Analysis of classification learning algorithms

Hana Rasheid Esmaeel
Department of Information & Communication, Al Nahrain University, Iraq

Article Info

Article history:

Received Apr 27, 2019
Revised Jul 7, 2019
Accepted Aug 8, 2019

Keywords:

Classification
Data mining
Decision tree
Teacher evaluation
Weka

ABSTRACT

The paper attempts to apply data mining technique, to estimate the teacher performance of college of Information Engineering (COIE) In Al Nahrain University in Baghdad/Iraq, Five classifications algorithms were used to build data they are (ZeroR, SMO, Naive Bayesian, J48 and Random Forest). The analysis implemented using WEKA (3. 8. 2) Data mining software tool. Information was collected from within the variety of form using "Referendum"; it was stored in Excel file CSV format then regenerate to ARFF (Attribute-Relation File Format). Many criteria like (Time is taken to create models, accuracy and average error) was taken to evaluate the algorithms. Random forest and SMOPredicts higher than alternative algorithms since their accuracy is the highest and have the lowest average error compared to others, "The teacher clarification, and wanting to be useful to students", was the strongest attribute. Further, removing the bad ranked attributes (10, 11, 12, and 14) that have a lower contact on the Dataset can increase accuracies of algorithms.

Corresponding Author:

Hana Rasheid Esmaeel,
Department of Information & Communication,
Al Nahrain University, Iraq.
Email: hanauom@yahoo.com

1. INTRODUCTION

"Extracting the knowledge from the massive set of information is called Data mining." [1], The data ought to be new, not obvious, and one should be ready to use it, that ought to be helpful when deciding to find the buried patterns and interaction that should be useful in deciding. Data mining consists of five elements [1].

a) Extract, transform and load dealing information.
b) Store and manage the info in exceedingly two-dimensional info systems.
c) Provide information access to business analysts and knowledge technology professionals.
d) Analyze the info by application code.
e) Reward the info in an exceedingly helpful format like a graph or table.

The aim of this paper is to evaluate the teacher performance in (COIE) in Al Nahrain University in Baghdad/Iraq within the Referendum to the students to estimate the teacher performance by using five classifications algorithms, and then discover which attribute has the strongest effect in teacher evaluation, and which algorithm was the best.

Several works have used data mining to enhance teacher performance; Asanbe M. O. Osofisan A.O. and William W. F [2] have designed Artificial Neural Network (ANN) and Decision Tree scheme their system Was tested using data from a "Nigerian University". Renuka Agrawal1, Jyoti Singh, and Zadgoankar [3] "suggests type to estimate the performance victimization data mining" like ("association," classification rules "Decision Tree", "Rule Induction," K-NN, and "Naïve Bayesian") to search out traditions to assist them to reinforce supply the educational process". Hemaid and El- Halees [4] in their study they use Questionnaire which has questions on the classes, they intend a form to "test teacher performance during data mining techniques like, classification, association rules to find out behavior to aid them toward enhanced the learning procedure and improve the presentation of teachers in classroom”, to enhance the educational process and expand the contribution of teachers in the classroom”.Works published by Ahmeda, Rizanerc and Ulusoyc [5] using The sequential Minimal Optimization, Naïve Bayes, J48 Decision Tree, and Multilayer Perception to Evaluate Student records to predict the teacher performance and investigates factors that have affected students achievements to develop the teaching system,. In [6], in his study to Predict students' performance he finds that the classification scheme is repeatedly used in educational data mining area, it includes, NeuralNetwork and Decision Tree, the two methods greatly used by the researchers for predicting students' performance. Ms. A.Pavithra, Mr. S. Dhanaraj [7] In their study they examine the prediction accurateness of the academic performance of teaching the students using different classification algorithms like, “MLP, Naïve Bayes, Decision tree, REP tree, and J48 tree”, they concluded that “many factors will influence the student performance, and it may differ to the different locality of students". Farid Jauhari, Ahmad Afif Supianto [8] proposes three boosting algorithms (C5. 0, AdaBoost. M1, and AdaBoost. SAMME) to build the classifier for predicting student's performance. They used three scenarios of evaluation, the first scenario employs 10-fold cross-validation to compare the performance of boosting algorithms. The second scenario was accustomed to evaluate boosting algorithms below the various varieties of coaching information within the third scenario, they build models from one subject Dataset, and test using another subject Dataset. They conclude that the third scenario results indicate that they can build a prediction model using one subject to predict another. Bin Mat and N. Buniyamin [9] using neuro-fuzzy tool to classify and predict electrical engineering students graduation achievement based on mathematics competency. It's supported longitudinal progress and cross-validation model on two arithmetic subjects, semesters’ performance, and graduation achievement of electrical students. They conclude that the mixture of statistical associate analysis and machine learning will facilitate to extract data, and alter university management to assist low achievers at an early stage. They hoped that their findings can help faculty management to review mathematics curriculum with respect in the increasing range of engineering field. S. Hussain, N. A. Dahan, F. M. Ba-Alwib, and N. Ribata [10] used for classifications methods, (J48, PART, Random Forest and Bayes Network Classifiers). The high influential attributes were selected using the data mining tool Weka. They conclude that the Random Forest Classification method was the most suited algorithm for the Dataset.

