International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 1 (2015) pp. 1113-1142

http://www.ripublication.com

Mining Students' Record to Predict Their Performance in Undergraduate Degree

L. Ramanathan, Dr. Angelina Geetha, Dr. M. Khalid

Assistant Professor (Senior), SCSE, VIT University, Vellore,
Professor,Department of Computer Science and Engineering,
B.S Abdur Rahman University, Chennai,
Dean,School of Computing Science & Engineering,Galgotias University.
E-mail:{lramanathan@vit.ac.in,angelina@bsauniv.ac.in, mkhalidbs@yahoo.com}

Abstract

Data mining is the process of discovering useful patterns and relationships from the data. It facilitates making decision in education field. The quantity of data educational database stores is growing rapidly. Higher educational institute can use this database which holds hidden information. This hidden information is helpful in improving student's performance in their academics. Successful prediction of student's performance leads to assure and achieve better quality in higher education system. The performance of student is determined by different attributes. In this paper, the classical techniques and the evolutionary data mining technique called grammar based genetic algorithm (GGA) is used to build a model for predicting the performance of undergraduate students by using their past educational and general record. The result of different classifiers is compared to discover the final prediction model. Further, the result of prediction is analyzed and its relationship with different predictors is studied. This kind of analysis facilitates improvement in a specific area of educational system and also helps studying the various general factors responsible for the performance of student.

Keywords: Data Mining, Educational Data Mining, Students Performance, Classification, Prediction Model

1. Introduction

The aim of data mining is extracting the knowledge out of huge set of data. The knowledge that is mined should be useful and advantageous. In today's world of competition, institutes are looking for the student who can pass the course

successfully. Institute put efforts to maintain the admission system of student in a good manner. The management should take a quick and proper decision, thus time to time student's accurate information is required. In an educational institute it is very crucial having the ability for predicting student's performance. Student's performance in academics is the main criteria set by the company for their recruitment. So apart from institute, this performance also concerns to corporations. Academic performance of student depends on different factors which may be related to personal life, social life, psychology and variables related to environment.

Data mining proposes various techniques and tools for analysis of different kinds of data available in the database. Using these data mining techniques and algorithms in the educational field is known as Educational Data Mining abbreviated as EDM [1]. Higher quality education plays a major role developing a good attitude and knowledge in students. There are students who do not perform good in their academics and these results in their low grades and failure in graduation; therefore they take more time for completing their degree. Thus it is necessary to study the factors that are responsible for predicting the good and bad performance of student.

For prediction of students' performance, the popular technique of data mining used is classification. Classification is the technique used for prediction of unknown values of data by using the known values of data. The aim of classification is developing the model which can predict the required result from available dataset. Further, predefined class groups are mapped to data by employing classification technique. It is also called as supervised learning as determination of classes is done before data is examined. Different methods of classification e.g. decision tree, rule mining and so on can be employed on students' records to predict their performance in academics and final examination. This kind of prediction facilitates teachers for identifying students who scored low marks, who is weak in a particular area and reasons behind the failure. So this helps in taking the steps for improving students' performance and makes them passes with good marks.

The educational institutes approach towards achieving higher performance and producing better graduates by providing quality teaching. Education plays a major and important role in progress and development of the country. Many social factors and living habits of students are also important in the prediction of their performance in educational institute. In higher education institute, it is not possible to accomplish quality results without detail and sufficient knowledge. Data mining facilitate filling such gap in knowledge, and helps in prediction of students' performance. In this paper, the different models are built using classical data mining classifiers and evolutionary grammar based genetic programming to predict students' performance in first year of under graduation in the institute. Moreover, result of different data mining techniques are analyzed to discover the final prediction model; and the relationship between input attributes and output attribute is studied to identify the contribution of each input factor in predicting the grades/marks. Further steps can be taken by institute, management, teachers and parents to improve the performance of student by improving the various factors.

2. Literature Review

Studying various factors (predictor attributes) which are responsible for student's performance in their academics is crucial in analyzing and making better educational system. This approach is examined by different studies that are carried out in this field, which include collecting data from educational institute or university and from different surveys. The impact of important factors on students' performance is determined by various data mining algorithms. Researches are done on predicting student's future performance on test using automatic detector for learning in future [5], [Ryan S.J.D. Baker, Sujith M. Gowda and Albert T. Corbett, 2011], prediction of final marks of students depending on their involvement in web based forums [6], [M.I. López, J.M Luna, C. Romero, S. Ventura]. Apart from this, researchers determined the performance of learning of each student using adaptive learning in e-learning platform [7], [Ya-huei Wang, Hung-Chang Liao, 2011] and some focused on impact of multiple graphical representations on student's learning [8], [Martina A. Rau, Zachary A. Pardos, 2012], Some studies carried by researchers proposed the method to improve the accuracy of knowledge tracing (KT) model using student's first response time to the question [9], [Yutao Wang, Neil T. Heffernan, 2012], whereas some proposed method to predict perceived disorientation in e-learning using student's eye movement and log metrics [10], [Akcapinar, G., Cosgun, E., &Altun, A., 2011].

Moreover, the research is done on predicting student's performance using his/her own past information instead of using other's performance for individual's performance prediction [11], [Nguyen Thai-Nghe, Tomas Horvath, Lars Schmidt-Thieme, 2011]. Also forecasting placement result of secondary education and discovering important factors/predictors with the use of sensitivity analysis [2], [B. Sen, E. Ucar, D. Delen, 2012], thus focusing on students' academic achievement. Some discovered high relationship in between family income and test score, which clears that high income result in positive impact on score in test and ultimately on educational achievement. Besides that researchers mentioned the point that achievement in academics is not affected directly by income, instead attitude and notions of family get affected which is responsible for result outcome. In addition, other suggest parameter tuning approach using meta learning, that is increasing the classifier accuracy by tuning default parameters for various datasets of educational system [3]. For example, [M.M. Molina, J.M. Luna, C.Romero, S.Ventura, 2012] have done study on finding different parameter and their values which improves accuracy rather than using default parameter values. Their work focused on how the accuracy of classification algorithm increases and decreases when default value of parameter confidence factor (confidenceFactor) and minimum number of objects (minNumObj) is changed. Further researchers (C. Marquez-vera, C. Romero and S. Ventura, 2012) using data mining for prediction of school failure [4], which aims at increasing accuracy for student failure prediction. For this study, they gathered students' personal, family as well as marks information as input variables (attributes) and set output variable as PASS or FAIL. Feature selection is performed to identify the most important attributes contributing to output. Again they applied an algorithm for data balancing and cost sensitive classification for achieving better accuracy.

Educational Data Mining (EDM) deals with building methods which explains different nature of data and different students who learns in different settings. Classification algorithms that have a form like IF-THEN rules called as rules of prediction which makes the classification and develop essential relationship between data. Grammar based technique can be used to obtain better accuracy by generating optimal rules of classification. This approach of grammar based algorithm leads in improved prediction model which performs better prediction of students' data class variable as compared with other classification algorithms.

3. Methodology

The method or architecture used in this paper to predict students' performance dwells to Knowledge Discovery and Data Mining (KDD) process, as shown in Figure 1.

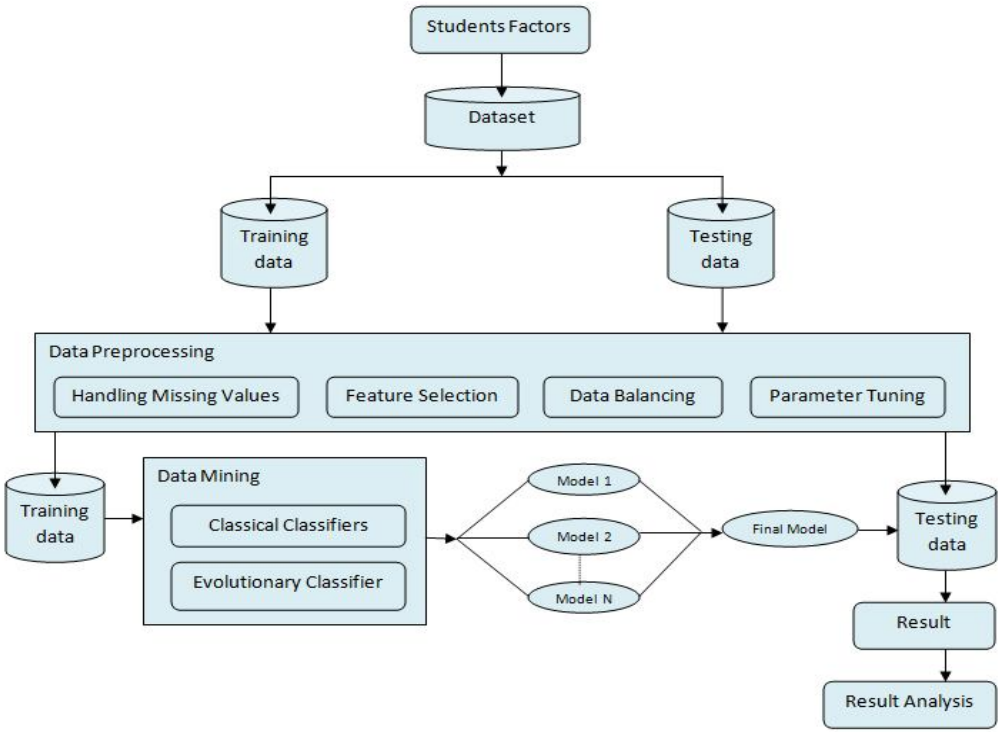


Figure. 1:. Architecture of the scenario

The architectural approach and its principal stages are as follows:

1. Data Collecting: In this stage, all necessary information (attributes/factors) of students is collected. For this, the factors which may affect to the performance of student are considered. Then a dataset is created from this collected information.
2. Data Preprocessing: This composed of preparing the dataset for employing data mining approach. This stage includes data integration, data cleaning, and data transformation. Further, different techniques are used like handling missing values, feature selection and parameter tuning for imbalanced data.

3. Data Mining: In this stage, data mining algorithms such as classification is used in prediction of students' performance. For doing this, a grammar based genetic programming is used and its comparison is carried out with other algorithms of classification. Moreover, the problem of data imbalance is overcome.
4. Result and Analysis: Finally, the model is built on the basis of classification performance and this model is used for prediction of students' grade in first year of under graduation. The factors i.e. predictors that are responsible for grade prediction are analyzed and their relationship with class variable (output attribute) is interpreted.

4. Data Source Collection

The source of data collection used is from VIT University, Vellore, India. The dataset is formed by collecting the information of B.Tech (Bachelor of Technology) students of second, third and final year. Initially record of 200 students is taken from various branches of B.Tech. The total 25 attributes i.e. features or variables; and their description and values that are used in this study are shown in Table 1. The output attribute or class variable that is to be predicted is grade obtained in B.Tech 1st year (Btech1stYearGrade).

