A MULTI-LEVEL CLASSIFICATION MODEL PERTAINING TO THE STUDENT'S ACADEMIC PERFORMANCE PREDICTION

Mrinal Pandey[1] and S.Taruna[2]
Research Scholar[1] and Associate Professor[2]
Department of Computer Science, Banasthali University, Jaipur (Rajasthan), India

ABSTRACT
The students' performance monitoring and evaluation is an essential activity of an education system to keep track of the success and failure records of the students. The objective of this research is to provide the best classification model to predict the students' academic performance. In this paper we propose a Multilevel Classification Model (MLCM) based on Decision Tree Algorithm for the predictions of the academic performance of the undergraduate engineering students. The multi-level classification model consists of two levels. In level one, the four classification models namely Decision Tree (J48), Lazy Learner (IBK), Neural Network (MLP) and Naïve Bayes Tree (NBT) were constructed, evaluated and compared. The decision tree classifier was selected for the model construction in this step. In level 2, the overall accuracy of the classification model as well as the accuracy of individual class was enhanced by eliminating the outliers from the original dataset and by constructing Multilevel Classification Model (MLCM) using filtered dataset.

KEYWORDS: *Classification, Prediction, Decision Tree, Lazy Learner, Neural Network, Cross Validation, K-Nearest Neighbors, Multi-Layer Perceptron.*

I. INTRODUCTION

Educational data mining is becoming popular in the modern educational era to improve the quality of the education system. A typical education system consists of various activities like planning, admission, enrollment, examination, placement, student monitoring and evaluation etc. Today students' retention and placement is becoming a great area of concern for any higher educational organization and with it, the necessity of student performance analysis is required to predict the academic performance of the student so that the right feedback can be given to students related to their progress in the course and instructors can evaluate their teaching learning process [1]. Authors of [2] cite "The problem arises when instructors with little or no knowledge about data analysis want to predict the students' performance or want to prevent student drop out among others". In response to this, Data mining plays a vital role in decision making in the area of higher education. This can be achieved by extracting meaningful, previously unknown information from the educational data set and providing it to the decision makers. A student performance system can be useful for everyone in the education system to predict the performance of the students. In this research we provide the multilevel classification model for the student performance system therefore four classification techniques namely Decision Tree, Lazy Learner, Neural Network and Naïve Bayes Tree were used for constructing the classifier.

The organization structure of this paper is as follows. Section II discusses the data mining classification techniques applied in this study. Section III describes the related work. Section IV represents the propose work for this study. Section V shows the results and analysis. Finally conclusions and the future scope of this research are discussed in section VI and VII respectively.

II. DATA MINING TECHNIQUES

Data mining is the convergence of multiple disciplines such as Database Technology, Machine Learning, Information Science, Statistics, Visualizations and other disciplines. In academics, data mining involves predicting the students' performance, predicting placement probability, clustering of similar students and associating types of students with suitable courses. Machine learning techniques include classification and regression algorithms, association rules, sequential pattern analysis as well as clustering and web mining etc.

2.1 Classification Techniques

Classification techniques are supervised learning techniques which classifies the data items into the predefined class labels. The data classification process involves the model construction and the model usage. In model construction the training data are used to construct the model by classification algorithm and during model usage this model is used to predict the value of unseen data. Data mining encompasses various classification techniques like Decision Tree Algorithms, Bayesian Classification, Classification by Back Propagation, Support Vector Machine and K Nearest Neighbor etc.

2.1.1 Decision Tree C4.5

A decision tree algorithm (ID3) is given by Quinlan Ross [3]. It is a flowchart like tree structure where each internal node in a tree represents test on an attribute and each branch represents an outcome of the test while each leaf node holds a class value [4]. The C4.5 is an advanced version of ID3 algorithm. It overcomes the limitations of ID3 algorithm. C4.5 uses the gain ratio as an attribute selection measure to build a decision tree. An attribute having highest gain ratio value is used as a root node and the possible values of that attribute comes out as the branches of the tree and then again next higher gain ratio valued attribute is used as a next level of the tree. This process continues till the complete tree is generated. C4.5 also perform pruning to remove unnecessary branches in the decision tree to improve the accuracy of classification. In WEKA C4.5 is known as J48 Decision Tree (J48 is the Java version of C4.5).

2.1.2 Naïve Bayes Tree

NB Tree is a hybrid classification technique between Decision Tree and Naive Bayes classification. The algorithm is similar to the classical recursive partitioning schemes except that the leaf nodes created are Naive-Bayes categorizers instead of nodes predicting a single class [5]. It combines the advantage of both Decision Tree and Naïve Bayes Classification.

2.1.3 Multilayer Perceptron

The MLP is a multi layer feed forward Neural Network. It consists of an input layer, an output layer and one or more hidden layers. It uses back propagation algorithm for training purpose. The objective of this algorithm is to obtain a set of weights that makes almost all the tuples in the training data classified correctly [4]. It compares the output of the network with the desired output and the error is computed. Once the error is computed then this error is back propagated to the network to adjust the weights to reduce the error rate in each iteration.

2.1.4 K-Nearest Neighbor Algorithm

KNN is non parametric lazy learner algorithm for the classification and prediction. In order to classify a new instance, this algorithm checks the distance of its k neighbors from the training set to classify it. In general Euclidean Distance measure is used to find the distance. A training instance closest to the given test instance predicts the same class as this training instance [4]. In WEKA this algorithm is available as IBK.