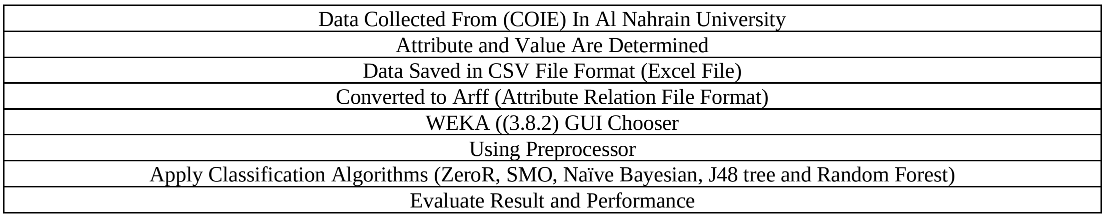
Saouabi Mohamed, Abdullah Ezzati [11] proposes a data mining process for employability data using classification techniques (the Decision Tree classifier, Logistic regression, and Naïve Bayes algorithms), apply them by Rapid Miner Studio Educational Version (8. 1. 000), using employability Dataset. They conclude that the Decision tree classifier is more accurate than Logistic regression and Naïve Bay.

2. RESEARCH METHOD

2.1. Prepare Data

For this analysis, data were collected from college students at totally different Departments within the faculty of College of Information Engineering (COIE) at Al Nahrain University in Baghdad/Iraq for the aim of the investigation; however their skilled improvement has taken Place throughout the term. The info was collected from the college within the type of form to judge teacher performance as shown in Table 1.

a) Preparing Information, Teacher's Data was Evaluated. Attributes and value were determined.
b) Data saved in excel file in CSV (Comma Separated Values).
c) To use Weka data must be converted to Arrf (Attribute-Relation File Format)
d) Using Weka (3.8.2) GUI Chooser and Explorer.
e) Using Wekapreprocessor and open Arff file.
f) Apply classification algorithms (ZeroR, SMO, Naïve Bayesian, J48 tree and Random Forest).
g) Evaluate the result and performance. Figure 1 shows the proposed system layout.



