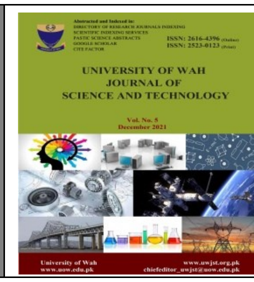
UW Journal of Science and Technology Vol. 5 (2021) 63-71
ISSN: 2523-0123 (Print) 2616-4396 (Online)

University of Wah
Journal of Science and Technology

www.uow.edu.pk



Prediction of Punjab Examination Commission (PEC) Student's Performance Using Educational Data Mining Approaches

Hira Afzal[1], Fareeha Anwar[2]

***Abstract*— Predicting the student performance is integral in the field of education. Conducted studies mainly focused on the higher education area due to its importance and validity but recognizing problems of students in early stages can be beneficial in the long run. Detecting problems at secondary level not only reduces the rate of students' failure in primary stages but also can be converted into teacher's pedagogical support. In Pakistan predicting students' performance at primary and secondary level have not been explored yet. Even though the province of Punjab conducts yearly assessment of students which can be utilized to study students' performance and behavior. This study focused on the primary and secondary level students' data who attended Punjab Examination Commission assessment, to predict their performance using educational data mining techniques. Dataset is created by collecting data from schools. Based on precision, accuracy and time taken to execute the model decision tree J48 outperforms others with accuracy of 99.3% with minimum execution time. The most significant factors which contributed to students' downfall were low attendance and lack of understanding of certain subjects.**

***Index Terms*— Educational Data Mining, Attribute Reduction, Decision Trees, Naïve Bayes, MLP, SVM.**

I. Introduction

DATA mining concedes the handlers to have an understanding of the data and the mined information can be helpful in making appropriate decisions. Educational Data Mining EDM is a powerful technique to predict and analyze student performance for research purposes and also in improving quality of education [1]. Educational data mining has surfaced as an independent research area in recent years and has been expanding due to its effectiveness and accurate predictions towards students and learning systems. Many educational institutes and school managements today, try unique and effective methods to advance their students' progress. It is desired to enhance number of students accepted in the yearly academics

Unfortunately, in Pakistan analyzing students' performance at school level has not been explored to its full extent. Punjab Examination Commission (PEC) conducts a yearly PEC exam for grade 5th and grade 8th students. The main objective of this exam is to prepare students for a more comprehensive examination that has yet to come. If thoroughly studied, these student data can be utilized by educational institutes to comprehend the conduct of a student. The purpose of this study is to predict students' performance and factors which affect it the most. The information extracted can be further valuable for the educators and heads of the institute, so they can take appropriate measures and accommodations to surge proficiency of students and improve the educational system overall.

II. Related Work

Analyzing and improving students' performance has always been the main concern of the educational institutes. Predicting students' performance not only improves the quality of education but it can also help identify the risks of failure and dropouts amongst students. Educational data mining is rapidly gaining popularity and is not only being used to improve students' performance but also to understand student behavior and patterns of failure.

Manuscript received: September 22,2021; Accepted: December 27, 2021

hiraafzal@outlook.com 1 ; fareehaanwar@iiu.edu.pk 2
Corresponding Authror : Fareeha Anwar
Affiliation : International Islamic Univeristy Islamabad Pakistan

A country's development can be determined by the quality of its education system. Day by day the education system is improving throughout the world. The researchers build a performance prediction model based on classification techniques. Attributes of this study were students' social interaction, academic integration and various emotional skills such as assertion, leadership, stress management etc. it was found that the result of the previous semester majorly influenced the next semester.[2], [2]–[5].

From related work, we assembled that the main data mining approaches used in most studies are: Classification and Clustering. Educational data mining was researched from three different aspects: 1) Students' performance prediction, 2) Students' failure prediction, 3) Pedagogical performance prediction. In general, the applied data mining approach is classification. A few researchers have used clustering and also a hybrid approach (classification + clustering) [3], [6]. Feature/attribute selection is also used in few papers to improve performance of a model [4], [5]. For predicting students' performance, the most effective and widely performed classification techniques are artificial neural networks, support vector machines, naïve Bayes and decision trees as they give best performance depending on the data [4], [7], [8]. Although there isn't enough material on the detection of performance factors at early stages of education. The foremost target of these studies has been the undergraduate/undergraduate students. Data set of these studies are primarily composed of higher education students. It is crucial to find problems that students face at early stages and vital to resolve them to pursue the education system. Punjab Examination Commission gives extensive data of students at secondary and primary level. Hence, given is the model to predict PEC students' performance and factors that affect their performance by using most efficient and commonly used mining techniques congregated from previous studies which are decision trees, artificial neural networks and support vector machine. This study also combined the classification technique with attribute selection/reduction to get better understanding of the results.