III. RELATED WORK

Over the past few years a large number of research were performed in the arena of educational data mining. Authors of [6] conducted a survey of applications in the area of data mining for the educational data mining since 1995 until 2005. Educational data mining is being used as a prediction tool and helps in various decision making situations in the education sector. An example of this is given in [7] to predict the accuracy whether students will become graduate or not. The information obtained can be used to help the low performing students. Another student's performance analysis was performed in [8] using classification and clustering data mining techniques particularly with ZEROR and DBSCAN algorithm respectively. In research [9] Decision Tree C 4.5, KNN, Naïve Bayes as well as Boosting and Bagging techniques were used to predict the accuracy of the classifiers and Bagging was found the best algorithm among all the five classifiers. A decision tree based classification model was proposed in [10] to predict the students' academic performance in the first year of engineering examination based on the past performance of the students. In an another study, authors of [11] proposed a classification model using ID3 Decision Tree to predict the division or grades of the students on the basis of previous database. They used attendance, Class tests, seminars and assignment marks as attributes to predict the performance of the students at the end of the semester. A comparative study of classifiers was conducted with [2] and the authors proposed a Meta Algorithm for pre-processing the data set to enhance the accuracy of the classifiers. The author [12] developed a prototype of decision support system based on regression techniques for predicting the student's future grades. A research was conducted in [13] to predict the students eligible for scholarship and to predict low performing students to assist them. This study was conducted for two educational institutes AIT and CTU in Vietnam. A comparative study of decision tree algorithms was conducted for predicting the student's grades for undergraduate engineering students in [14]. The authors compared four decision tree algorithms namely J48, Reptree, Simple Cart and NB Tree. The J48 algorithm was found the best suitable algorithm for the study. In [15] CRISP methodology was employed and three classifiers namely ID3, C4.5 and Naïve Bayes were compared to predict the performance of students particularly for C++ course. In [16] study, the authors proposed a simulation tool based on Fuzzy application for analyzing students' performance on the basis of their CPA and GPA. The authors inclined their study towards the development of Intelligent Planning System (INPLANS) using Fuzzy Systems, Neural Networks, and Genetic Algorithms for the Academic Advisory Domain in educational institutions by evaluating and predicting the students' performance. A software tool was developed based on neural network to predict the student performance in the course of mathematics. This tool helps educators to identify weak students [17]. A research was conducted in [18] using ensemble methods. The authors proposed Adaboost ensemble with the Genetic Algorithm to predict the performance of the students in early stages so that risk of failure can be controlled by providing appropriate advising to the students those are at high risk. In [19] authors proposed a generalized Student Success System (S3) that provides an ensemble-based analytical system for tracking student academic success. This system consists of a flexible predictive modelling engine that uses machine learning techniques to identify student who are at risk. The author of [20] compared the Bayesian Network classifiers for predicting the student's academic performance and developed a model which helps to identify the drop out students and students who require special attention and counselling from their teachers. A comparative study was presented in [21] to find the prediction ability of neural network particularly for student performance prediction. The authors trained the neural network based on two heuristic algorithms namely, the cuckoo search and gravitational search algorithms and conclude that the neural network trained by the cuckoo search is better for the prediction of students' academic performance in this particular study. In [22] study association a rule mining was used to select relevant attributes on the data set of 60 students and a Multi-Layer Perceptron Neural Network was employed to compare and evaluate the results after applying the attribute selection method to evaluate the students' performance.

IV. PROPOSED METHOD

Figure 1 shows the class distribution. During the student performance prediction analysis, it was observed that this is a class imbalance problem. The number of students who obtains grade 'A' and grade 'C' is relatively smaller than the total number of students while the number of students who obtains grade 'B' are very large in the dataset and the students who obtain grade 'F' are on average in number, therefore it is essential to develop a model that can maximize the classification accuracy as well as the individual class accuracy so that future instances can be predicted correctly. With this objective we performed two levels of experimentation on our dataset and proposed a Multi Level Classification Model (MLCM). Section 4.1 describes the working of proposed model and Figure 2 shows the working of MLCM.

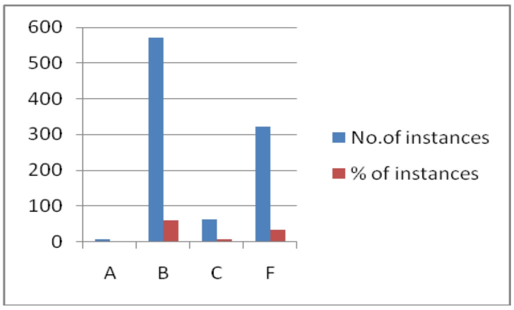

Figure 1: Class Distribution

4.1 Working of Multi level Classification Model

Initially the bootstrap method was applied to the dataset to maintain the class distribution. Bootstrap method is a resample function which oversamples the minor classes and under samples the major classes. Thereafter four classifiers namely J48 Decision Tree, IBK, MLP and NB Tree were constructed using 10 fold cross validation and compared to determine the best model for the student performance prediction.

In the next level with the objective of improving the accuracy of this model we performed the data analysis on the predicted results obtained from DT J48. We observed that there exist some instances which can be considered as outliers in statistical terms (i.e. Student obtains grade 'A' or grade 'B' and fails in results or student obtaining lower grades 'C 'or 'F' and achieves a higher grade such as 'A' or 'B'). These misclassified instances were identified and removed randomly from the dataset. To eliminate such instance from the dataset we selected the highest gain ratio value attribute 'agg-g-7th'. The predicted values of misclassified instances were compared with the values of attribute 'agg-g-7th'. If a student obtains higher grade value for attribute 'agg-g-7th' (i.e. 'A' or ' B' grade) and fails in prediction results or student obtains lower grades 'C' or 'F' and achieves higher grade in predictions such as 'A' or 'B' then we remove such instances from the original data set. In this study near about 4% misclassified instances were randomly removed from the data set. In this level the new filtered dataset with the bootstrap method was used to construct the DT J48 (Multilevel Classification Model) model, which is better than the model constructed in the level 1 in terms of model performance as well as in individual class performance.