Data Collected From (COIE) In Al Nahrain University
Attribute and Value Are Determined
Data Saved in CSV File Format (Excel File)
Converted to Arff (Attribute Relation File Format)
WEKA ((3.8.2) GUI Chooser
Using Preprocessor
Apply Classification Algorithms (ZeroR, SMO, Naïve Bayesian, J48 tree and Random Forest)
Evaluate Result and Performance

Figure 1. The roposed system

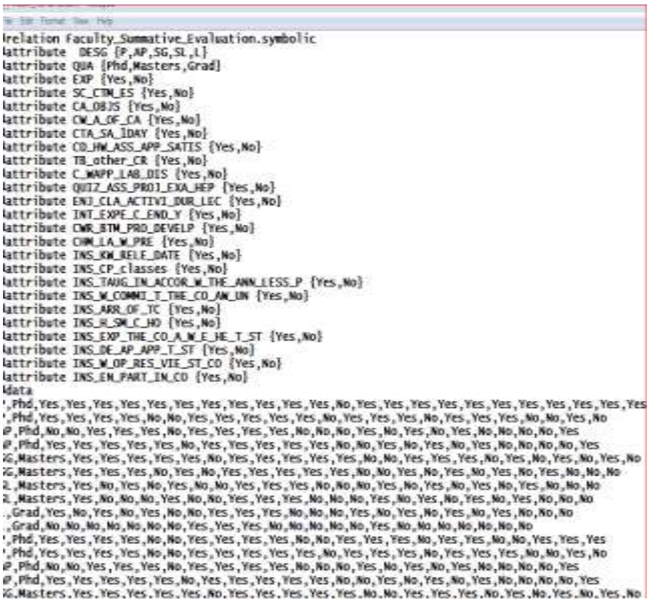
Table 1. The Questions and Their Abbreviations

Seq	Attribute	Description
1	DESG.	Title
2	QUA.	Degree
3	EXP.	Experience
4	SC_CTM_ES	The semester course content pedagogic and analysis were provided at the beginning
5	CA_OBJ.S	Atthe beginning the teacher were clearlyspecificthe aims and purpose of the course.
6	CW_A_CA	Theamountof creditallotted tothe course waspositively significance
7	CTA_SA_AY	The course was instructed in step with the information proclaimed on the primary day of sophistication
8	CD_HW_ASS._APP._SAT	The class discussions homework assignment applications and studies were satisfactory
9	TB_other_CR	The text book and different courses resources were enough and up to this point
10	C_WAPP._LAB_DIS	The course acceptable conversation of the laboratory applications and different studies.
11	QUIZ.ASS.PROJ.EXA._HEP.	The quizzes assignment comes and exams contributed to serving to the educational.
12	ENJ.CLA.ACTIVI.DUR.LEC.	The lecture allows students to participate their knowledge.
13	INT.EXPE.C.END.Y	The course were met all student prospect
14	CWR_BTM_PRO._DEVE	The course was relevant and helpful to my skilled development.
15	CHM_LA_W_PRE.	The course helped Maine check up on life and world with my new perspective.
16	INS._KW_RELE._DATE	The lecturer information was relevant and up to this point.
17	INS._CP_clas	The lecturer came ready for categories
18	INS._TAUG.IN_ACCOR._	The lecturer instructed in accordance with the proclaimed lesson set up.
19	INS._W_COMMI._T_THE._CO.	The lecturer was committed to the course and was comprehendible
20	INS._ARR._OF_T.C	The lecturer arrived of your time for categories
21	INS._H.SM._C_HO.	The lecturer had a sleek associate of sophistication hours
22	INS._EXP._THE_CO._A.W.E._HE._T_	The teacher clarification and was wanting to be useful to students
23	INS._DE_AP_APP._T_ST.	The lecturer incontestable appositive approach to students
24	INS._W.OP._RES._VIE._ST._CO.	The teacher was considerate of the views of student on the topic of the course.
25	INS._EN._PART._IN_CO.	The lecturer inspired participation within the course.

2.2. Data Collected

The knowledge was gathering for making ready the model, the fields that are needed for data processing was taken, this includes pre-processing or extracts vital info from it then produce correct format file of the info like inweka.arff file format (Attribute Relation File Format) [3], as shown in Figure 2.