Table 1: Attribute Names, Description and Values

Students' Attribute Name	Description	Values
Gender	Sex	{Male, Female}
AgeDuringBtechAddmission	Age during B.Tech 1st year admission	{sixteen, seventeen, eighteen, nineteen}
10thGrade	Grade in 10th class (SSC)	{S 90% - 100%, A 80% - 89%, B 70% - 79%, C 60% - 69%, D 50% - 59%, E 40% - 49%, F < 40%}
12thGrade	Grade in 12th class (HSSC)	{S 90% - 100%, A 80% - 89%, B 70% - 79%, C 60% - 69%, D 50% - 59%, E 40% - 49%, F < 40%}
10thBoard	Board in 10th class (SSC)	{CBSE, ICSE, SB}
12thBoard	Board in 12th class (HSSC)	{CBSE, ICSE, SB}
12thPCMorPCBGrade	Grade in PCM/PCB in 12th class (HSSC)	{Distinction 75% - 100%, First-class 60% -74%, Second-class < 60%}
VITEEERank	Rank in VITEEE entrance exam of college	{Very-good 1 - 1000, Good 1001 - 2000, Average 2001 - 3000, Below-average 3001 - 4000, Greater-than-4000, NA}
YearGapBetween12thAnd1stYear	Number of year gap between 12th and B.Tech 1st year	{nil, zero, one, two}

NumberOfArrears	Number of arrears/backlogs faced in B.Tech	{zero, one-two-three, four-five-six, seven-eight-nine}
TimeManagement	Time management (speed) in exam	{Good, Average, Poor}
FamilyAnnualIncome	Family annual income status	{High, Middle-class, Poor, Below-poor}
FamilySize	Number of members in family	{one, two, three, greater-than-three}
FatherQualification	Father's qualification (education)	{no education, elementary, secondary, graduate, post-graduate, NA}
MotherQualification	Mother's qualification (education)	{no education, elementary, secondary, graduate, post-graduate, NA}
ExtracurricularActivityStatus	Extracurricular activity status	{Good, Average, Poor}
FoodHabit	Students' Food Habit	{Vegetarian, Non-Vegetarian}
OtherHabit	Students' Other Habit	{Smoking, Drinking, Both, NA}
HometownLocation	Students' Hometown Location	{Big-city, City, Village}
10thEnglishGrade	Grade in English in 10th class (SSC)	{Distinction 75% - 100%, First-class 60% -74%, Second-class < 60%}
10thMathsGrade	Grade in Maths in 10th class (SSC)	{Distinction 75% - 100%, First-class 60% -74%, Second-class < 60%}
12thEnglishGrade	Grade in English in 12th class (HSSC)	{Distinction 75% - 100%, First-class 60% -74%, Second-class < 60%}
12thMathsGrade	Grade in Maths in 12th class (HSSC)	{Distinction 75% - 100%, First-class 60% -74%, Second-class < 60%}
CommunicationSkill	Students' communication skill	{Good, Average, Poor}
Btech1stYearGrade	Grade obtained in B.Tech 1st year	{S 90% - 100%, A 80% - 89%, B 70% - 79%, C 60% - 69%, D 50% - 59%, E 40% - 49%, F < 40%}

5. Data Processing

Data preprocessing is very important task in KDD. It aims at cleaning the data for data mining process, so that to obtained efficient final result.

A. Data Integration, Data Cleaning and Data Transformation

Data Integration: First, one single dataset is created from all collected data. Whole data is integrated into one single dataset which consists of 200 students.

Data Cleaning: Data record which contains missing values can be filled with imputation method. In ourdataset only one record is found to have missing value in one attribute. So it is eliminated, as the complete information is not provided by a student.

Data Transformation:Some numeric attributes used while data collecting are transformed afterwards into nominal attributes. Further, transformation is done by transforming continuous attributes into discrete attributes. It gives better readability and understanding of the data. For example, grade output attribute values are transformed into nominal values as follows:

S: marks between 90 % to 100%
A: marks between 80 % to 89%
B: marks between 70 % to 79%
C: marks between 60 % to 69%
D: marks between 50 % to 59%

10 fold cross-validation is used for dividing the dataset into 10 parts of training data and test data. Thus once preprocessing is done dataset is partitioned into 10 folds having 25 attributes (variables) of 150 students. The normal problem that arises when dealing with dataset is its high dimensionality which means very large numbers of attributes/features are present, but some attributes/features are not required for classification. Also the problem of data imbalance may arise because one class contains very large number of records/instances and other class contains very less number of records/instances. Therefore, data instances cannot be classified correctly in case of class having very less number of instances. These problems can be solved as described next.

B. Feature Selection

The goal of feature selection is find out the highly ranked features/attributes/variables which results in prediction of output variable (Btech1stYearGrade). In this way, the identification of important features is carried out so that to solve high dimensionality problem of data for better classification. Feature selection is performed in WEKA using various algorithms and attributes which are not relevant are removed from the dataset. WEKA offers many attribute selection algorithms. So among these algorithms, ten algorithms are chosen as follows: Cfs Subset Eval, Consistency Subset Eval, Chi Squared AttributeEval, One RAttribute Eval, Filtered Attribute Eval, Filtered Subset Eval, Info Gain Attribute Eval, Gain Ratio Attribute Eval, Symmetrical Uncert- Attribute Eval, ReliefF Attribute Eval. These algorithms rank the attributes for best selection of attributes among 25 attributes in the dataset. Ranking is done by identifying which attributes are chosen by three or more algorithms i.e. finding the frequency of attribute selection. Table 2 shows the frequency count of attributes selected by feature selection.

Table 2: Ranking of Attributes

Students' Attribute Name	Frequency count
10thBoard	10
NumberOfArrears, 12thBoard	8
FamilyAnnualIncome, 10thGrade	7
12thGrade	3
MotherQualification, AgeDuringBtechAddmission	2
FamilySize	1

C. Data Balancing

Data imbalance is said to be one class (minor class) has much lesser instances than that of another class/classes (major class). Traditional (classical) algorithms for classification results in good accuracy of prediction but are less sensitive to the minor class having much lesser instances. This problem can be resolve at data preprocessing stage by balancing or sampling distribution of class. Many algorithms are available for data balancing. The popular algorithm used for data balancing is Synthetic Minority Over-Sampling Technique, abbreviated as SMOTE [15]. In SMOTE, over-sampling of minor class is done by choosing k nearest neighbours having minor class. In our work, the attributes obtained after feature selection are balanced by applying SMOTE algorithm. Figure 2 shows the imbalance in students' data; where class A (grade A) is major class which is 44 % of the dataset and other classes are minor classes specially class B (grade B) and class D (grade D) both are 12.5 % of the dataset. Figure 3 shows the balanced data obtained after using SMOTE algorithm.

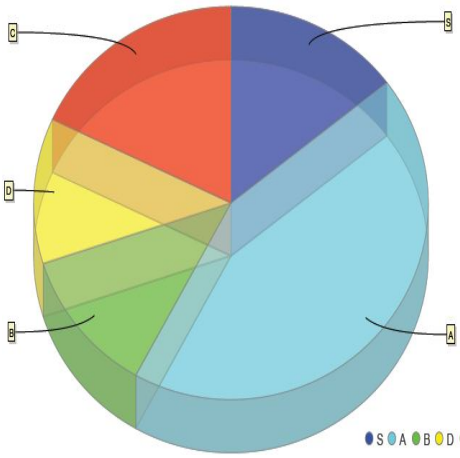


Figure 2: View of imbalanced data

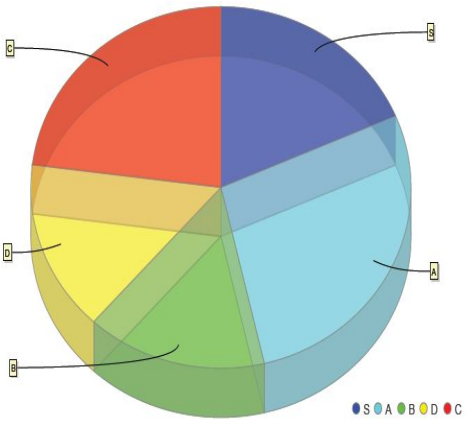


Figure 3: View of balanced data

D. Automatic Parameter Tuning

In recent years the area of automatic parameter tuning has attained a lot of interest [17]. It is the performance optimization technique used to obtain the best results by

adjusting parameters of the algorithm [18]. The use of appropriate parameters increases the accuracy of the algorithm. Parameter tuning involves methods such as using default values, reducing parameters and automatic parameter tuning. In this study, automatic parameter tuning is used which finds the better values of parameters as compared to the defaults. Using automatic parameter tuning, the classifiers which improved accuracy after data balancing step are shown in this study. Rule mining classifiers JRip, PART and ICRM; and decision tree classifiers J48 and RandomForest and REPTree are found to gain better accuracy, easily readable and understandable results.

6. Data Mining Using Classification Algorithms

A. Classical Classification Algorithms

After successful completion of data preprocessing stage, the dataset is now ready for mining process and the next stage to be performed is Data mining. Data mining provides various classification algorithms for prediction of final result. For prediction of students' performance, 19 classification algorithms are applied using WEKA Data mining tool. They are as follows based on:

- Bayes: BayesNet, NaiveBayes
- Functions: Logistic, MultilayerPerceptron, SimpleLogistic, SMO
- Rules: DecisionTable, JRip, NNge, PART, Ridor
- Trees: DecisionStump, J48, LMT, NBTree, RandomForest, RandomTree, REPTree, SimpleCart

B. Grammar based Genetic Programming Algorithm

The grammar based genetic programming algorithm is an evolutionary algorithm grounded on evolution of Darwinian Theory. Genetic programming (GP) codifies individuals by employing genetic operators as mutation and crossover. The main objective of GP is computer program evolution. Genetic algorithms (GAs) are used for rule learning i.e. genetic rule based system. GP is a variant of GA which is a technique of machine learning for optimizing computer program population. Genetic programming is used in problem optimization, searching and in problems of classification. In grammar based genetic programming grammar is specified and process of generating every legal individual is according to the grammar. The grammar based genetic algorithm which is focused in our work is Interpretable Classification Rule Mining (ICRM) algorithm.

ICRM Algorithm:

ICRM algorithm is a technique of rule mining which is grounded on rule based classifiers and gives better classification performance [16]. We have used an efficient ICRM algorithm. There are three stages in ICRM. First stage, in which rules are created based on attributes. In the second stage, algorithm process iteratively to construct the classifier by obtaining the best rule of classification. In the final third stage, classifier accuracy is optimized. These stages of ICRM algorithm are shown in Fig 4.

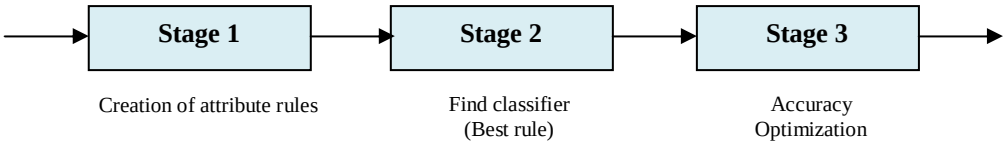


Figure.4. Stages in grammar based genetic programming

6. Experiments and Model Building

The dataset consists of 200 students and the output attribute is grade in B.Tech first year which is to be predicted. Prediction of students' performance based on input attributes is carried out in WEKA by executing all mentioned 19 algorithms, by using 10 fold cross-validation and default parameters.