III. PROPOSED SYSTEM

This section provides a proposed approach using most commonly used data mining techniques, which are selected based on the findings of the related work in the previous section. The data mining techniques are applied to the dataset gathered from various schools of Punjab. The dataset is based on the PEC students' and the extracted attributes also compliant with the previous studies. The last step of this DM implementation is analyzing the obtained results and findings and compare it with the literature

The methodology is based on four steps: Data Collection, Data Preprocessing, Classification using Data mining and Result Comparison. Once the dataset is collected it is manually stored in the MS Excel spreadsheets, data preprocessing will transform the data into suitable format. Further in this stage data will be observed for other imbalances and errors. Data mining process is divided into two phases and is discussed in detail. At the end the result will be compared in between the phases and with the previous studies in the literature. Detailed flow of the proposed model is shown in Fig. 1.

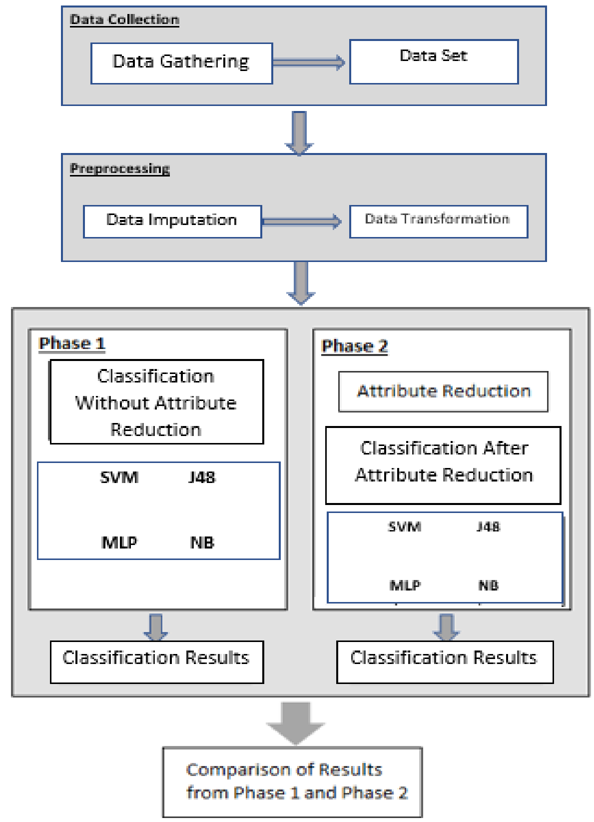


Fig. 1. Method Proposed for Improving Students' Performance

A. Data Collection

The data in this study is collected from various schools of Rawalpindi district, Punjab. Data includes the students who appeared in PEC (Punjab Examination Commission) examination at primary and middle level. Punjab Examination Commission (PEC) is an independent association setup by the Government of Punjab to evaluate the students' learning achievements specifically of grade 5 and 8. Punjab Examination Commission conducts an annual exam of the students of grade 5 and grade 8. In this step the data was gathered and processed into an actual data set. Data collection was done in two steps i.e., data gathering and data set preparation.

A gazette is provided by the PEC which only contains data of schools of the district and students' total marks. It did not provide the necessary attributes of the students which were needed for this study. In this scenario, the data should be collected from the schools manually. In the schools, data is stored in the class registers and PEC registers. PEC registers only contain the student evaluation related attributes (i.e., marks and percentage etc.) and general attributes can be taken from the class register. First of all, the attributes of data were defined on the basis of which data was collected. The attributes were defined with help of previous studies and were divided into two categories i.e, General Attributes and Academic Attributes (as shown in table 1). There are 17 attributes from which 5 attributes belong to general attributes and 12 attributes belong to academic attributes. Table 1 shows the detailed description of the attributes.

TABLE I
ATTRIBUTE DESCRIPTION

Attribute Type	Attribute Name	Description
General	Name	Name of the student
	Gender	The gender of the student
	Age	Age of students extracted from date of birth
	Number of students per class	Total students in the class
	Attendance	Total number of presents of student throughout the year
Academic	Urdu	Score in Urdu
	English	Score in English
	Mathematics	Score in Mathematics
	Science	Score in Science
	Islamiat	Score in Islamiat
	Obtained Marks	Sum of all subject marks scored by student
	Total Marks	Sum of all subjects' total marks.
	Percentage	Percentage evaluated based on obtained marks
	Status	Status of the students' performance
	Number of subjects Failed	Total number of subjects the student failed in.
	Class In charge	The Homeroom teacher
	Class	Grade in which student is studying

Student status is categorized in three types. Passed, promoted and failed. If the student has obtained above 33% marks in every subject, he\she is considered passed. If the student gets below 33% in one subject, he/she is promoted to the next grade with grace marks. If the students scored below 33% in two subjects, he/she also gets promoted to the next class. The students who get below 33% in 3 or more subjects are considered failed.