@relation *teacher evaluation*
@attribute *Name1* string
@attribute *Name2* numeric
@attribute Namen?
@data
Yes, 1, yes….



```
@relation Faculty_Summative_Evaluation.symbolic
@attribute  DESG {P,AP,SG,SL,L}
@attribute QUA {Phd,Masters,Grad}
@attribute EXP {Yes,No}
@attribute SC_CTM_ES {Yes,No}
@attribute CA_OBJS {Yes,No}
@attribute CW_A_OF_CA {Yes,No}
@attribute CTA_SA_1DAY {Yes,No}
@attribute CD_HW_ASS_APP_SATIS {Yes,No}
@attribute TB_other_CR {Yes,No}
@attribute C_WAPP_LAB_DIS {Yes,No}
@attribute QUIZ_ASS_PROJ_EXA_HEP {Yes,No}
@attribute ENJ_CLA_ACTIVI_DUR_LEC {Yes,No}
@attribute INT_EXPE_C_END_Y {Yes,No}
@attribute CWR_BTM_PRO_DEVELP {Yes,No}
@attribute CHM_LA_W_PRE {Yes,No}
@attribute INS_KW_RELE_DATE {Yes,No}
@attribute INS_CP_classes {Yes,No}
@attribute INS_TAUG_IN_ACCOR_W_THE_ANN_LESS_P {Yes,No}
@attribute INS_W_COMMI_T_THE_CO_AN_UN {Yes,No}
@attribute INS_ARR_OF_TC {Yes,No}
@attribute INS_H_SM_C_HO {Yes,No}
@attribute INS_EXP_THE_CO_A_W_E_HE_T_ST {Yes,No}
@attribute INS_DE_AP_APP_T_ST {Yes,No}
@attribute INS_W_OP_RES_VIE_ST_CO {Yes,No}
@attribute INS_EN_PART_IN_CO {Yes,No}
@data
P,Phd,Yes,Yes,Yes,Yes,Yes,Yes,Yes,Yes,Yes,Yes,Yes,No,Yes,Yes,Yes,Yes,Yes,Yes,Yes,Yes,Yes,Yes,Yes
P,Phd,Yes,Yes,Yes,Yes,No,No,Yes,Yes,Yes,Yes,Yes,No,Yes,Yes,Yes,No,Yes,Yes,Yes,No,No,Yes,No
AP,Phd,No,No,Yes,Yes,Yes,No,Yes,Yes,Yes,Yes,No,No,No,Yes,No,Yes,No,Yes,No,No,No,No,Yes
AP,Phd,Yes,Yes,Yes,Yes,Yes,No,Yes,Yes,Yes,Yes,Yes,No,No,Yes,No,Yes,No,Yes,No,No,No,No,Yes
SG,Masters,Yes,Yes,Yes,Yes,Yes,No,Yes,Yes,Yes,Yes,Yes,No,No,Yes,Yes,Yes,No,Yes,No,Yes,No,Yes,No
SG,Masters,Yes,Yes,Yes,No,Yes,No,Yes,Yes,Yes,Yes,Yes,No,No,Yes,No,Yes,No,Yes,No,Yes,No,No,No
SL,Masters,Yes,No,Yes,No,Yes,No,No,Yes,Yes,Yes,No,No,No,Yes,No,Yes,No,Yes,No,Yes,No,No,No
SL,Masters,Yes,No,No,No,Yes,No,No,Yes,Yes,Yes,No,No,No,Yes,No,Yes,No,Yes,No,Yes,No,No,No
L,Grad,Yes,No,Yes,No,Yes,No,No,Yes,Yes,Yes,No,No,No,Yes,No,Yes,No,Yes,No,Yes,No,No,No
L,Grad,No,No,No,No,No,No,No,Yes,Yes,Yes,No,No,No,No,No,Yes,No,No,No,No,No,No,No
P,Phd,Yes,Yes,Yes,Yes,No,No,Yes,Yes,Yes,Yes,No,No,Yes,Yes,Yes,No,Yes,Yes,No,No,Yes,Yes,Yes
P,Phd,Yes,Yes,Yes,Yes,No,No,Yes,Yes,Yes,Yes,Yes,No,Yes,Yes,Yes,No,Yes,Yes,Yes,No,No,Yes,No
AP,Phd,No,No,Yes,Yes,Yes,No,Yes,Yes,Yes,Yes,No,No,No,Yes,No,Yes,No,Yes,No,No,No,No,Yes
AP,Phd,Yes,Yes,Yes,Yes,Yes,No,Yes,Yes,Yes,Yes,Yes,No,No,Yes,No,Yes,No,Yes,No,No,No,No,Yes
SG,Masters,Yes,Yes,Yes,Yes,Yes,No,Yes,Yes,Yes,Yes,Yes,No,No,Yes,Yes,Yes,No,Yes,No,Yes,No,Yes,No
```