A. Analyzing Accuracy Performance

The accuracy achieved by these algorithms in case 1) using all attributes, 2) using feature selection and 3) using data balancing is shown in Table 3.

Table 3: Accuracy of Classifiers using Case 1, 2, 3

Algorithm (Classifier)	**Accuracy (%) using All attributes**	**Accuracy (%) using Feature Selection**	**Accuracy (%) using Data Balancing**
BayesNet	66.5	70	69.968
NaiveBayes	65.5	67	71.246
Logistic	56	66	70.287
MultilayerPerceptron	64.5	62.5	**74.44**
SimpleLogistic	67.5	63.5	70.287
SMO	61.5	**71.5**	69.968
DecisionTable	68	66	66.773
JRip	69.5	**71.5**	70.607
NNge	63	62.5	70.287
PART	58.5	68	71.885
Ridor	65	63	68.051
J48	**70.5**	**71.5**	70.607
LMT	67.5	63.5	71.565
NBTree	67.5	65	72.204
RandomForest	62.5	64	73.801
RandomTree	59.5	65.5	72.524
REPTree	65	65.5	70.287
SimpleCart	68	69	70.287
ICRM	64	62.5	69.329

From Table 3, it is clear that

- Using all attributes, the highest accuracy achieved is by J48 algorithm which is 70.5 %.
- Using feature selection, some classifiers have improved their accuracy. The highest accuracy achieved is by J48, JRip and SMO algorithm which is 71.5 % for all three.
- Using data balancing, many classifiers achieved better accuracy as compared to case 1 and 2. The highest accuracy achieved is by MultilayerPerceptron which is 74.44 %. After that RandomForest obtained accuracy of 73.8 %. Then RandomTree and NBTree have obtained accuracy of 72.52 % and 72.204 % respectively.

In all the three cases, algorithms which yield consistence performance are J48 and JRip. Though in the third case both algorithms obtained accuracy of 70.607 % (less than highest i.e. 74.44 % by MultilayerPerceptron), they achieve much better accuracy in all the three cases maintaining consistency.

The ICRM algorithm that used in this study has achieved the accuracy of 69.329 % using data balancing (case 3). This accuracy is less as compared to the algorithms that yielded higher prediction accuracy and found efficient i.e. MultilayerPerceptron, RandomForest, RandomTree, NBTree, JRip, and J48. But as the benefits of ICRM algorithms are considered, this accuracy is acceptable at this stage which can be further checked using parameter tuning. Also the rules discovered by ICRM (Fig. 9) are optimized rules showing least number of rules possible for efficient prediction of students' grade.

Figure 5 represents the comparison of accuracy of all 19 algorithms showing classical algorithms and evolutionary ICRM algorithm using all three cases.

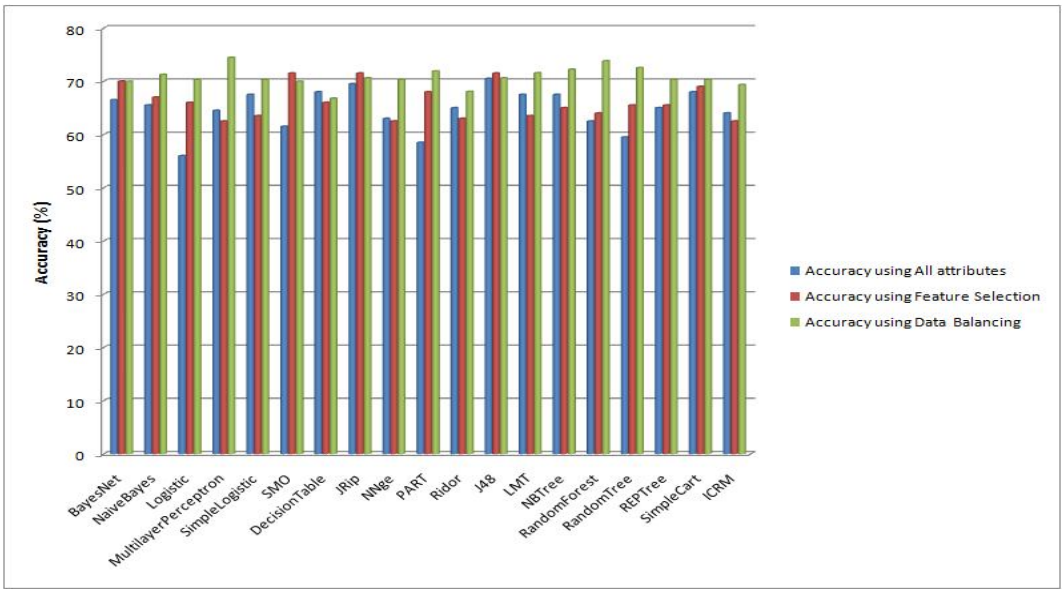


Figure 5: Comparison of classifiers accuracy using Case 1, 2, 3

From Table 5, it is clear that the applying automatic parameter tuning method on balanced data has improved the accuracy of some classifiers. Thus using parameter tuning, RandomForest achieved the highest accuracy of 76.0383 %, and this is the highest accuracy obtained as compared to case 1, 2 and 3; whereas an accuracy of ICRM classifier remains same as 69.32 % (less as compared to other classifiers), thus it is not included in Table 4 and Table 5. Cross-validated Parameter selection is shown in Table 4 as follows:

Table 4: Cross-validated Parameter Selection

Algorithm (Classifiers)	Default Parameters	Selected Parameters	Meaning
JRip	-F 3 -N 2.0 -O 2 -S 1	-F 3 -N 1.0 -O 2 -S 1	F: folds, N: minNo, O: optimizations, S: seed
PART	-M 2 -C 0.25 -Q 1	-M 1 -C 0.25 -Q 1	M: minNumObj, C: confidenceFactor, Q: seed
J48	-C 0.25 -M 2	-C 0.25 -M 1	C: confidenceFactor, M: minNumObj
RandomForest	-I 10 -K 0 -S 1 -print -num-slots 1	-I 7 -K 1 -S 1 -print -num-slots 1	I: numTrees, K: numFeatures, S: seed, num-slots: numExecutionSlots
REPTree	-N 3 -S 1 -V 0.001 -M 2 -L -1 -I 0.0	-N 4 -S 3 -V 0.001 -M 1 -L -1 -I 0.0	N: numFolds, S: seed, V: minVarianceProp, M: minNum, L: maxDepth, I: initialCount

Table 5: Accuracy of Classifiers using Data Balancing and Parameter Tuning

Algorithm (Classifiers)	Accuracy (%) using Data Balancing	Accuracy (%) using Parameter Tuning
JRip	70.607	72.843
PART	71.885	72.204
J48	70.607	72.524
RandomForest	73.801	**76.038**
REPTree	70.287	71.565

Figure 6 represents the comparison of improved classifiers accuracy using parameter tuning. Classification model JRip, PART, J48, RandomForest and REPTree

has improved their accuracy. RandomForest achieved the highest accuracy of 76.038 %.

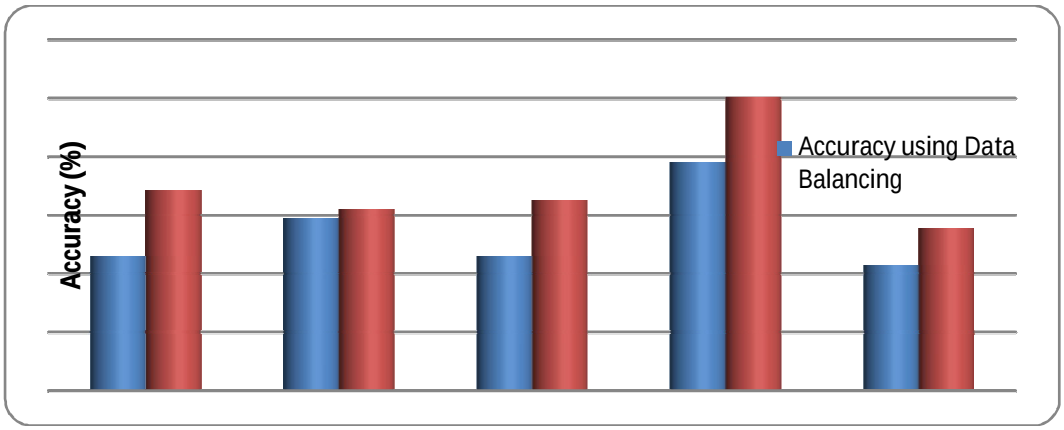


Figure 6: Comparison of improved accuracy using Parameter Tuning

B. Rules Discovery

Figure 7 shows the JRip rules discovered using parameter tuning. Number of rules discovered is 10. All rules predict the grade of students in B.Tech 1st year. For example, the second rule shows that if 10th Grade is C, then Grade in B.Tech 1st year is D. Fifth rule shows that if 10th Grade is A and Number of arrears is zero and 12th Grade is B, then Grade in B.Tech 1st year is B. In this way, all grades are predicted by these rules.

```
JRIP rules:
===========

(FamilyAnnualIncome = High) and (12thGrade = S) and (10thGrade = S) => Btech1stYearGrade=D (48.0/2.0)
(10thGrade = C) => Btech1stYearGrade=D (1.0/0.0)
(FamilyAnnualIncome = High) and (NumberOfArrears = seven-eight-nine) => Btech1stYearGrade=D (1.0/0.0)
(10thGrade = B) => Btech1stYearGrade=B (13.0/4.0)
(10thGrade = A) and (NumberOfArrears = zero) and (12thGrade = B) => Btech1stYearGrade=B (17.0/8.0)
(10thBoard = CBSE) and (12thGrade = S) and (10thGrade = S) => Btech1stYearGrade=S (46.0/10.0)
(NumberOfArrears = one-two-three) and (10thBoard = SB) => Btech1stYearGrade=C (46.0/2.0)
(NumberOfArrears = seven-eight-nine) => Btech1stYearGrade=C (24.0/0.0)
(NumberOfArrears = four-five-six) => Btech1stYearGrade=C (2.0/0.0)
 => Btech1stYearGrade=A (115.0/50.0)

Number of Rules : 10
```