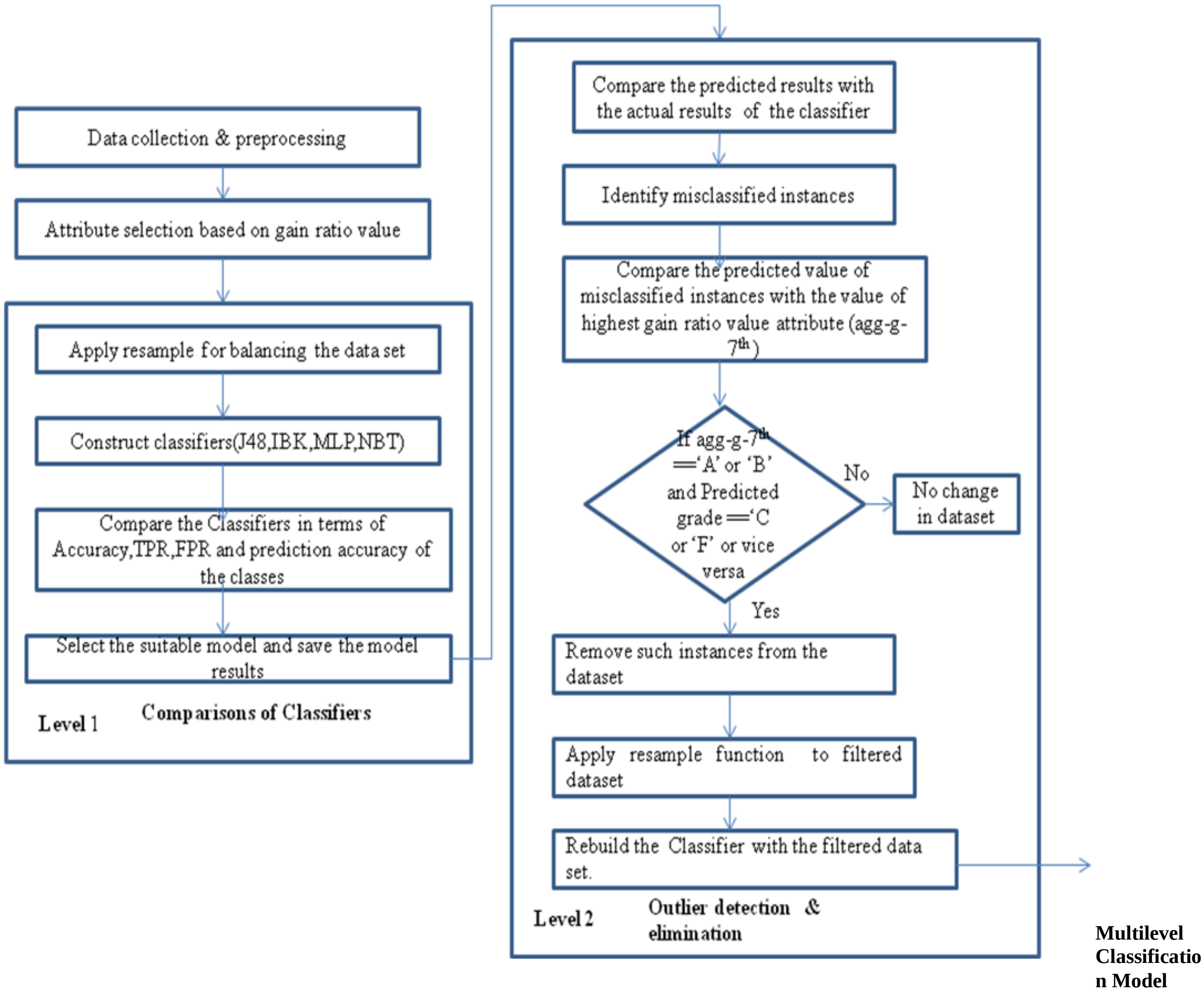

Figure 2: Working of Proposed Model

V. EXPERIMENTS AND RESULTS

This section starts with the experimental setup consisting of the data set, parameters of algorithms and attributes description used for the study and proceeds with the classification algorithms comparison and Outlier identification and removal.

5.1 Experimental Setup

The WEKA toolkit is selected for this study. WEKA is an open source tool [23] available which supports a wide range of data mining tasks and machine learning algorithms. It is useful for large datasets. This research is based on classification techniques. The classification algorithms available in WEKA can be directly applied to the dataset.

5.1.1 Data set

The data for the classification problem is collected from an engineering college in an EXCEL sheet. The data set consists of the academic performance and demographic information of the students. This is a multiclass problem where the grades of students are divided into four classes namely A, B, C and F according to their academic performance starting from high school to the 7th semester (prefinal year) of the engineering graduation. The original dataset consists of around 1000 instances and 18 attributes. The student‘s performance data was the continuous data. It was converted into nominal (categorical) data and the initial data cleaning as well as transformation was performed. Finally total 11 numbers of attributes were selected on the basis of the gain ratio attribute selection measure for this study. Remaining 7 attributes were removed from the data set. Table 1 from [24] shows attribute along with their gain ratio values.