Figure 2. Tacher.arff file

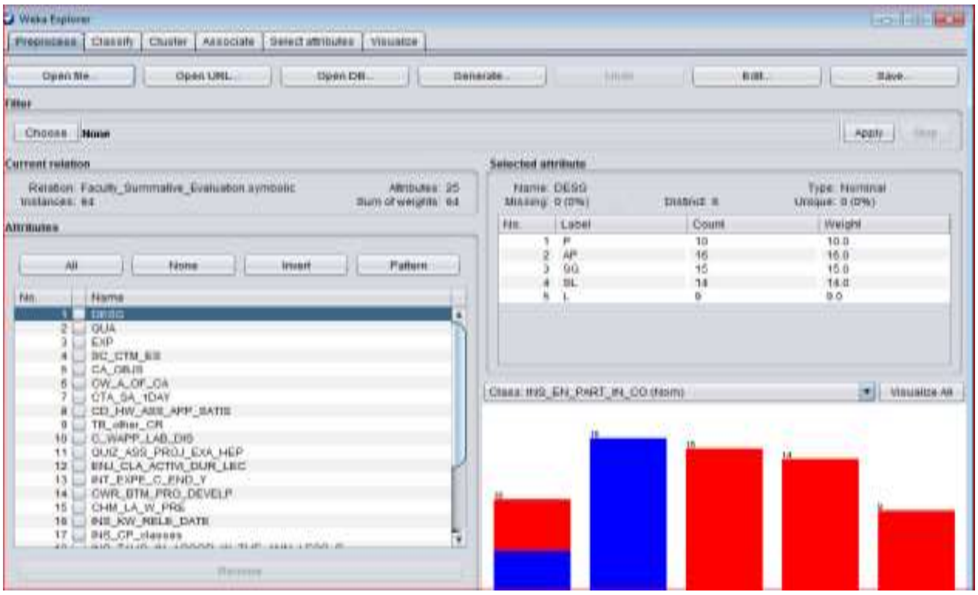
2.3. Methodology

The classification technique was used for the forecast of teachers' evaluation. Five classifications algorithms were used (ZeroR, SMO, Naïve Bayesian, J48 tree and Random Forest) and implemented using Weka (3. 8.2) Data mining software tool.

2.4. Building Models

2.4.1. Building the Trivial model ZeroR [12, 13]

a) In Preprocess panel, click "Open file" button, choose the file named (teacher.arff) as in Figure 3.
b) Select ZeroR by Clicking "Choose" Button.
c) Invoke classifier by clicking "start" button to make a model. The analytical performance of the model characterized by the right-hand classifier output frame.
d) The Confusion Matrix for the model is bestowed at the underside part of the Classifier output window. It is seen from it that compounds are classified as (21) affirmative and (43) No.
e) The accuracy of the model is (67.187) for No and (32.813) for affirmative as in Figure 4 and Figure 5.