Figure 7: JRIP rules discovered using Parameter Tuning

Figure 8 shows J48 decision tree discovered using parameter tuning. All attributes predict the grade of students B.Tech 1st year. For example, the first rule shows that if Number of arrears is zero, Family annual income is middle-class, 10th Grade is S, 12thGrade is S and 10th board is CBSE then Grade in B.Tech 1st year is S. In this manner, the pruned tree predicts all grades of students.

```
J48 pruned tree
------------------

NumberOfArrears = zero
|   FamilyAnnualIncome = Middle-class
|   |   10thGrade = S
|   |   |   12thGrade = S
|   |   |   |   10thBoard = CBSE: S (44.0/10.0)
|   |   |   |   10thBoard = ICSE
|   |   |   |   |   12thBoard = CBSE: S (1.0)
|   |   |   |   |   12thBoard = ICSE: A (4.0/1.0)
|   |   |   |   |   12thBoard = SB: A (0.0)
|   |   |   |   10thBoard = SB: A (3.0/1.0)
|   |   |   12thGrade = A: A (22.0/4.0)
|   |   |   12thGrade = B: A (5.0/1.0)
|   |   |   12thGrade = C: B (1.0)
|   |   10thGrade = A
|   |   |   12thGrade = S
|   |   |   |   10thBoard = CBSE: A (5.0/1.0)
|   |   |   |   10thBoard = ICSE: A (1.0)
|   |   |   |   10thBoard = SB: B (9.0/3.0)
|   |   |   12thGrade = A: A (43.0/23.0)
|   |   |   12thGrade = B: B (15.0/6.0)
|   |   |   12thGrade = C: A (0.0)
|   |   10thGrade = C: A (0.0)
|   |   10thGrade = B
|   |   |   12thBoard = CBSE: A (2.0)
|   |   |   12thBoard = ICSE: B (0.0)
|   |   |   12thBoard = SB: B (3.0)
|   FamilyAnnualIncome = High
|   |   12thGrade = S: D (49.0/3.0)
|   |   12thGrade = A: A (6.0/1.0)
|   |   12thGrade = B: A (2.0)
|   |   12thGrade = C: D (0.0)
|   FamilyAnnualIncome = Below-poor
|   |   10thGrade = S: S (1.0)
|   |   10thGrade = A: B (1.0)
|   |   10thGrade = C: S (0.0)
|   |   10thGrade = B: S (0.0)
|   FamilyAnnualIncome = Poor
|   |   10thBoard = CBSE: S (2.0)
|   |   10thBoard = ICSE: S (0.0)
|   |   10thBoard = SB: A (1.0)
NumberOfArrears = seven-eight-nine
|   FamilyAnnualIncome = Middle-class: C (25.0)
|   FamilyAnnualIncome = High: D (1.0)
|   FamilyAnnualIncome = Below-poor: C (0.0)
|   FamilyAnnualIncome = Poor: C (0.0)
NumberOfArrears = four-five-six
|   10thGrade = S: C (0.0)
|   10thGrade = A: C (2.0)
|   10thGrade = C: D (1.0)
|   10thGrade = B: C (0.0)
NumberOfArrears = one-two-three
|   10thGrade = S
|   |   12thGrade = S: S (1.0)
|   |   12thGrade = A: B (1.0)
|   |   12thGrade = B: S (0.0)
|   |   12thGrade = C: S (0.0)
|   10thGrade = A
|   |   10thBoard = CBSE
|   |   |   12thBoard = CBSE: A (5.0/2.0)
|   |   |   12thBoard = ICSE: B (1.0)
|   |   |   12thBoard = SB: B (2.0)
|   |   10thBoard = ICSE: B (1.0)
|   |   10thBoard = SB: C (46.0/2.0)
|   10thGrade = C: C (0.0)
|   10thGrade = B
|   |   12thBoard = CBSE: A (1.0)
|   |   12thBoard = ICSE: B (0.0)
|   |   12thBoard = SB: B (6.0)
```

Figure 8:. Decision tree J48 structure discovered using Parameter Tuning

Figure 9 shows rules discovered by ICRM algorithm using data balancing. ICRM algorithm obtains as many numbers of rules as there as number of classes. Thus, it can be seen that 5 rules are generated for 5 output classes. These are optimized rules. Each rule predicts unique grade (class). For example, the third rule shows that if Grade in 12^{th} is S and 12^{th} Board is CBSE, then Grade in B.Tech 1^{st} year is S. In this way, 5 unique rules predict 5 unique grades.

```
=== Classifier model (full training set) ===

 1 Rule: IF (AND = FamilyAnnualIncome High = 12thGrade S ) THEN (Btech1stYearGrade = D)
 2 Rule: ELSE IF (!= NumberOfArrears zero ) THEN (Btech1stYearGrade = C)
 3 Rule: ELSE IF (AND = 12thGrade S = 12thBoard CBSE ) THEN (Btech1stYearGrade = S)
 4 Rule: ELSE IF (AND != 10thGrade S != 12thBoard CBSE ) THEN (Btech1stYearGrade = B)
Default: (Btech1stYearGrade = A)
```

Figure 9: ICRM rules discovered using Data Balancing

C. Analyzing Different Performance Measures

Effect of Data Balancing: In addition to obtaining prediction accuracy, some other important performance measures are Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Receiver Operating Characteristics (ROC) Area. Classification models are evaluated based on these widely known measures. Table 6 shows the values of these measures obtained by 19 classification models using data balancing. Moreover, the comparative result of models based on these measures is shown in Figure 10.

Table 6: MAE, RMSE and ROC area Values of Classifiers using Data Balancing

Algorithm (Classifier)	**MAE**	**RMSE**	**ROC Area**	**Accuracy (%) using Data Balancing**
BayesNet	0.1381	0.2798	0.905	69.968
NaiveBayes	0.1403	0.2797	0.906	71.246
Logistic	0.1411	0.2899	0.885	70.287
MultilayerPerceptron	0.1292	0.2857	0.903	**74.44**
SimpleLogistic	0.154	0.2831	0.896	70.287
SMO	0.2567	0.3407	0.864	69.968
DecisionTable	0.1865	0.2899	0.886	66.773
JRip	0.1512	0.2884	0.864	70.607
NNge	**0.1188**	0.3447	0.813	70.287
PART	0.1396	0.2826	0.883	71.885
Ridor	0.3575	**0.1278**	0.796	68.051
J48	0.1407	0.287	0.88	70.607
LMT	0.1442	0.2887	0.89	71.565
NBTree	0.1335	0.2807	0.904	72.204
RandomForest	0.1316	0.2738	**0.907**	**73.801**
RandomTree	0.1314	0.298	0.867	72.524
REPTree	0.147	0.289	0.88	70.287
SimpleCart	0.1484	0.2895	0.87	70.287
ICRM	**0.1227**	0.3503	0.806	**69.329**

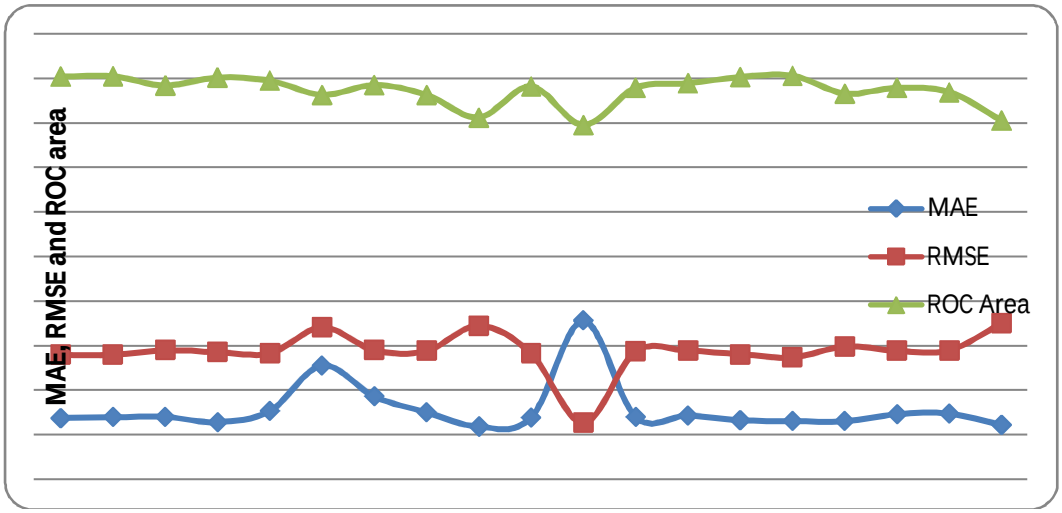


Figure 10: Comparison of MAE, RMSE and ROC area values of classifiers using Data Balancing

Mean Absolute Error (MAE) is the measure of goodness of classification model and prediction accuracy. It measures the mean magnitude of the residuals or errors. In case of large errors, MAE is less sensitive. Lower the MAE value, more accurate the model is. The lowest MAE value is achieved by NNge (0.1188) and then by ICRM which is 0.1227. The MAE is given as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$

Where, 'y_j' is the predicted value, 'y'_j'is the actual value and 'n' denotes the number of predictors.

Root Mean Squared Error (RMSE) is the square root of the mean squared distance of fitted line from data points. RMSE is more sensitive to large errors. Lower the RMSE value, more accurate is the prediction model. The lowest RMSE value is achieved by Ridor which is 0.1278. The RMSE is given as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2} \qquad .(1)$$

Where, similarly 'y_j' is the predicted value, 'y'_j'is the actual value and 'n' denotes the number of predictors.

The ROC area curve graph shows true positives versus false positive. It represents the area under curve which stands for the accuracy of classification model. The model is said to be more accurate if area (value) is larger. The ROC area curve for 19 classification models (Table 6, Fig. 10) shows the ROC area value greater than 0.7, this clears that all the classification models performs well and have reliability in

prediction. The highest value is gained by the RandomForest which is 0.907, whereas the ICRM has obtained less ROC area value i.e. 0.806 using data balancing.

Effect of Automatic Parameter Tuning:Using automatic parameter tuning some classifiers has obtained better values of MAE, RMSE and ROC area. Table 7 shows the values of these measures using data balancing (case 3) and parameter tuning. Using automatic parameter tuning, RandomForest achieved lowest value of MAE, RMSE and ROC area. Thus the most efficient classifier is RandomForest; also it has achieved highest accuracy 76.038 %, whereas the accuracy of ICRM classifier remains same as 69.32 %. So ICRM is not shown in Table 7.

Table 7: MAE, RMSE and ROC area values of classifiers using Data Balancing and Parameter Tuning

Algorithm (Classifiers)	MAE		RMSE		ROC Area	
	Data Balancing	**Parameter Tuning**	**Data Balancing**	**Parameter Tuning**	**Data Balancing**	**Parameter Tuning**
JRip	0.1512	0.1426	0.2884	0.2791	0.864	0.882
PART	0.1396	0.1342	0.2826	0.2844	0.883	0.879
J48	0.1407	0.1355	0.287	0.2869	0.88	0.878
RandomForest	0.1316	**0.1331**	0.2738	**0.2778**	0.907	**0.897**
REPTree	0.147	0.1463	0.289	0.2914	0.88	0.88

ROC curve is a graphical plot between the true positive rate (TP Rate) and the false positive rate (FP Rate). The good ROC curve is that which is closer to the top left corner. Connecting points (0, 0) and (1, 1) forms a straight line which shows random classifier having even odds. The benefit of ROC curve is to identify region for which the model is better as compared to another. Using ROC curve, the performance of learner can be summarized to represent it in a single quantity. Larger area under curve (AUC) is always better. The classifier that gives good performance in the region of higher TP rate is favoured over others which don't.