Table 1: Attribute description

S. No	Name	Value#	Description	Possible Values	Gain Ratio
1	ag-g-7th	13	Aggregate grades up to 7th semesters	A,B,C,F	0.52861
2	ag-G-3rd	11	Aggregate grades up to 6th semesters	A,B,C,F	0.32879
3	BACKLOGS	15	Backlogs	Yes, No	0.2624
4	BACKLOGS-No.	14	Total no of Backlogs (till 7th semester)	0,1-5,6-10,>10	0.21569
5	Iyear-Grade3	8	Aggregate grades of 1st and 2nd semesters	A,B,C,F	0.21244
6	2nd-year-Grade	9	Aggregate grades of 3rd and 4th semesters	A,B,C,F	0.21101
7	Gap	16	Gap in study	0,1,2	0.18789
8	Grade-12th	6	Student 's Grades in Class Standard 12th	A,B,C	0.10596
9	10th-Grade	4	Student 's Grades in Class Standard 10th	A,B,C	0.09801
10	7TH-SEM-GRADE	12	Grades of 7th semesters	A,B,C,F	0.05374
11	Gender	1	Student Gender	Male, Female	0.04924
12	Board of 12th	7	Name of Senior secondary board	CBSE,ICSE,HBSE	0.04262
13	Age	3	Age of student	22,23,24,25,26	0.04106
14	region	17	Region from where a student belongs	NCR,FARIDABAD,OUTER-ZONE	0.02928
15	Board of 10th	5	Name of High school board	CBSE,ICSE,HCSE	0.02022
16	3rd-year-Grade	10	Aggregate grade of 5th and 6th semesters	A,B,C,F	0.01596
17	Branch	2	Student Branch	CSE,IT,MECH,ECE	0.00573
18	Final/Class	17	Predicted Value	A,B,C,F	

Note: Grade values: A=81-100, B=61-80, C=41-60, F<=40

5.1.2 Attribute Selection Measure

Attribute selection measure [4] is used to select the best attribute for splitting purpose. The most popular attributes selection measures are gini index, information gain and gain ratio for Decision Tree algorithms. Attribute selection measure determines the ranking of each and every attribute in a dataset and selects the attribute which has the highest ranking (score). The selected attribute is labelled as the root node and branches grow for each possible value of the attribute. The tuples are partitioned according to the splitting criterion. This is a recursive process and continues until the entire tree is constructed. The C4.5 Decision Tree Algorithm (J48) use the gain ratio method. The attribute with the maximum gain ratio is used to select as the splitting attribute. The gain ratio is based on Entropy and information gain given in (1) and (2) respectively from [4]. Gain ratio and split information is given in (3) and (4) respectively.

$$\mathrm{Info}(D) = -\sum_{i=1}^{m} p_i \log_2(p_i) \tag{1}$$

$$\mathrm{Gain}(A) = \mathrm{Info}(D) - \mathrm{Info}_A(D) \tag{2}$$

$$\mathrm{GainRaito}(A) = \mathrm{Gain}(A)/\mathrm{SplitInfo}(A) \tag{3}$$

$$\mathrm{SplitInfo}_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times \log_2 \frac{|D_j|}{|D|} \tag{4}$$

Where $\frac{|D_j|}{|D|}$ is the weight of j^{th} partition.

5.1.3 Parameters for algorithms

The J48 Decision tree algorithm is available in the class"weka. classifiers. trees. J48". It has four main parameters named minNumObj, confidence Factor, unpruned, and BinarySplits. The Naïve Bayes

Tree is available in the class “weka. classifiers. trees. NBTree”. It is a Decision Tree with Naïve Bayes classifier. There is a single parameter called Debug in this algorithm. The values of it can be either true/false. True value leads to generate additional information to the console. K-Nearest Neighbor Algorithm (IBK) is available in the class “weka. classifiers. lazy. Ibk”. It has four main parameters named KNN (number of neighbors), Cross Validate, Distance Weighting and Nearest Neighbor Search Algorithm. The multilayer perceptron neural network is available in the “weak. classifiers. functions. Multilayer Perceptron”. It has fifteen parameters e.g. GUI, autoBuild, debug, delay, hidden Layers, learning Rate, momentum, training time etc.
The four aforementioned algorithms were tested for different sets of parameter settings and different numbers of folds to find the effects on the results. Finally the result obtained by default parameter setting and 10 fold cross validation were the best and selected for the experiment.

5.2 Classification Algorithm Comparison

This section describes the construction and evaluation of level 1 classification model. Initially the bootstrap method was applied to the original dataset and after that four aforementioned classifiers were constructed, evaluated and compared for the construction of level 1 classifier.
Figure 3 shows the complete J48 decision tree construction. The attribute ag-g-7th having the highest gain ratio value (0.52861) therefore it was selected as the root node and the four possible values for this attribute A, B, C, and F are coming out as the four branches from this root node. The attribute ag-G-3rd is the next highest gain ratio valued attributes so it selected as a next node for splitting purpose. This process continues till the generation of complete tree.