The data was extracted manually from data records of students from different primary, middle and secondary schools. The data set was limited to the past 4 years, from 2015 to 2019. Total data of 1439 students were collected from various schools.

B. Data set

Attributes spanned on two different sources; class register and PEC register; were carefully merged into a single datasheet. For this Microsoft Excel was used to record the data digitally. As a result, an integrated data set was prepared consisting 6 general and 11 academic a total of 17 related attributes.

C. Data Preprocessing:

Low-quality data will lead to low-quality mining results [9]. Most datasets are highly noisy, missing and inconsistent due to their typically huge size and likely to be collected from different multiple and heterogeneous sources. Preprocessing of data is mandatory before applying the classification to avoid inapt results. Data is quality data if they satisfy the requirements of the intended use. The factor that compromises the quality of the data includes inaccuracy (having incorrect attribute values), incompleteness missing point value), inconsistency (containing divergences to categorizing the data). There are two steps of preprocessing performed to improve the quality of the dataset.

D. Data Imputation:

The data should also be examined regarding consecutive rules. A consecutive rule says that there can be no missing values between lowest and highest values for the attribute [9]. Reason for the missing may include 1) a party originally asked to provide a value for the attribute refuses to give the information. 2) the party does not know the correct value. 3) or the value is to be provided in the later process. Whatever the reasons behind missing values are, these missing values contribute to data inaccuracy. In data preprocessing, data imputation is the process of replacing missing values with the substituted values. It is important to identify, mark and handle missing data is crucial for better predicament of the data. Missing values can be replaced with some other values. This is called imputing missing values. To impute missing values, the Replace Missing Values filter was applied. Replace Missing Values is the mean imputation which is the replacement of a missing observation with the mean/median of the non-missing observations of that variable. The collected dataset had 11% missing values. A Replace Missing Value filter is applied to recover the error.

E. Data Transformation:

Data transformation is the process of transferring data from one format to another, typically from source data format into destination data format. It includes tasks like data integration which is the process of collecting different data types (different databases and datasets) and merging the data into the same structure or same schemas. In this preprocessing step data was collected from different schools of the district, putting data into their relative attributes and then merging the different sources of the data into a single dataset. After that, convert the source format of the dataset into the destination data file in this case ARFF format.

F. Classification Techniques:

Based on the finding of the literature review, the most frequently used data mining approaches were classification. Additionally, widely used data mining algorithms by category were Decision Tree, Naïve Bayes, Artificial Neural Networks and Support Vector Machine. Classification is done using two ways i.e., applying classifiers directly on the dataset and applying classifiers after reducing attributes.

F.1. Data set Data mining implementation on full dataset:

In this phase all the dataset with 17 attributes was used to predict students' performance. The 10-fold cross validation process was utilized. Training was done using nine folds and for testing the remaining one-fold was used. For Decision Tree J48 and Support Vector Machine SMO, 10-cross validation criteria are applied. And for Multilayer Perceptron MLP the training data and testing data was split into 90:10 ratios respectively for better results. Following are the steps performed in phase 1. Details are given in Algorithm 1.

Algorithm 1: Data Mining with Dataset D

- **Input:** Dataset **D**= $\{d_1, …., d_n\}$, Classifier set **C**= $\{c_1, …, c_n\}$
- **Output:** Accuracy Set X = $\{x_1, …, x_n\}$, Best accuracy **A**
- ***Step 1****:* Preprocessing
  - Transform source data to **.arff** format
  - Apply ***Replace Missing Value*** function
- ***Step 2:*** Classifier Implementation
  - Accuracy set **X**
  - ***for*** i= 1 to 4, ***do***
    - Apply ci on **D**
    - Get xi,
- ***Step 3:*** Accuracy comparison
  - Best accuracy **A**
  - If $x_1 > x_2$¬ Then A= x_1 else A= x_3
  - If A > x_3 Then A=A else A = x_3
  - If A> x_4 Then return A else A=x_4
- return **A**

Initially the dataset contained 11% missing values which were resolved in step one of the algorithms by using Replace Missing Value algorithm. This algorithm replaces all the missing values for nominal and numeric attributes in the data with the modes and means from the training data. In the next step: classification was applied on the whole dataset. J48 is a decision tree which uses top-down recursive, divide and conquer strategy. It uses a measure called information gain to choose the attribute at each stage. Root node of the tree is no. of subjects failed which leads to internal nodes math, science, no of subjects failed, no of students per class, obtained marks. All these internal nodes result in leaf nodes which is the status of the students being passed, promoted or failed.
SMO learns a linear model on the given data set. It attempts to find a dividing plane where examples of one class fall on one side and examples of other classes fall on the other side of the plane. SMO breaks the problem into possible subproblems in this case promoted, failed and passed which are solved analytically. MLP is a class of feedforward artificial neural networks. It usually consists of three layers: input layer, hidden layer and output layer. In this approach MLP was first trained on 10% of the dataset and then tested with 90% of the dataset. Graphical representation of the Network was difficult to visualize due its complex and massive size.

