu, future studies can involve the implementation of machine learning in prediction and classification of tasks with more datasets [1].

References

1. Sekeroglu B, Dimililer K, Tuncal K. Student performance prediction and classification using machine learning algorithms. InProceedings of the 2019 8th International Conference on Educational and Information Technology 2019 Mar 2 (pp. 7-11).
2. Khan I, Al Sadiri A, Ahmad AR, Jabeur N. Tracking student performance in introductory programming by means of machine learning. In2019 4th MEC International Conference on Big Data and Smart City (ICBDSC) 2019 Jan 15 (pp. 1-6). IEEE.
3. Burman I, Som S. Predicting students academic performance using support vector machine. In2019 Amity International Conference on Artificial Intelligence (AICAI) 2019 Feb 4 (pp. 756-759). IEEE.
4. Kumar, M., Singh, A. J. Performance Analysis of Students Using Machine Learning & Data Mining Approach, 2019.