Figures 11.1, 11.2, 11.3, 11.4 and 11.5 shows the ROC curve of class S, class A, class B, class C and class D for improved classifiers (JRip, PART, J48, RandomForest, REPTree) using automatic parameter tuning respectively. It can be seen from the graphs that the AUC for RandomForest has higher TP rate at low FP rate as compared to other classifiers. Thus it is visible from the graph that RandomForest performs best in this study.

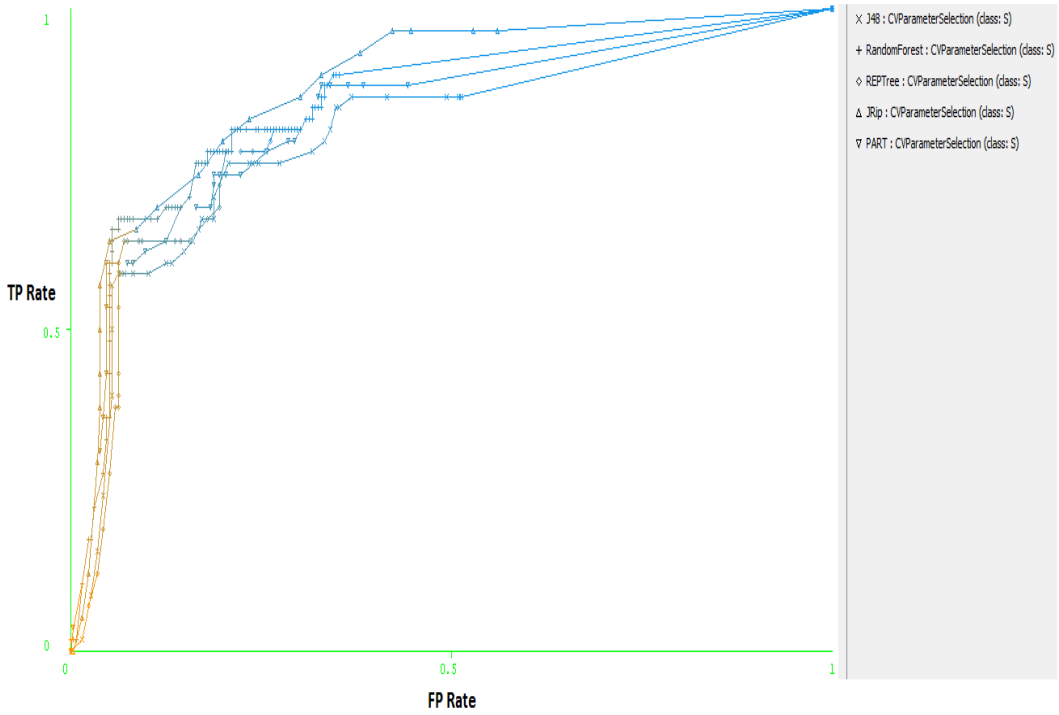


Figure 11: 1 ROC curve of class S for improved classifiers

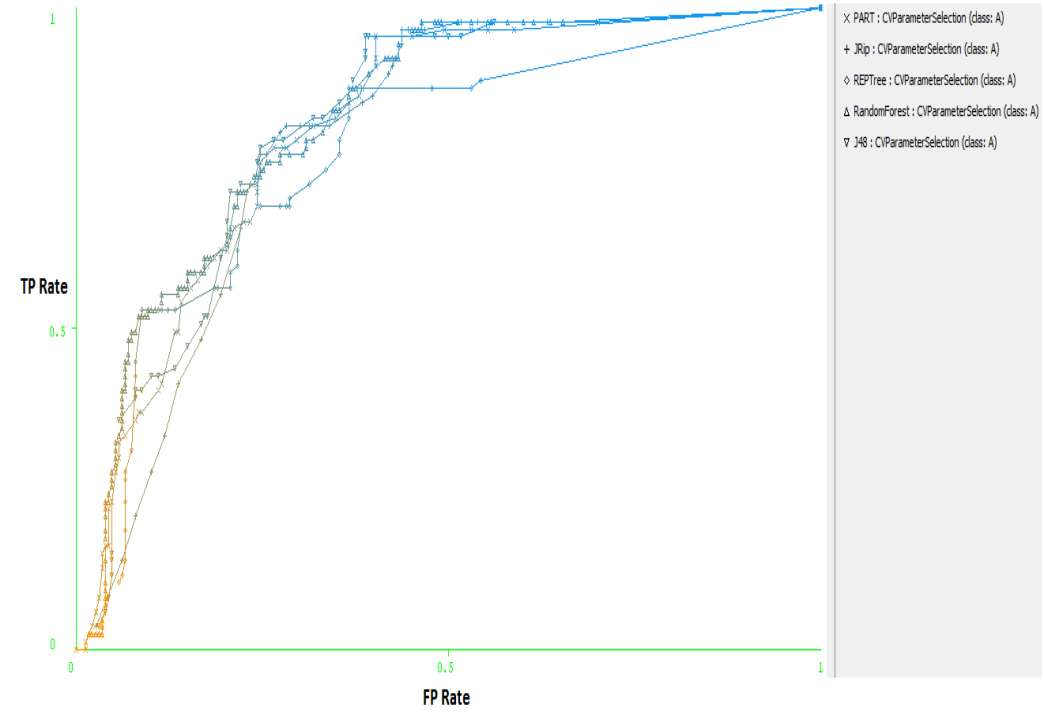


Figure 11: 2 ROC curve of class A for improved classifiers

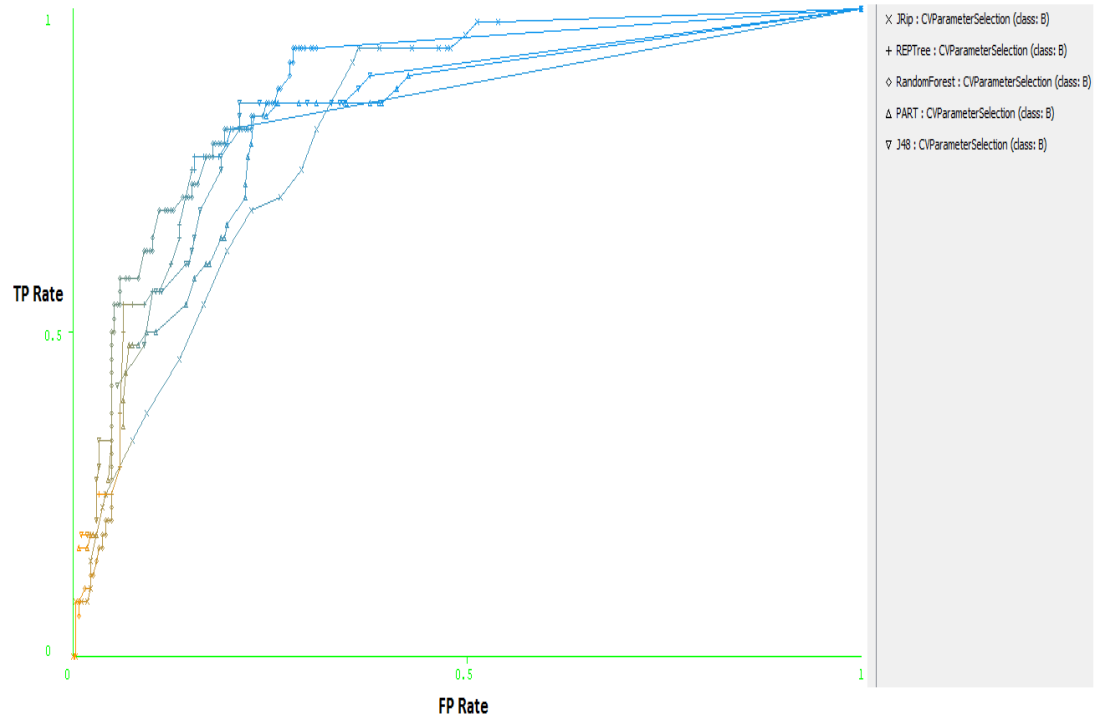


Figure 11:3 ROC curve of class B for improved classifiers

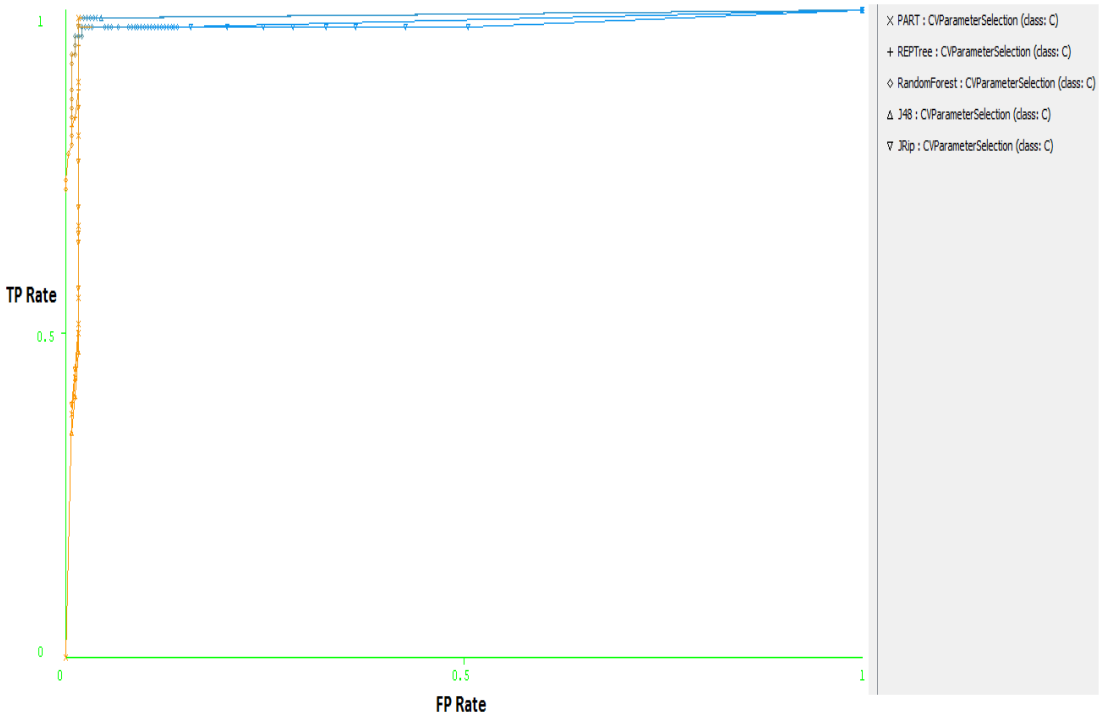


Figure 11:4 ROC curve of class C for improved classifiers

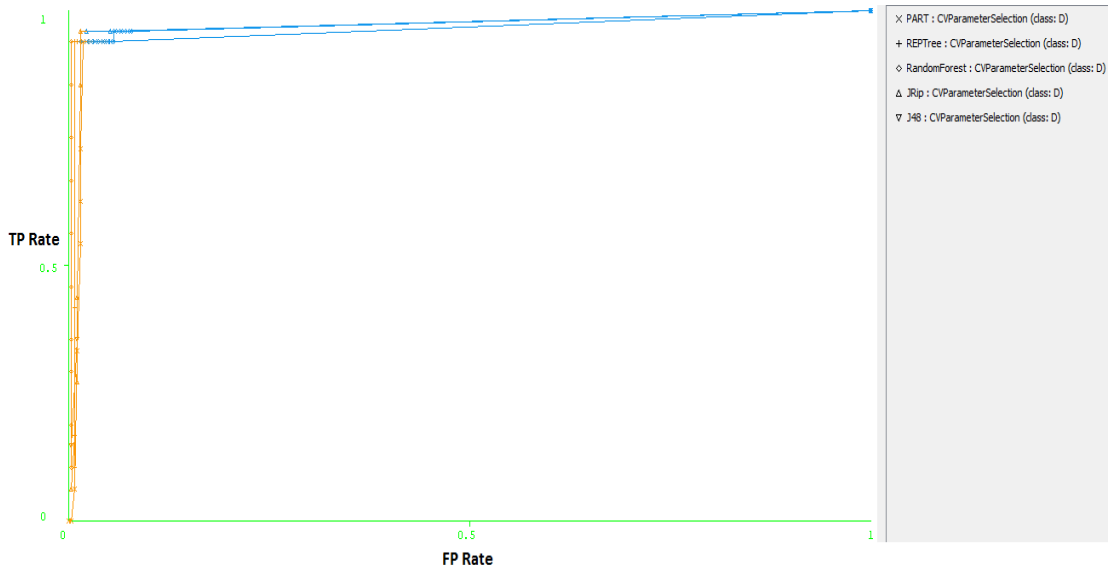


Figure 11:5 ROC curve of class D for improved classifiers

D. Relationship of Predictors with Output (Class) Attribute

In our study of predicting students' performance, the relationship of input attributes (predictors) selected by feature selection algorithms with the output attribute shows the contribution or effect of each predictor on students' grade prediction in B.Tech 1st year. This relationship is analyzed and it is found that what values of predictors are responsible for predicting specific grade. Predictors with high frequency count (Table 2) i.e. 10thBoard, NumberOfArrears, 12thBoard, FamilyAnnualIncome and 10thGrade have much greater influence on prediction. So their relationship is discovered.

Relationship between Btech1stYearGrade and 10thBoard:

Figure 12 shows the relationship between B.Tech 1st year grade and 10th board. The following observations are performed for each grade of output attribute:

Grade S: In this category, CBSE students are more than ICSE and SB students.

i.e. CBSE > ICSE > SB

Grade A: In this category, ICSE students are more than CBSE and SB students.

i.e. ICSE > CBSE > SB

Grade B: In this category, ICSE students are more than SB students. CBSE students are least in this category.

i.e. ICSE > SB > CBSE

Grade C: In this category, SB students are more than CBSE students. No ICSE student is there in this study.

i.e. SB > CBSE

Grade D: In this category, CBSE students are more than SB students. No ICSE student is there in this category.

i.e. CBSE > SB.

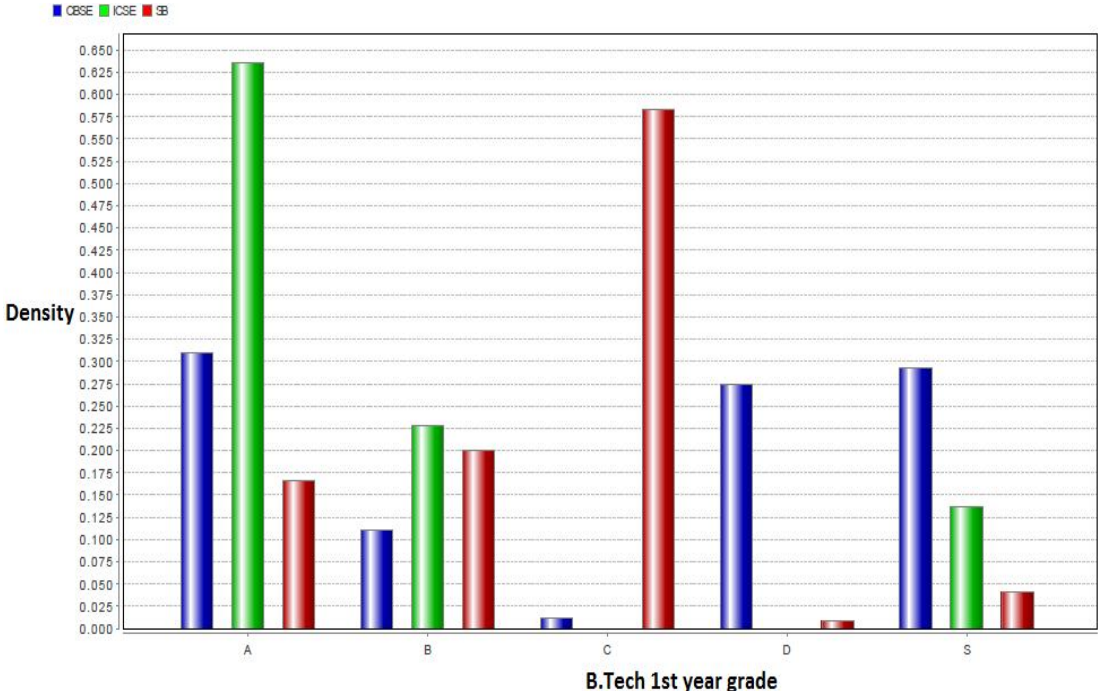


Figure 12: Relationship between Btech1stYearGrade and 10thBoard