Filter: None

Current relation

Relation: Faculty_Summative_Evaluation.symbolic Attributes: 25
Instances: 64 Sum of weights: 64

Attributes

No.	Name
1	DESG
2	QUA
3	EXP
4	SC_CTM_ES
5	CA_OBJS
6	CW_A_OF_CA
7	CTA_SA_1DAY
8	CD_HW_ASS_APP_SATIS
9	TB_other_CR
10	C_WAPP_LAB_DIS
11	QUIZ_ASS_PROJ_EXA_HEP
12	ENJ_CLA_ACTIVI_DUR_LEC
13	INT_EXPE_C_END_Y
14	CWR_BTM_PRO_DEVELP
15	CHM_LA_W_PRE
16	INS_KW_RELE_DATE
17	INS_CP_classes

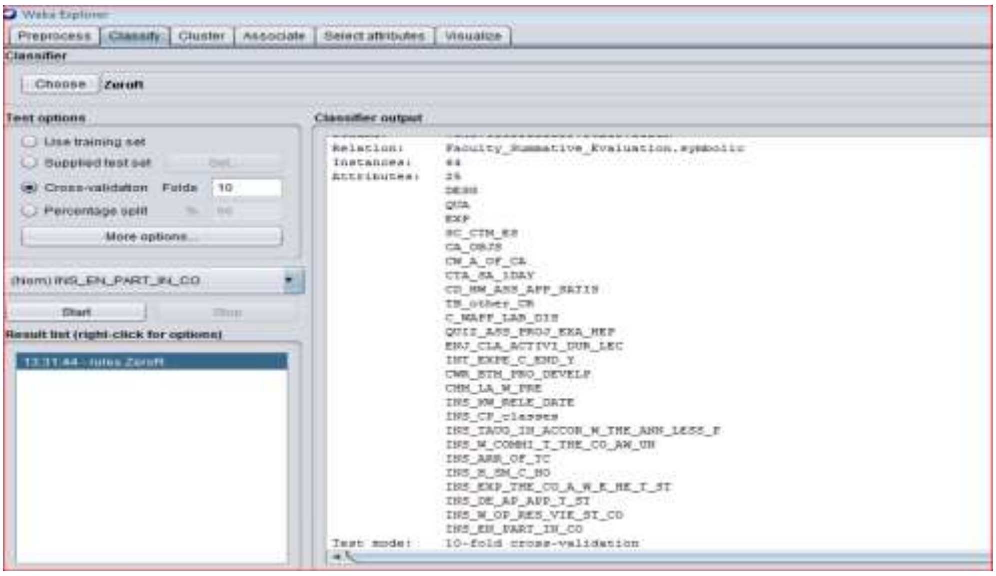
Selected attribute

Name: DESG Type: Nominal
Missing: 0 (0%) Distinct: 5 Unique: 0 (0%)

No.	Label	Count	Weight
1	P	10	10.0
2	AP	16	16.0
3	SG	15	15.0
4	SL	14	14.0
5	L	9	9.0

Class: INS_EN_PART_IN_CO (Nom)