F.2. Data Mining Implementation after Attribute Reduction:

In this phase attributes reduction is applied to the dataset to see if the model achieves better accuracy. For this purpose, CFS Correlation feature Selection and Info Gain Attribute Eval were used.

Attribute Reduction:
Attribute reduction is a technique which is used for data reduction in the data mining process. The method of data reduction may achieve a condensed description of the original data which is much smaller in quantity but keeps the quality of the original data. Classification accuracy is improved by removing irrelevant and redundant features from the dataset. In this study, for the attribute reduction process, CFS Subset Eval feature selection algorithm with Best first method and Info Gain Attribute Eval with Ranker method were used. Feature selection or attribute selection is a process by which one automatically searches for the best subset of attributes in one dataset.
CFS Subset Eval, evaluates the significance of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Subsets of features that are highly correlated with the class while having low intercorrelation are preferred. Key benefits of performing feature selection are: reduced overfitting, which means less redundant data which leads to less opportunity to make decisions based on noise. Other benefits include improved accuracy and reduced training time. The CFS is computed using Eq. 1.

$$F = \frac{s\overline{f_p}}{\sqrt{s + (s-1)\overline{f_q}}} \qquad \text{Eq. 1}$$

Where Fs is correlation between the summed features subset and class variables, **'s'** is the number of subset features, $\mathbf{f_p}$ is the average of the correlation between the subset feature and the class variable, $\mathbf{f_q}$ is the average inter correlation between subset features.
Info Gain Attribute Eval evaluates the worth of an attribute by measuring the information gain with respect to the class. It is used for feature selection tasks. It measures how each feature contributes in decreasing the overall entropy. Entropy is the measure of degree of impurity. The closest it is to 0, the less impurity is in the dataset. Hence, a good attribute is an attribute that reduces the entropy at most and also contains more information. To calculate the Info Gain Eq 2 is used.

$$IG\ (Cx, Ax) = H(Cx) - H\ (Cx \mid Ax) \qquad \text{Eq. 2}$$

where C is the Class and A is the attribute.

The Entropy H(C) is defined in Eq 3,

$$H(Cx) = sum\ (Pi * log2\ (Pi)) \qquad \text{Eq. 3}$$

P_i being the probability of the class in the data set.

Algorithm of classification with attribute reduction is given below

Algorithm 2: **Attribute Reduction**

- ***Input***: Dataset **D** = {d_1, …, d_n},
- ***Output***: $\mathbf{S_1}$ = {a_1, …, a_n}, $\mathbf{S_2}$ = {b_1, …, b_n}, Selected features **F** = {f_1, …, f_n}
- ***Step 1:*** Apply **CFS** with **BF** on **D**
  - Get feature set $\mathbf{S_1}$
- ***Step 2:*** Apply InfoGain with Ranker on **D**
  - Get feature set $\mathbf{S_2}$
- **Step 3:** Extract common features
  $F \leftarrow S_1 \cap S_2$
- return ***F***

The CFS Subset Eval algorithm searches for a subset of features that work well together. Proposed algorithm used CFS Subset Eval with Best first method. InfoGainAttributeEval searches for the attribute with more information. Each method gave a set of selected attributes. By merging together, a subset of 6 attributes was formed from 17 attributes. The reduced attributes which are selected by feature selection attribute are given in table 2:

TABLE 2
BEST ATTRIBUTES

Attribute	Description
Name	Name of the student
Attendance	Number of presents of students throughout the year
Math	Score in Mathematics
Science	Score in Science
Status	Pass, Promoted or Failed
No of subjects failed	Total number of subjects' student failed

After attribute reduction, a new dataset was formed using only above six attributes. This is explained in algorithm 3 to get more efficient and accurate results from classifiers.