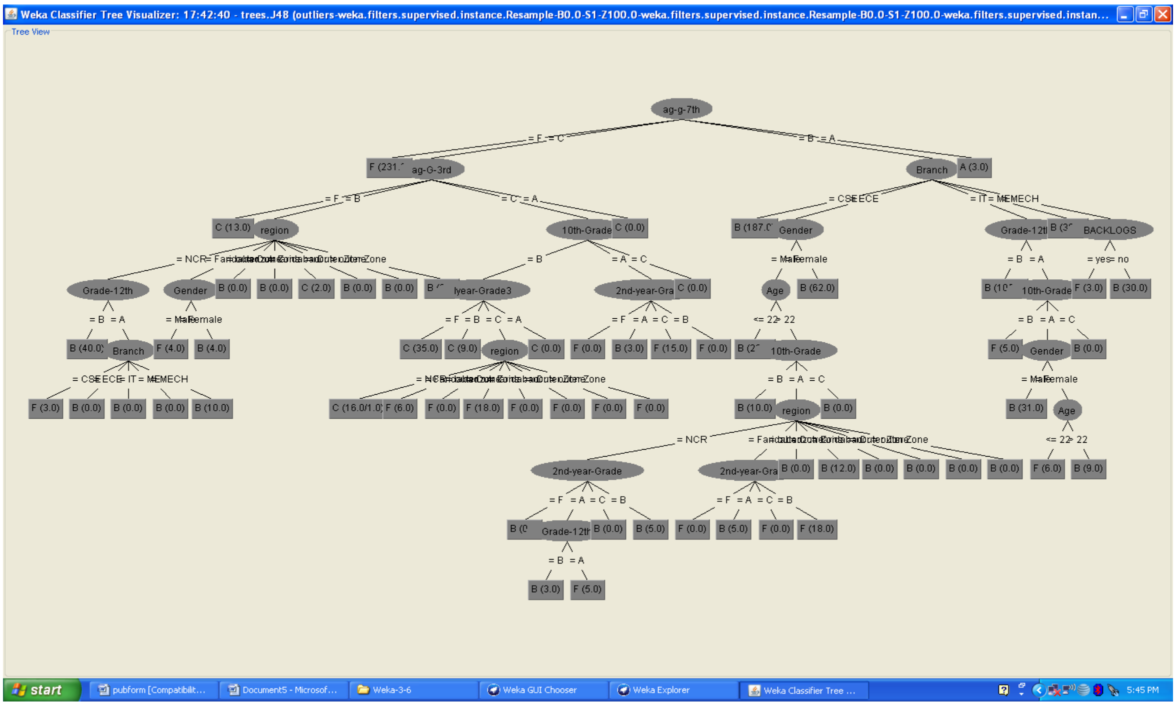

Figure 3: Decision Tree Construction

Table 2 depicts the performance comparisons of classifiers in terms of time taken to construct the classifiers in seconds, traditional model accuracy in percentage, the number of correctly classified instances and number of incorrectly classified instances. Although the accuracy rate is most commonly used empirical measure for model evaluation but it is not sufficient enough when working with imbalance datasets. It does not differentiate among the numbers of correctly and incorrectly classified examples of the respective classes and may lead to erroneous conclusions [25] therefore in this study Confusion Matrix, True Positive Rate (TPR) and False Positive Rate (FPR) measures are used for model evaluations.

Table 2: Performance Comparisons of Classifiers

Classifiers	Original Dataset			
	Time	Efficiency	Correctly classified Instances	Incorrectly Classified Instances
DT	0.03	95.49%	953	45
MLP	21.9	94.88%	947	51
IBK	0.02	93.59%	934	64
NBT	2.11	90.08%	899	99

A typical confusion matrix for binary class is shown in Table 3 from [4]. The column represents the predicted class and the row represents the actual class. In the confusion matrix TP is the number of positive instances classified correctly and TN is the number of negative instances classified correctly where FP is the number of negative instances incorrectly classified as positive and FN is the number of positive instances incorrectly classified as negative.

Table 3: Confusion Matrix

	Predicted Class		
Actual Class	Class	C1	C2
	C1	TP	FN
	C2	FP	TN

True Positive Rate is the number of correctly classified instances in a given class. It is also known as sensitivity or hit ratio and can be defined as

$$TPR=TP/TP+FN \qquad (5)$$

False Positive Rate is the number of incorrectly classified instances of a given class. It is also known as false alarm and can be defined as

$$FPR=FP/FP+TN \qquad (6)$$

Accuracy of the classifier is the total number of correctly classified instances. It can be defined as

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN} * 100 \qquad (7)$$

Table 4 shows the confusion matrix for the classifiers constructed in this part of the experiment. It states that MLP is excellent for predicting class A and class C (with 100% predicted accuracy) and DT is good for all four A, B, C and F classes (100%, 97.26%, 87.14%, 96.34%) While IBK stands between these two algorithms where as NBT is achieving lowest class accuracy among the all four classifiers.

Table 4: Confusion Matrix for Classifiers

Classifiers	DT				IBK				MLP				NBT			
Class	B	A	C	F	B	A	C	F	B	A	C	F	B	A	C	F
B	**569**	0	3	12	**537**	2	7	8	**618**	0	0	9	**530**	4	11	9
A	5	**7**	0	0	4	**8**	0	0	12	**7**	0	0	6	**6**	0	0
C	3	0	**61**	8	9	0	**74**	9	2	0	**40**	8	10	0	**72**	10
F	8	0	6	**316**	14	0	11	**315**	20	0	0	**282**	38	0	11	**291**
Predicted Accuracy %	**97.26**	**100**	**87.14**	**96.34**	**95.21**	**80**	**80.43**	**94.87**	**94.79**	**100**	**100**	**94.31**	**97.26**	**100**	**87.14**	**96.34**

Table 5 shows the true positive rate and false positive rates for all the four classes. It shows that TP for class A is the highest (. 667) in IBK algorithm while MLP achieves the highest TP rate for class B (. 986) and the TP rate for class C is almost same for DT, IBK and MLP classifiers where as DT is the best in terms of TP rate for class F (.958). FP rate can also be compared for aforementioned classifiers. The MLP has a lowest FP rate for both the classes A and C (i. e 0) where as DT is good for all classes i.e. FP rate for classes A, B, C and F are (0, .039, .01 and.03) respectively. IBK lies between the two classifiers in terms of FP rate.