Relationship between Btech1stYearGrade and NumberOfArrears: Figure 13 shows the relationship between B.Tech 1st year grade and Number of arrears. The following observations are performed for each grade of output attribute:

Grade S: In this category, students having only zero and on-two-three arrears are present. Number of students having zero arrears is more than that of having one-two-three arrears.

i.e. zero > one-two-three

Grade A: In this category also, number of students having zero arrears is more than that of having one-two-three arrears. As compared to Grade S observation, here more number of students has one-two-three arrears.

zero> one-two-three

Grade B: In this category, number of students having one-two-three arrears is more than that of having zero arrears.

i.e. one-two-three > zero

Grade C: In this category, number of students having seven-eight-nine arrears is more than that of having one-two-three and four-five-six arrears.

i.e. seven-eight-nine > one-two-three > four-five-six

Grade D: In this category number of students having four-five-six arrears is more than that of having zero and seven-eight-nine arrears.

i.e. four-five-six > zero > seven-eight-nine

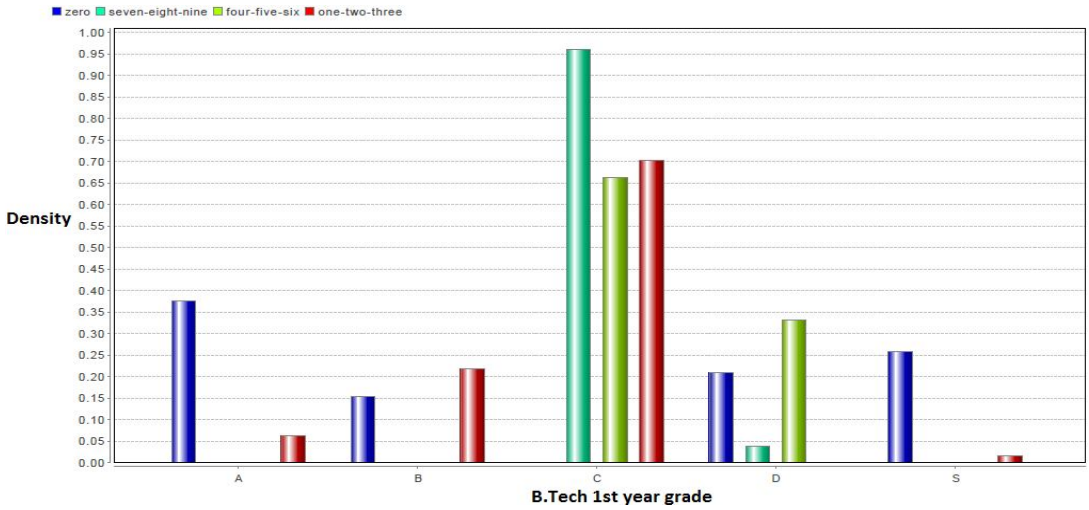


Figure 13: Relationship between Btech1stYearGrade and NumberOfArrears

Relationship between Btech1stYearGrade and 12thBoard: Figure 14 shows the relationship between B.Tech 1st year grade and 12th board. The following observations are performed for each grade of output attribute:

Grade S: In this category, CBSE students are more than ICSE students. SB students are least.

i.e. CBSE > ICSE > SB

Grade A: In this category, ICSE students are more than that of CBSE students. SB students are least.

i.e. ICSE > CBSE > SB

Grade B: In this category, ICSE students are more than that of SB. CBSE students are least.

i.e. ICSE > SB >CBSE

Grade C: In this category, SB students are more than that of CBSE. No ICSE student is found in the study.

i.e. SB > CBSE

Grade D: In this category, SB students are more than that of CBSE. No ICSE student is found in the study.

i.e. SB > CBSE

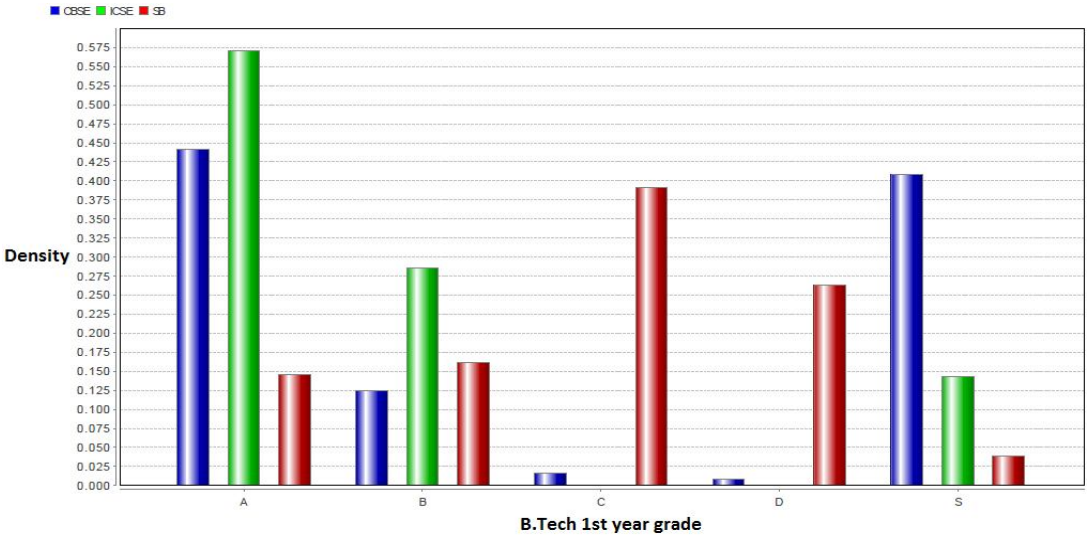


Figure 14: Relationship between Btech1stYearGrade and 12thBoard

Relationship between Btech1stYearGrade and FamilyAnnualIncome: Figure 15 shows the relationship between B.Tech 1st year grade and Family Annual Income. The following observations are performed for each grade of output attribute:

Grade S: In this category, Below-poor and poor students are more than that of middle-class and high class students.

i.e. Below-poor, Poor > Middle-class > High

Grade A: In this category, middle-class students are more than poor and high class students.

i.e. Middle-class > Poor > High

Grade B: In this category, below-poor students are more than middle-class and high class students.

i.e. Below-poor > Middle-class > High

Grade C: In this category, middle-class students are more than high class students.

i.e. Middle-class > High

Grade D: In this category, High class students are more than middle-class students.

i.e. High > Middle-class

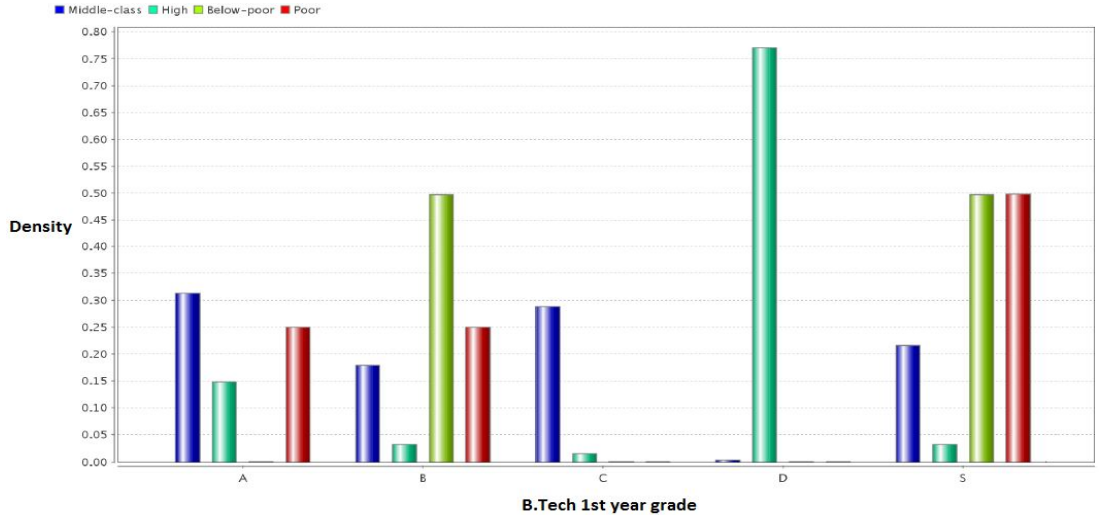


Figure 15: Relationship between Btech1stYearGrade and FamilyAnnualIncome

Relationship between Btech1stYearGrade and 10thGrade: Figure 16 shows the relationship between B.Tech 1st year grade and 10th grade. The following observations are performed for each grade of output attribute:

Grade S: In this category, number of students having grade S is more than that of having grade A.

i.e. S > A

Grade A: In this category, number of students having grade S is more than that of having grade A and grade B.

i.e. S > A > B

Grade B: In this category, number of students having grade B is more than that of having grade S and grade A.

i.e. B > A > S

Grade C: In this category, number of students having grade A is more than that of having grade B and grade S.

i.e. A > B > S

Grade D: In this category, number of students having grade C is more than that of having grade S and grade A.

i.e. C > S > A

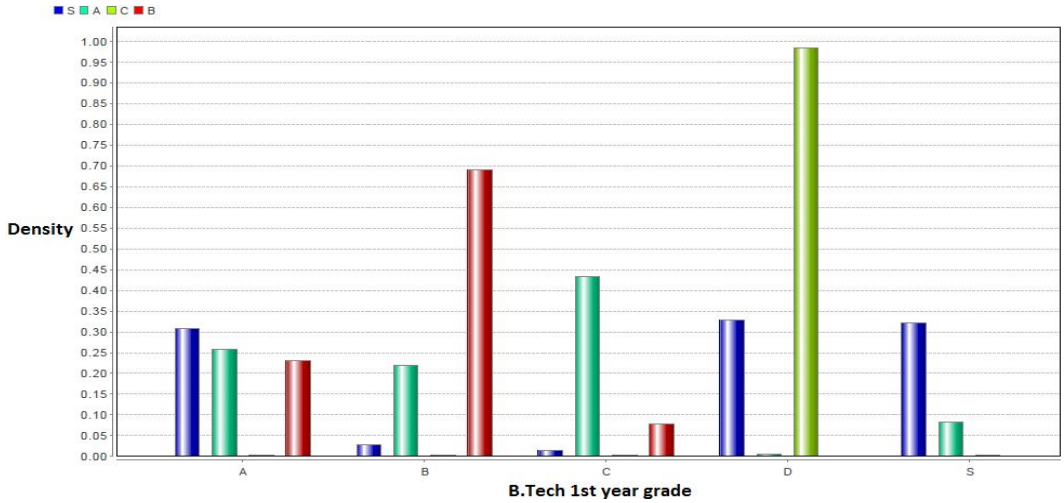


Figure 16: Relationship between Btech1stYearGrade and 10thGrade

7. Results and Discussion

A. Summarization of Relationships

Summarization of all relationships in previous section is shown in Table 8. Thus, it can be easily seen that how each predictor is contributing for the prediction of students' performance (B.Tech1stYearGrade) in VIT University. From this the institute can easily recognize the student who can score the specific grade based on his/her past academic and general record.

Table 8: Relationship between Btech1stYearGrade and Predictors

B.Tech1stYearGrade	10thBoard	NumberOfArrears	12thBoard	FamilyAnnualIncome	10thGrade
Grade S	CBSE > ICSE > SB	zero > one-two-three	CBSE > ICSE > SB	Below-poor, Poor > Middle-class > High	S > A
Grade A	ICSE > CBSE > SB	zero > one-two-three	ICSE > CBSE > SB	Middle-class > Poor > High	S > A > B
Grade B	ICSE > SB > CBSE	one-two-three > zero	ICSE > SB >CBSE	Below-poor > Middle-class > High	B > A > S
Grade C	SB > CBSE	seven-eight-nine > one-two-three > four-five-six	SB > CBSE	Middle-class > High	A > B > S
Grade D	CBSE > SB	four-five-six > zero > seven-eight-nine	SB > CBSE	High > Middle-class	C > S > A

B. Relationship between Top Two Predictors

Figure 17 shows the scatter graph of the relationship between the top two predictors (10^{th} board and number of arrears) for prediction of students' final grade (Btech1stYearGrade). It can be clearly seen that prediction result is almost grade S if student is from 10thBoard as CBSE and NumberOfArrears as zero. Similarly prediction result is almost grade C if student is from 10thBoard as SB and

NumberOfArrears as one-two-three and seven-eight-nine. In this manner, the top two predictors contribute for reaching towards the particular final grade of student.

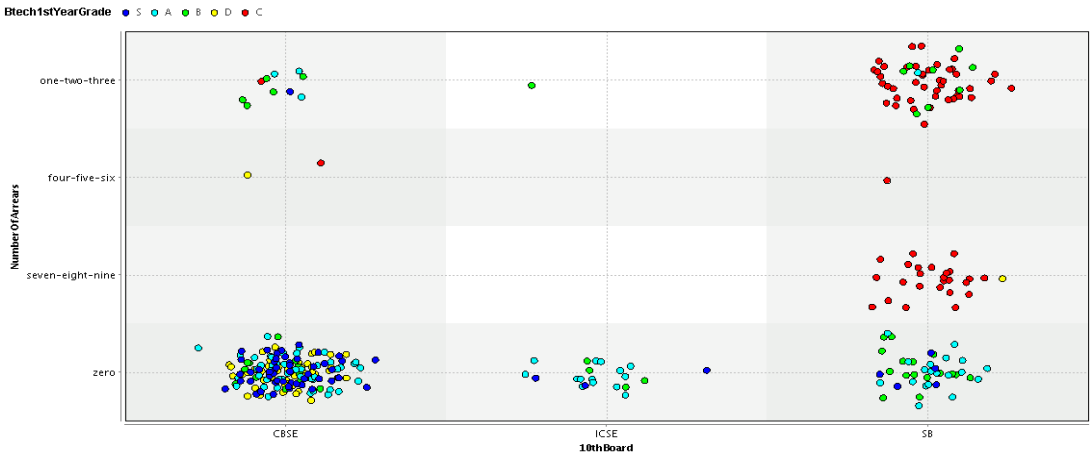


Figure 17: Relationship between 10th Board and NumberOfArrears in terms of final grade (Btech1stYearGrade)

C. Final Prediction Model Building

We have used 19 classifiers for predicting students' performance. Each classifier's performance is studied and they are compared with other. Figure 18 shows the first part of our final model which is built in WEKA. The model consists of all good classifiers, processes, stages and results that are carried out in this paper for evaluation of accuracy and performance. In other word, it shows all the workflow at one place. The model is run and it produces all the results within 10 seconds. Then, it becomes easy see outputs of each classifier in terms of accuracy achieved, trees and rules generated and performance.