***Algorithm 3*:** **Data Mining with Dataset R**

- ***Input***: Dataset **D** = {d_1, …, d_n}, Selected features **F** = {f_1, …, f_n}, Classifier set **C**= {c_1, …, c_n}
- ***Output***: Reduced Data set **R** = {r_1, …, r_n}, Accuracy set **Y** = {y_1, …, y_n], Best accuracy **B**
- ***Step 1:*** Create reduced dataset **R**
  - R $\leftarrow$ extract **R** from **D** based on **F**
- ***Step 2***: Preprocess
  - Transform reduced data ***R*** to ***.arff*** format
  - Apply ***Replace Missing Value*** function on ***R***
- ***Step 3***: Classifier Implementation
  - Accuracy set Y
  - ***for*** i= 1 to 4, ***do***
    - Apply $\mathbf{c_i}$ on **D**
    - Get $\mathbf{y_i}$.
- ***Step 4***: Accuracy comparison
  - Best accuracy **B**
  - If y_1> y_2 Then B= y_1 else B= y_2
  - If B > y_3 Then B=B else B = y_3
  - If B> y_4 Then return B else B=y_4
- return **B**

In this algorithm, a reduced dataset based on the selected feature set. The reduced dataset contains the data which includes only attributes gained from the feature set. The reduced dataset was put through the classification algorithms to examine accuracy change. Again, the classifiers are executed on reduced dataset using 10-cross validation.
This section proposed a classification methodology which was applied to the selected data set. Next section will provide the analysis of the results obtained from the proposed solution and will compared to find the efficiency of the proposed methodology

IV. SIMULATION AND RESULTS

This section provides the results and analysis which will help us understand how the students are learning and how the classifiers performed in terms of their evaluation measures. To evaluate and compare classifier performance we used accuracy, precision and time taken. Performance of classifiers executed on a full dataset is given in table 3.

TABLE 3
RESULTS WITHOUT ATTRIBUTE REDUCTION

Classifier	Total Instances	Correctly Classified Instances	Incorrectly Classified Instances	Precision	Recall
SMO	1438	1415	23	0.984	0.984
J48	1438	1424	14	0.991	0.990
MLP	1438	1242	196	0.891	0.88
NB	1439	1341	97	0.949	0.933

Highest number of correctly classified instances are 1424 out of 1438 which is achieved by J48 decision tree leaving only 14 instances which are incorrectly classified. SMO closely follows J48 by predicting 1415 correctly classified instances. NB gave 1341 correctly classified instances out of 1439. MLP comes at the end with the highest incorrectly classified instances of 196. Confusion matrix is given in table 4.

TABLE 4
CONFUSION MATRIX PHASE 1

	SMO			J48			MLP			NB		
Classifier as →	a	b	c	a	b	c	a	b	c	a	b	C
a= Promoted	171	2	12	183	1	1	32	2	137	175	3	7
b= Failed	3	59	1	1	62	0	1	50	3	5	58	0
c= Pass	4	1	1185	10	1	1179	1	1	1072	79	3	1108

In Confusion Matrix, **a** is used for promoted, **b** is for Failed and **c** is for Pass. In SMO, promoted classified as promoted are 171, promoted classified as failed are 2 and promoted which are classified as pass are 12. Failed which are classified as promoted are 3. Failed classified as failed are 59. Failed classified as pass is 1. Pass classified as promoted are 4. Pass classified as failed is 1 and pass classified as pass are 1185.

In J48, promoted classified as promoted are 183, promoted classified as failed is 1 and promoted classified as pass are 1. Failed classified as promoted is 1. Failed classified as failed are 62. Failed classified as pass is 0. Pass classified as promoted are 10. Pass classified as failed is 1 and pass classified as pass is 1179. In MLP, promoted classified as promoted are 32, promoted classified as failed are 2 and promoted classified as pass are 137. Failed classified as promoted are 1. Failed classified as failed are 50. Failed classified as pass are 3. Pass classified as promoted is 1. Pass classified as failed is 1 and pass classified as pass are 1072. In NB, promoted classified as promoted are 175, promoted classified as failed are 3 and promoted classified as pass are 7. Failed classified as promoted are 5. Failed classified as failed are 58. Failed classified as pass is 0. Pass classified as promoted are 79. Pass classified as failed are 3 and pass classified as pass are 1108.

J48 clearly out performs other classifiers closely followed by SMO and NB classifiers. MLP performs the worst. Accuracy results of classifiers without attribute reduction are given in fig 2.

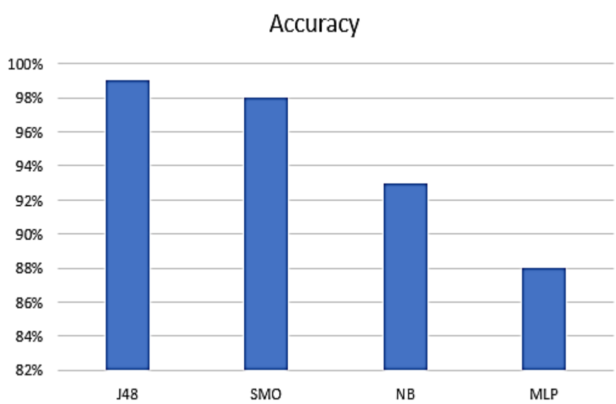


Fig. 2. Accuracy and classifier comparison without attribute reduction

It can be seen from Fig 2, that decision tree classifier J48 provides the best accuracy of 99% in the shortest amount of time. Where Artificial Neural network MLP gives the worst output with taking most time as compared to other executed algorithms.