Table 5: TP rate and FP rate of Classifiers

	DT		IBK		MLP		NBT	
Class	TP Rate	FP Rate	TP Rate	FP Rate	TP Rate	FP Rate	TP Rate	FP Rate
A	0.583	0	0.667	0.002	0.368	0	0.5	0.004
B	0.974	0.039	0.969	0.061	0.986	0.092	0.957	0.122
C	0.847	0.01	0.804	0.02	0.8	0	0.783	0.024
F	0.958	0.03	0.926	0.026	0.934	0.024	0.856	0.029

The performance of NBT is relatively poor in terms of TP rate and FP rate. Here we can see that the three classifiers namely DT, IBK and MLP are good enough for modelling. It is also observed from the table 3 to 5 that the Decision Tree J48 has the highest accuracy rate as well as good individual class accuracy. Moreover, it has higher TPR and lower FPR. It gives the best results for all 4 classes as well as for overall model performance. Therefore we saved the prediction results using DT J48 model.

5.2.1 Outliers detection and removal

This section discusses the experiment and results of the level 2 experiment. Level 2 starts with the identification of outliers by comparing the predicted results with the actual results and proceeds by comparing the values of attribute 'agg-g7th with the values of predicted results particularly for the wrongly classified instances. If the wrongly (misclassified) instance has a higher grade value i.e. 'A' or 'B' and predicted as lower value i.e. 'C' or 'F' or vice versa then such instance were removed from the dataset and this filtered dataset was used for model construction. Initially DT J48 model is constructed with this new dataset. It was observed that there is a significant improvement in the accuracy of 2nd level of the classifier as well as the accuracy of individual classes for DT J48 (MLCM). To validate and compare the results the remaining three classifiers IBK, MLP and NBT were constructed and evaluated for the filtered dataset. The results show that the accuracies of the classifiers and class accuracies for these algorithms were improved during this part of the experiment.

To evaluate the efficiency of the classifiers, 10 fold cross validation and standard performance techniques such as confusion matrix, true positive rate, false positive rate, recall, precision, classifier efficiency, kappa statistic and RMS etc. were used.

Table 6 shows the performance summary of the classifiers. It can be observed from this table that both MLP and IBK classifiers are equal in all the parameters except the time taken to build the model. Time taken by MLP is too high as compare to IBK. It is also clear from the table that DT is best in terms of model accuracy as well as in prediction of individual class accuracy. The table also depicts the kappa statistic values. The DT J48 is leading with the highest value (i.e. 0.9961). A comparative study of the classifiers is presented in Table 7 for both the datasets (original dataset and filtered dataset). It shows that DT J48 MLCM is the best classifier in this particular study.

Table 6: Summary of the Classifiers

Experimental Statistics									
Classifies	Time in Seconds	Efficiency (%)	Correctly Classified Instances	Incorrectly Classified Instances	Kappa Statistic	Mean Absolute Error	Root mean Squared Error	Relative Absolute Error(%)	Root relative Squared Error (%)
DT	0.02	99.79	959	2	0.9961	0.0016	0.0329	0.60	19.12
IBK	0	99.69	958	3	0.9942	0.002	0.036	0.73	20.14
MLP	24.59	99.69	958	3	0.9942	0.004	0.386	1.48	23.16
NBT	1.95	99.27	954	7	0.9864	0.0108	0.0651	4.03	17.80

Table 7: Comparisons of classifier performance with Original data set and Filtered dataset

Classifiers	Original Dataset				Filtered Dataset			
	Time (Sec)	Efficiency	Correctly Classified Instances	Incorrectly Classified Instances	Time (Sec)	Efficiency	Correctly Classified Instances	Incorrectly Classified Instances
DT	0.03	95.49	953	45	0.02	99.79%	959	2
MLP	21.9	94.88	947	51	24.59	99.68%	958	3
IBK	0.02	93.59	934	64	0	99.69%	958	3
NBT	2.11	90.08	899	99	1.95	99.27%	954	7

Table 8 depicts the predicted class accuracy. It can be observed from the Table 8 itself that predicted accuracy for all the classes is improved with the filtered dataset. The accuracy of class 'C' has a remarkable improvement. It is clear from the table that all four classifiers produce excellent results for each class prediction.

Table 8: Predicted Class Accuracy for Original dataset and Filtered dataset

	Efficiency of classes in % for Original dataset				Efficiency of classes in %for Filtered dataset			
Classifiers	A	B	C	F	A	B	C	F
DT	100	97.26	87.14	96.34	100	99.82	98.67	100
IBK	80	95.21	80.43	94.87	99.69	99.82	98.67	100
MLP	100	94.79	100	94.31	99.69	99.82	98.67	100
NBT	100	94.79	87.14	96.34	99.05	99.47	98.67	100

Table 9 represents the summary of the performance measures of classifiers in terms of TP rate, FP rate, Precision, Recall and ROC Area for all aforementioned classifiers and for all four classes. This table also shows that DT J48 is the best performing classifier among the all four classifiers and NB tree is the lowest performer in this study.