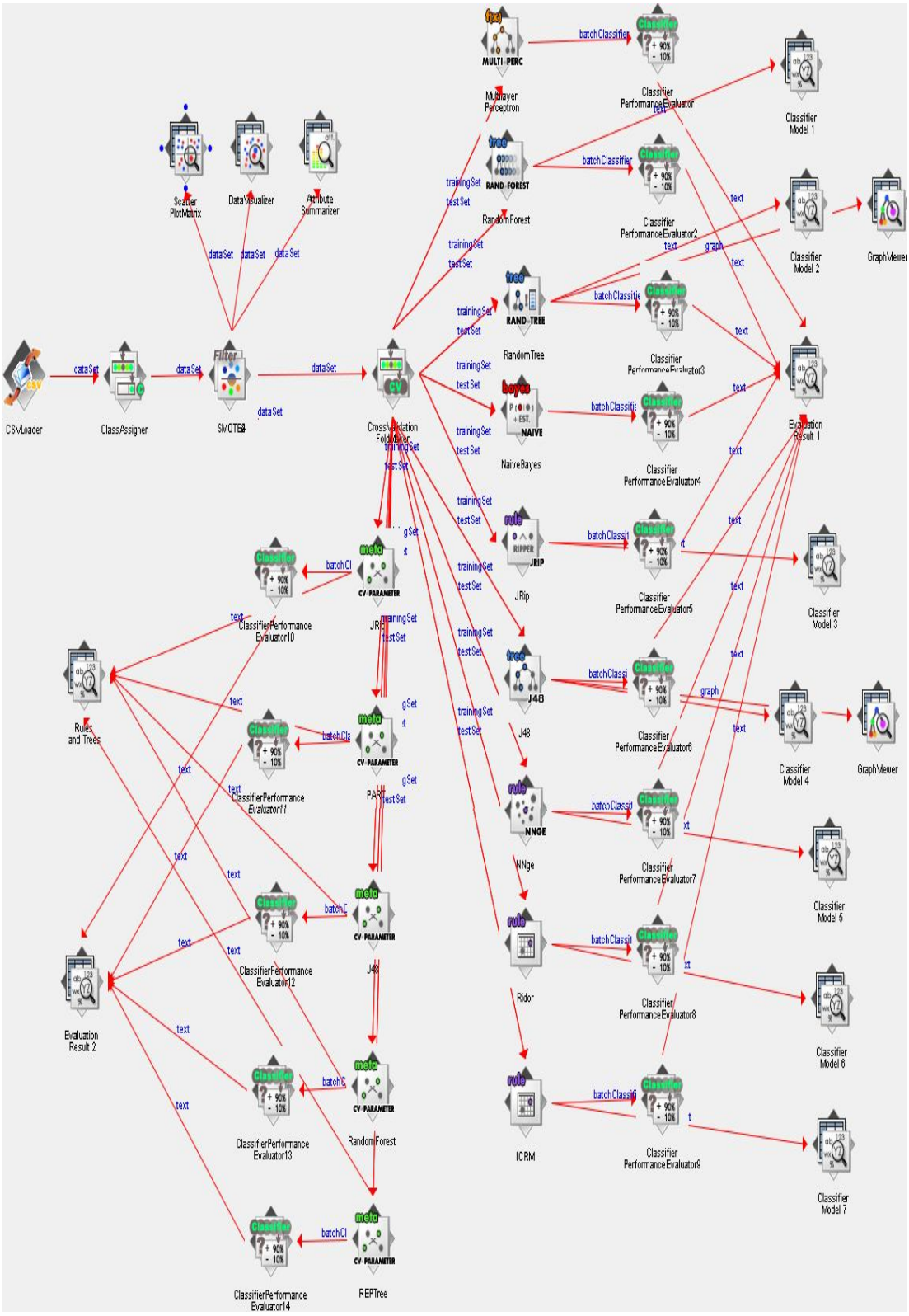
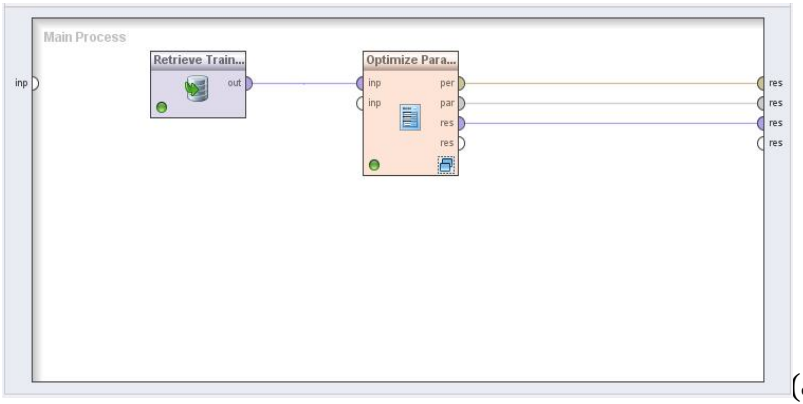


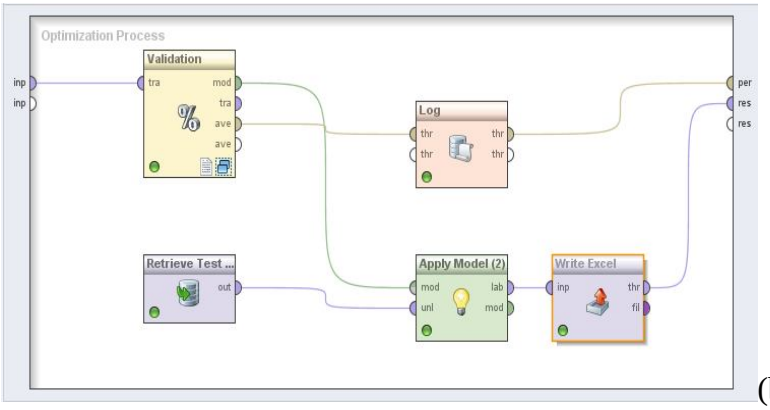
Figure 18: Final Prediction Model (Part 1)

From our study of the classical and evolutionary data mining technique (ICRM classifier), we finally discover that RandomForest performs best (accuracy of 76.038 %) for prediction of students' performance (Grade) in B.Tech 1st year. Thus, it is efficient to use RandomForest classifier model for predicting the grades of students (Btech1stYearGrade). Figure 19.a.b.c shows a view of final model built in RapidMiner5.3 using RandomForest which predicts the students' performance by using the past records of earlier and current students.

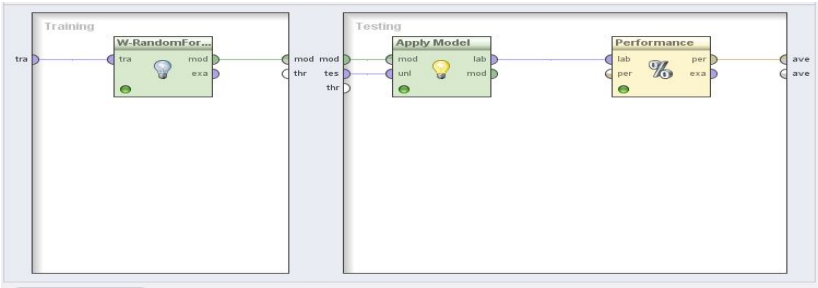
The model accepts the input dataset as test data for which the grades are to be predicted and based on the past academic result of student (Training data); it outputs the students' grades. The result of prediction is then stored in separate Excel file. So, this prediction result now can be used by the institute for identifying how a particular student will perform if he/she joins the institute.



(a)



(b)



(c)

Figure 19: Final Prediction Model (Part 2)

8. Conclusion and Future Work

In our study, we have evaluated the performance of various classifiers for mining the students' records to predict the performance in undergraduate degree. Results of the classical classification models are compared with each other and in turn with the evolutionary grammar based genetic programming model. The performance parameters of each model are analyzed. The grammar based genetic algorithm (GGA), ICRM produces optimized rules for prediction. But the RandomForest achieves highest accuracy. Thus, final model is built using RandomForest for predicting B.Tech 1st year grade. Further, the relationship of these grades with different predictors is determined and the contribution of each predictor in predicting students' grade is studied. Future work here is to gather some more number of attributes and identify their prediction capability and various effects by having an amended prediction model for predicting students' performance in academics.

9. References

[1] Alejandro Pena-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works", In ExpertcSystems with Applications, ScienceDirect, 2013.

[2] B. Sen, E. Ucar, D. Delen, "Predicting and analyzing secondary education placement-test scores: A data mining approach, In Expert Systems with Applications 39, 2012, pp. 9468 – 9476.

[3] Molina, M. M., Luna, J. M., Romero, C., & Ventura, S., "Meta-learning approach for automatic parameter tuning: a case of study with educational datasets", In Proceedings of the 5th international conference on educational data mining, 2012, pp. 180–183.

[4] Marquez-Vera, C., Romero, C., & Ventura, S., "Predicting school failure using" data mining. In Proceedings of the 4th international conference on educational data mining, 2012, pp. 271–275.

[5] Baker, R. S. J. D., Gowda, S. M., & Corbett, A. T., "Automatically detecting a student's preparation for future learning: Help use is key, In Proceedings of the 4th international conference on educational data mining, 2011, pp. 179–188.

[6] M.I. López, J.M Luna, C. Romero, S. Ventura, "Classification via clustering for predicting final marks based on student participation in forums", International Conference on Educational Data Mining (ICOEDM), 2012, pp. 212 – 215.

[7] Ya-huei Wang, Hung-Chang Liao, "Data mining for adaptive learning in a TESL-based e-learning system", In Expert Systems with Applications 38, 2011, pp. 6480-6485.

[8] Rau, M. A., &Pardos, Z. A., "Interleaved practice with multiple representations: analyses with knowledge tracing based techniques", In Proceedings of the 5th international conference on educational data mining, 2012, pp. 168–171.

[9] Wang, Y., & Heffernan, N. T. , "Leveraging first response time into the knowledge tracing model", In Proceedings of the 5th international conference on educational data mining, 2012, pp. 176–179.

[10] Akcapinar, G., Cosgun, E., &Altun, A., "Prediction of perceived disorientation in online learning environment with random forest regression", In Proceedings of the 4th international conference on educational data mining", 2011, pp. 259–263.

[11] Nguyen Thai-Nghe, Tomas Horvath, Lars Schmidt-Thieme, "Personalized Forecasting Student Performance", In 11th IEEE International Conference on Advanced Learning Technologies, 2011, pp. 412-414.

[12] Goldin, I. M., Koedinger, K. R., &Aleven, V., "Learner differences in hint processing", In Proceedings of the 5th international conference on educational data mining, 2012, pp. 73–80.

[13] Wang, Y., & Beck, J. E., "Using Student Modeling to Estimate Student Knowledge Retention", In Proceedings of the 5th international conference on educational data mining, 2012, pp. 200–203.

[14] Yoo, J. S., & Cho, M. H., "Mining concept maps to understand university students' learning", In Proceedings of the 5th international conference on educational data mining, 2012, pp. 184–187.

[15] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) "Synthetic minority over-sampling technique", Journal of Artificial Intelligence Research 16, 2002, pp. 321– 357.

[16] Cano A, Zafra A, Ventura S., "An EP algorithm for learning highly interpretable classifiers", In: Proceedings of the 10th international conference on intelligent systems design and applications, ISDA, 2011, pp. 325–330.

[17] Konen, W., Koch, P., Flasch, O., Bartz-Beielstein, T., Friese, M. and Naujoks, B., "Tuned data mining: a benchmark study on different tuners", Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation, 1995–2002, 2011.

[18] Maimon, O., Rokach, L. and Edel, I., "Parameter tuning for classification algorithms in data mining using meta learning", 13th Israeli Conference of Industrial Engineering and Management, 2004.

[19] Brijesh Kumar Bhardwaj, Saurabh Pal, "Data Mining: A prediction for performance improvement using classification", In IJCSIS, Vol. 9, No. 4, April 2011.

[20] O. N. Pratiwi, "Predicting student placement class using data mining", In IEEE International conference on TALE, 26-29 Aug 2013, pp. 618-621.