In phase 2 the data was reduced to selected attributes and classifiers were applied to the reduced dataset. Table 5 shows the process of data mining with reduced attributes. After reduction the correctly classified instances improved in all three classifiers. Confusion matrix shows the changes after the reduction. instances are more efficiently classified as promoted, failed and passed. J48 predicted four more instances than before attribute reduction, making a total of 1428 out 1438 correctly classified instances.

SMO and NB prediction are also improved by more correctly classified instances whereas in spite of improved correctly classified instances in MLP the precision of the classifier has worsened as compared to other classifiers.

TABLE 5
RESULTS WITH ATTRIBUTE REDUCTION

Classifier	Total Instances	Correctly Classified Instances	Incorrectly Classified Instances	Precision	Recall
SMO	**1438**	**1417**	**22**	**0.985**	**0.985**
J48	**1438**	**1428**	**10**	**0.993**	**0.993**
MLP	**1438**	**1274**	**164**	**0.827**	**0.873**
NB	**1438**	**1392**	**46**	**0.969**	**0.968**

The confusion matrix after attribute reduction shown is in table 6.

TABLE 6
CONFUSION MATRIX PHASE 1

	SMO			**J48**			**MLP**			**NB**		
Classifier as →	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**B**	**c**
a= Promoted	170	1	14	183	1	1	1	2	159	170	3	12
b= Failed	0	62	1	1	62	0	1	50	2	4	59	0
c= Pass	4	1	1185	6	1	1183	0	1	1074	25	2	1168

Significant changes can be seen in the confusion matrix after attribute reduction. SMO improved failed classified as failed by 3 instances. J48 improved pass classified as pass by 4 instances. NB improves pass classified as pass by 57 instances whereas MLP classified promoted as pass by significant number of instances.

Accuracy results of classifiers without attribute reduction are given in fig 3.

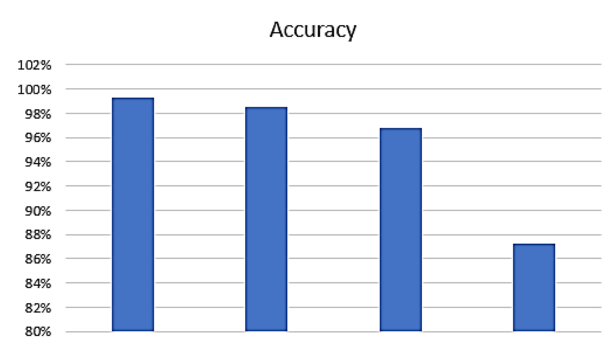


Fig. 3. Accuracy and classifier comparison after attribute reduction

From the above results it can be observed that accuracy of decision tree j48 is highest closely followed by SMO classifier. SMO took more time to build the model as compared to J48. MLP takes maximum time as compared to other classifiers.

Result Comparison Before and After Reduction of Attributes:

Table 7 shows results of each classifier before and after the reduction of attributes.

TABLE 7
RESULTS COMPARISONS

Classifier	Without Attribute Reduction			After Attribute Reduction		
	Accuracy	Precision	Time Taken	Accuracy	Precision	Time Taken
SMO	98.4%	0.984	10 sec	98.5%	0.985	10 sec
J48	99%	0.991	3 sec	99.3%	0.993	3 sec
NB	93%	0.949	3 sec	96.9%	0.969	3 sec
MLP	88.7%	0.891	40 sec	87.2%	0.827	40 sec

In the above results, it can be observed that the time taken to build the model in phase 1 and phase 2 is constant. All classifiers took the same amount of time to build the model even after the reduced data set. Accuracy of SMO is slightly improved after the reduction of attributes with a minute change of 0.1%. J48 also performed better with a reduced set of attributes and was a bit higher than SMO. NB gave significant results after attribute reduction improving the accuracy by 3%. Most interesting results were provided by MLP, it not only took the maximum time to build the model but also its accuracy is reduced with the reduced attributes whereas other classifiers' accuracy improved.

The main focus of this study was on the approaches introduced in [3], [8] and [4].All the approaches collected different student attributes based on general, academic, demographic or other influencing attributes on which mainly three classifiers were applied i.e. NB,SVM, C4.5 and Neural Networks, However, [4] also introduced feature reduction after classification to improve the efficiency of the model. This study was inspired by attribute reduction in [4] also along with novel introduced approaches mentioned in [3] and [8]. A brief comparison of state of art techniques and proposed technology is given in table 8.