Table 9: Classifier performance measures

Classifiers	Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
DT	A	1	0	1	1	1	1
	B	0.998	0.003	0.998	0.998	0.998	0.998
	C	1	0.001	0.987	1	0.993	0.999
	F	0.997	0	1	0.997	0.998	0.998
IKB	A	0.667	0	1	0.667	0.8	1
	B	0.996	0.003	0.998	0.996	0.997	0.999
	C	1	0.001	0.987	1	0.993	1
	F	1	0.002	0.997	1	0.998	1
MLP	A	0.333	0	1	0.333	0.5	0.579
	B	0.998	0.003	0.998	0.998	0.998	0.997
	C	1	0.001	0.987	1	0.993	1
	F	1	0.002	0.997	1	0.998	1
NBT	A	0.333	0	1	0.333	0.5	0.997
	B	0.993	0.008	0.995	0.993	0.994	0.999
	C	1	0.001	0.987	1	0.993	1
	F	0.997	0.005	0.991	0.997	0.994	0.999

Figure 4 depicts the TPR comparison of all four algorithms for both stages of the classifiers. The figure shows that the TP rate is highest for all the classes in DT classifiers. It is noticeable that TP rate for NBT slightly goes down for class “A”. Similarly Figure 5 is a graphical representation of FP rate comparisons for all three classifiers as well as for all four classes. The FP rate is lowest for all A, B, C and F classes using Decision Tree Algorithm (MLCM).

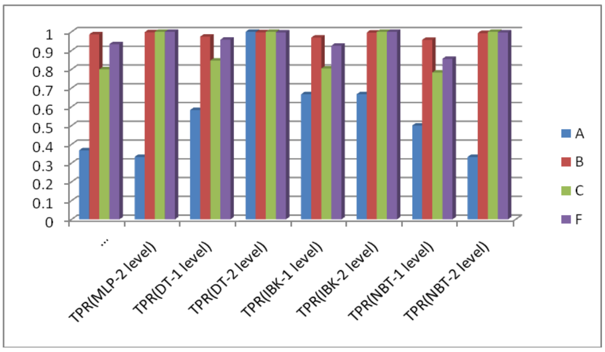

Figure 4: True Positive Rate Comparisons

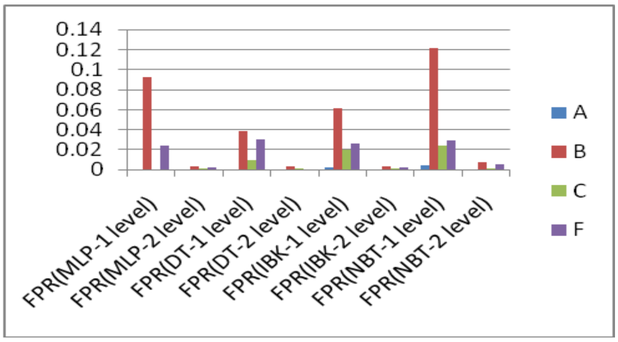

Figure 5: False Positive Rate Comparisons

Figure 6 depicts the comparisons of all the four classifiers for both the levels of classifiers with the individual class accuracy. It is observed and analysed that the method proposed in this research produce the significant results. It is also noticeable that the accuracy of individual classes is increased by all four algorithms, but the percentage of improvement in class accuracy is highest in NBT and NBT is reaching towards the other three classifiers. Thus we can conclude that by removing the outliers the accuracy of predicted class can be enhanced.

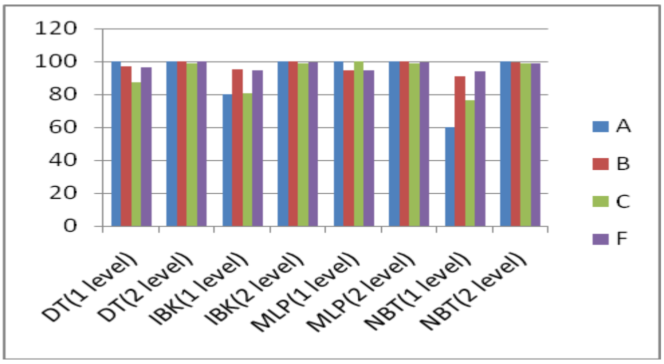

Figure 6: Class accuracy comparisons for classifiers

VI. CONCLUSIONS

Four classification algorithms were studied in this experiment and a multilevel classification model is proposed with the combination of the pre-processing techniques such as resample filter as well as removing the misclassified instances from the initial classifiers. This approach shows that DT(MLCM) classifier achieved highest accuracy of 99.79% and the highest individual class accuracy for classes A,B,C and F are 100%, 99.82%, 98.67% and 100% respectively i.e the DT classifier is 100% accurate for predicting higher score and failure students. The TPR is highest and the FPR is lowest in DT for all four classes among the four classifiers. Therefore DT J48 (MLCM) is the best suitable classifier for predicting the grades of the students. This can be used as a base for developing the student performance system.

VII. FUTURE WORK

This research can be further enhanced by applying the proposed method on different levels of academic performance dataset of the students and a robust student performance decision support system can be developed. This will help educators to find the students falling among the four categories without having any prior knowledge of data analysis techniques. The proposed approach in this research can be used for different dataset to establish the results.

REFERENCES

[1]. Angel A. Juan, Thanasis Daradoumis, Javier Faulin, and Fatos Xhafa, (2009). “A data analysis model based on control charts to monitor online learning processes”. Int. J. Bus. Intell. Data Min., 4:159{174, July 2009. ISSN 1743-8195.

[2]. Garica and Zorrilla Marta ,(2011) “Comparing classication methods for predicting distance students' performance”. Published in Workshop and Conference Proceedings 17.

[3]. Quinlan, J.R. 1993. C4.5: Programs for machine learning, Morgan Kaufmann, San Francisco.

[4]. Han, J., Kamber M. and Pie,(2006) , Data Mining Concepts and Techniques, 2nd edition, Morgan Kaufmann Publishers 2006.

[5]. Kohavi, R. 1996. Scaling Up the Accuracy of Naïve-Bayes Classifiers: A Decision-tree Hybrid. Proceedings of the 2th International Conference on Knowledge Discovery and Data Mining, pp.202-207, 1996.

[6]. Romero, C., Ventura, S. 2007. “Educational Data Mining: A Survey from 1995 to 2005”. Expert Systems with Applications 33, 2007, pp.135-14Applications and Reviews, IEEE Transactionson, 40(6):601{618, 2010. ISSN 1094-6977. doi: 10.1109/TSMCC.2010.2053532

[7]. Luan. 2004. Data Mining application in higher education . SPSS Inc, 2004.

[8]. Aher, Sunita B., LOBO L.M.R.J,(2011). “Data mining in educational system using WEKA”. International conference on emerging technology trend . Proceeding published by International Journal of computer applications .

[9]. Jantan , H. and Othman, AR Hamdan,(2010). “Classification and prediction of academic talent using data mining techniques”. KES 2010, Part 1,LNAI 62/6,pp 491-500.

[10]. R.R Kabra, R.S Bichkar, "Performance Prediction of engineering students using Decision trees", International Journal of computer applications (0975-8887), Vol-36-No.11, December 2011.

[11]. Brijesh Kumar Baradwaj and Saurabh Pal, "Mining Educational Data to Analyze Student Performance", International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6,2011.

[12]. S.B Kotsiantis,"Use of machine learning techniques for educational purposes: a decision support system for forecasting students' grades", Springer Science,2011.

[13]. Nguyen Thai Nghe, Paul Janecek, and Peter Haddawy', "A Comparative Analysis of Techniques for Predicting Academic Performance", 37th ASEE/IEEE Frontiers in Education Conference, IEEE, 2007.

[14]. M. Pandey and V.K.Sharma," A Decision Tree Algorithm Pertaining to the Student Performance Analysis and Prediction", International Journal of Computer Applications (0975 – 8887) Volume 61– No.13, January 2013.

[15]. Qasem A. Al-Radaideh, Emad M. Al-Shawakfa, and Mustafa I. Al-Najjar " Mining Student Data Using Decision Trees", The 2006 International Arab Conference on Information Technology.

[16]. Saxena, Urvashi R. , Singh, S.P. 2012. "Integration of Neuro-Fuzzy Systems to Develop Intelligent Planning Systems for Predicting Students' Performance". International Journal of Evaluation and Research in Education (IJERE), Vol.1, No.2, December 2012, pp. 61~66 ISSN: 2252-8822

[17]. Livieris Ioannis E. , Drakopoulou Konstantina, Panagiotis Pintelas. 2012. " Predicting students' performance using artificial neural networks", www:hcicte2012.uth.gr/main/sites/default/files/proc/.../LivierisEtAl.pdf

[18]. Sharaf, Ahmed., Malaka, ElDen., Moustafa, A., Harb, Hany M., Emara, Abdel H.(2013). "Adaboost Ensemble with Simple Genetic Algorithm for Student Prediction Model", International Journal of Computer Science & Information Technology (IJCSIT) Vol 5, No 2, April 2013.

[19]. Essa, Alfred. and Ayad, Hanan. (2012) "Improving student success using predictive models and data Visualizations", Research in Learning Technology Supplement: ALT-C 2012 Conference Proceedings, ISBN 978-91-977071-4-5 (print), 978-91-977071-5-2(online), http://dx.doi.org/10.3402/rlt.v20i0.19191

[20]. Sundar (2013) "A Comparative Study for Predicting Student's Academic Performance using Bayesian Network Classifiers", IOSR Journal of Engineering (IOSRJEN) e-ISSN: 2250-3021, p-ISSN: 2278-8719 Vol. 3, Issue 2 (Feb. 2013), ||V1|| PP 37-42.

[21]. Eng-Fung Chen and Quang Hung Do (2014) 'Training Neural Networks to Predict Student Academic Performance: A comparison of Cuckoo search and Gravitational search Algorithms", Int. J. Comp. Intel. Appl. 13, 1450005 (2014) [18 pages] DOI: 10.1142/S1469026814500059

[22]. Borkar and Rajeswari (2014),' Attributes Selection for Predicting Students' Academic Performance using Education Data Mining and Artificial Neural Network', International Journal of Computer Applications (0975 – 8887) Volume 86 – No 10, January 2014

[23]. Weka, University of Waikato, New Zealand,http://www.cs.waikato.ac.nz/ml/weka/

[24]. Pandey M. and Taruna ,S. 2014. "An Empirical Analysis of Classification Techniques for Predicting Academic Performance". published in proceding of 4th IEEE International Advance Computing Conference February 21st -22nd , 2014.

[25]. Galar, Mikel., Fern´andez, Alberto., Barrenechea, Edurne., Bustince Humberto, and Herrera, Francisco .2011."A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches". IEEE Transcations on systems, Man and Cybernetics—Part –c:Applications and Reviews, Digital Object Identifier 10.1109/TSMCC.2011.21612

AUTHORS

Mrinal Pandey received her M Tech degree in Computer Science from Banasthali University. She is an Assistant Professor in Manav Rachna College of Engineering, Faridabad. She is pursuing PhD from Banasthali University in the area of Data Mining**.**

S Taruna is an Associate Professor in Computer Science Department in Banasthali University. She is PhD Supervisior of three research scholar. Her research Areas are Communication Network, Sensor Network and Data Mining. She has presented many papers in National and International Conferences, and published around 25 papers in various journals. She is also reviewer of various journals and conferences.