TABLE 8
COMPARISONS WITH PREVIOUS TECHNIQUES

Technique	Classifier Accuracy			
	SVM	Decision Tree	Neural Network	Naïve Bayes
Rustia, Cruz [8]	61.89%	73.10%	65.67%	62.98%
Zafar, Mueen [4]	-	80.5%	81.4%	85.7%
Francis, Babu [3]	64.15%	62.26%	48.42%	-
Proposed	98.5%	99.3%	87.2%	96.8%

It can be clearly observed that the proposed technology provided much better accuracy as compared to previous techniques. Decision tree provided the most accurate result close up to 99% which was not achieved in the previous techniques. Similarly, SVM, NB and Neural Network also performed better. The most influencing factors found by proposed techniques are much similar to the [4] influencing factors. Loss of participation in online forums which leads to the failure in the course in [4] corresponds to the low attendance of students and lack of understanding in subjects i.e. science and math which needs students to be physically and mentally present, (factors) as observed in proposed technique. Hence it is observed that J48 gave best results in terms of time and accuracy and MLP gave worst results.

As guided from the literature review, we extracted three major and most commonly used techniques that are said to be the most effective classifiers i.e, Decision tree J48, Support vector machine SMO and Artificial neural network MLP. However, our study shows that MLP does not perform well with the tabular data, as it takes too much time to execute the model which is understandable due to its deep learning approach. But when the results are compared MLP shows considerably low accuracy as compared to decision tree and support vector machine. Hence it can be deduced that deep learning models may not be the best be appropriate for tabular data, since it might contain simple enough relationships which a decision tree or support vector machine could lead to better predictions. As for the other two mining techniques it can be seen that decision tree J48 outperforms the others. In previous studies, the most awarded data mining approach was C4.5/C5.0/J48 decision tree with providing average accuracy up to 90% whereas our approach provides most accurate results of J48 capable of 99% as compared to previous methods. Data reduction technique also gave promising results by not only improving the evaluation measures but also giving significant insight of the influencing factors that affects the students' performance

V. CONCLUSIONS

The main objective of this study was to predict PEC (Punjab Examination Commission) Students' performance. The examination is held annually by the Punjab Examination Commission of the primary and middle level students. In order to perform the EDM techniques we needed to integrate the data from various schools of District Rawalpindi, Punjab. The aim of this study was to predict these students' performance and find the factors which affected it the most. Three data mining techniques applied were, Decision tree, Support Vector Machine and Artificial Neural Networks. J48, SMO and MLP were selected from the above categories. All these algorithms were applied to the collected dataset. Decision Tree J48 performed best in aspect to other classifiers. This study can help the teachers on the basic level of the students. It also assists teachers to identify students' who are most likely to fail the examination. Various studies have been conducted to identify the factors which influence students' performance. These factors can differ from one institute to another or in this case one education system to another. This study shows that students' lack of interest in school or other family factors leads to low attendance in school. Also, students' low learning capability of students in logical subjects like Mathematics and Science can lead to their loss of interest in these subjects and further to failure in examination. In this case the teachers or the administration staff should make the environment of the school healthy and attractive for students. Extracurricular activities can be introduced to improve students' morals and their interest in school. Also, the teachers should be more interactive with students and use aids to help them develop interest in these challenging subjects. Education system should also provide the school and staff with more facilities and should encourage the teachers so they accomplish this exigent task with high significance.

References

[1] H. Kumar, B. Saini, and P. Singh, "Academic Enhancement System using EDM Approach," Second Int. Conf. Adv. Comput. Commun. Paradig., 2019.

[2] A. Abu Saa, "Educational Data Mining & Students' Performance Prediction," no. (IJACSA) International Journal of Advanced Computer Science and Applications, 2016.

[3] B. K. Francis and S. Babu, "Predicting Academic Performance of Students Using a Hybrid Data Mining Approach," no. Journal of Medical Systems, 2019.

[4] A. Mueen, B. Zafar, and U. Manzoor, "Modeling and Predicting Students' Academic Performance Using Data Mining Techniques," no. I.J. Modern Education and Computer Science, p. 2016.

[5] Y. K. Shaheed, T. O. Oladele, A. O. Akanni, and W. M. Ibraham, "Students Performance prediction Based on Data Mining Classification Techniques," no. Nigerian Journal of Technology (NIJOTECH), 2018.

[6] U. O. Cagas, A. J. P. Delima, and T. L. Toledo, "PreFIC: Predictability of Faculty Instructional Performance through Hybrid Prediction Model," Int. J. Innov. Technol. Explor. Eng., vol. 8, no. 7, 2019.

[7] A. M Baby, "Pedagogue Performance Assessment (PPA) using Data mining Techniques," IOP Publ., 2018.

[8] R. A. Rustia, Ma. M. A. Cruz, Michael Angelo P., M. A. P. Burac, and T. D. Palaoag, "Predicting Student's Board Examination Performance using Classification Algorithms," p. 5, 2018.

[9] J. Han, M. Kamber, and J. Pei, Data Mining Concepts and Techniques, 3rd ed. .

[10] H. Abdullah, "Using Data Mining Techniques to Predict StudentbPerformance to Support Decision Making in University Admission Systems," IEEE, 2020.

[11] A. Abu Saa, M. Al-Emran, and K. Shaalan, "Mining Student Information System Records to Predict Students' Academic Performance," Springer Nat. Switz.

[12] K. Agarwal, E. Maheshwari, C. Roy, M. Pandey, and S. Rautray, "Analyzing Student Performance in Engineering Placement Using Data Mining," no. Proceedings of International Conference on Computational Intelligence and Data Engineering, 2019.

[13] L. Aulck, N. Velagapudi, J. Blumenstock, and J. West, "Predicting Student Dropout in Higher Education," ICML, 2016.

[14] J. Cunha, E. Moura, and C. Analide, "Data Mining in Academic Databases to Detect Behaviors of Students Related to School Dropout and Disapproval," Springer Int. Publ. Switz., 2016.

[15] E. Fernandes, M. Holanda, M. Victorino, V. Borges, R. Carvalho, and G. Erven, "Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil," J. Bus. Res., 2019.

[16] Dr. D. A. A. Gnana Singh and Dr. E. J. Leavline, "Dimensionality Reduction for Classification and Clustering," IJ Intell. Syst. Appl., 2019.

[17] G. Gutiérrez, J. Canul-Reich, A. Zezzatti, L. Margain, and J. Ponce, "Mining: Students Comments about Teacher Performance Assessment using Machine Learning Algorithms," Int. J. Comb. Optim. Probl. Inform., vol. 9, 2018.

[18] R. Hasan and S. Mahmood, "Student Academic Performance Prediction by using Decision Tree Algorithm."

[19] R. Hasan, S. Palaniappan, Abdul Rafiez, S. Mahmood, and K. Sarker, "Student Academic Performance Prediction by using Decision Tree Algorithm," Int. Conf. Comput. Inf. Sci., no. 4rth, 2018.

[20] R. Hemaid and A. El-Halees, "Improving Teacher Performance using Data Mining," Int. J. Adv. Res. Comput. Commun. Eng., vol. 4, no. 2, 2015.

[21] A. Ikhwan et al., "A Novelity of Data Mining for Promoting Education Based on FP-Growth Algorithm," no. International Journal of Civil Engineering and Technology (IJCIET), 2018.

[22] R. R. Kabra and R. S. Bichkar, "Performance Prediction of Engineering Students using Decision Trees," Int. J. Comput. Appl., vol. 36.

[23] L. Khobragade and P. Mahadik, "Students' Academic Failure Prediction Using Data Mining," no. International Journal of Advanced Research in Computer and Communication Engineering, 2016.

[24] A. Kumar Pal and S. Pal, "Evaluation of Teacher's Performance: A Data Mining Approach," Int. J. Comput. Sci. Mob. Comput., 2013.

[25] A. M.O, O. A.O, and W. W.F, "Teachers' Performance Evaluation in Higher Educational Institution using Data Mining Technique," Int. J. Appl. Inf. Syst., 2016.

[26] S. MARDIKYAN and B. BADUR, "Analyzing Teaching Performance of Instructors Using Data Mining Techniques," Inform. Educ., 2011.

[27] T. Mishra, Dr. D. Kumar, and Dr. S. Gupta, "Mining Students' Data for Performance Prediction," Fourth Int. Conf. Adv. Comput. Commun. Technol., 2014.

[28] A. Patil and S. R. Nagarmunoli, "Instructor"s Performance Evaluation System using Data Mining Techniques," Int. J. Adv. Res. Comput. Commun. Eng., vol. 6, no. 9, 2017.

[29] V. Paul, M. Bresfelean, and N. Ghisoiu, "Determining Students' Academic Failure Profile Founded on Data Mining Methods," Int Conf Inf. Technol. Interfaces, 2008.

[30] A. Pradeep, S. Das, and J. Kizhekkethottam, "Students

Dropout Factor Prediction Using EDM Techniques," Int. Conf. Soft-Comput. Netw. Secur., p. 2015.

[31] P. Shayan and M. van Zaanen, "Predicting Student Performance from Their Behavior in Learning Management Systems," no. International Journal of Information and Education Technology, 2019.

[32] R. Villwock, A. Appio, and A. Andreta, "Educational Data Mining with Focus on Dropout Rates," Int. J. Comput. Sci. Netw. Secur., 2015.

[33] M. Zaffar, M. A. Hashmani, and K. S. SAVITA, "Performance Analysis of Feature Selection Algorithm for Educational Data Mining," IEEE Conf. Big Data Anal. ICBDA, 2017.