Figure 3. Selected attributes of teacher.arff



Classifier: ZeroR

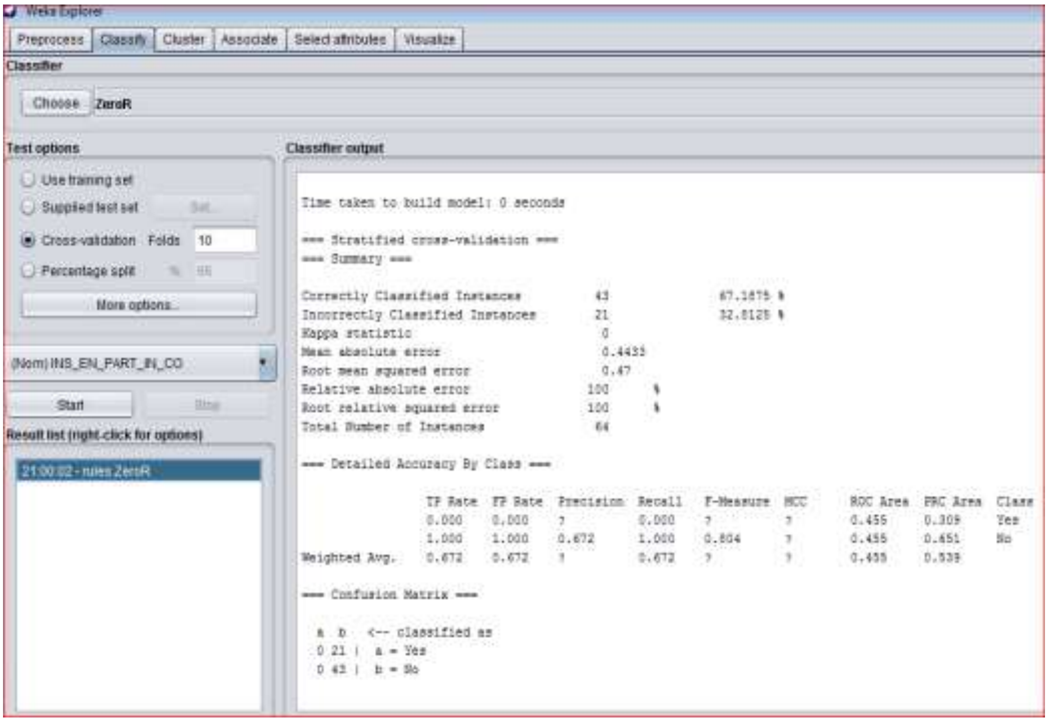
Test options
○ Use training set
○ Supplied test set Set...
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) INS_EN_PART_IN_CO

Result list: 13:31:44 - rules.ZeroR

```
Relation:     Faculty_Summative_Evaluation.symbolic
Instances:    64
Attributes:   25
              DESG
              QUA
              EXP
              SC_CTM_ES
              CA_OBJS
              CW_A_OF_CA
              CTA_SA_1DAY
              CD_HW_ASS_APP_SATIS
              TB_other_CR
              C_WAPP_LAB_DIS
              QUIZ_ASS_PROJ_EXA_HEP
              ENJ_CLA_ACTIVI_DUR_LEC
              INT_EXPE_C_END_Y
              CWR_BTM_PRO_DEVELP
              CHM_LA_W_PRE
              INS_KW_RELE_DATE
              INS_CP_classes
              INS_TAUG_IN_ACCOR_W_THE_ANN_LESS_P
              INS_W_COMMI_T_THE_CO_AW_UN
              INS_ARR_OF_TC
              INS_R_SM_C_HO
              INS_EXP_THE_CO_A_W_E_HE_T_ST
              INS_DE_AP_APP_T_ST
              INS_W_OP_RES_VIE_ST_CO
              INS_EN_PART_IN_CO
Test mode:    10-fold cross-validation
```

Figure 4. ZeroR classifier output



Choose ZeroR

Cross-validation Folds 10

(Nom) INS_EN_PART_IN_CO

21:00:02 - rules.ZeroR

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          43               67.1875 %
Incorrectly Classified Instances        21               32.8125 %
Kappa statistic                          0
Mean absolute error                      0.4433
Root mean squared error                  0.47
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances               64

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.000    0.000    ?          0.000    ?          ?        0.455     0.309     Yes
                 1.000    1.000    0.672      1.000    0.804      ?        0.455     0.651     No
Weighted Avg.    0.672    0.672    ?          0.672    ?          ?        0.455     0.539

=== Confusion Matrix ===

  a  b   <-- classified as
  0 21 |  a = Yes
  0 43 |  b = No
```

Figure 5. Continue of ZeroR classifier output

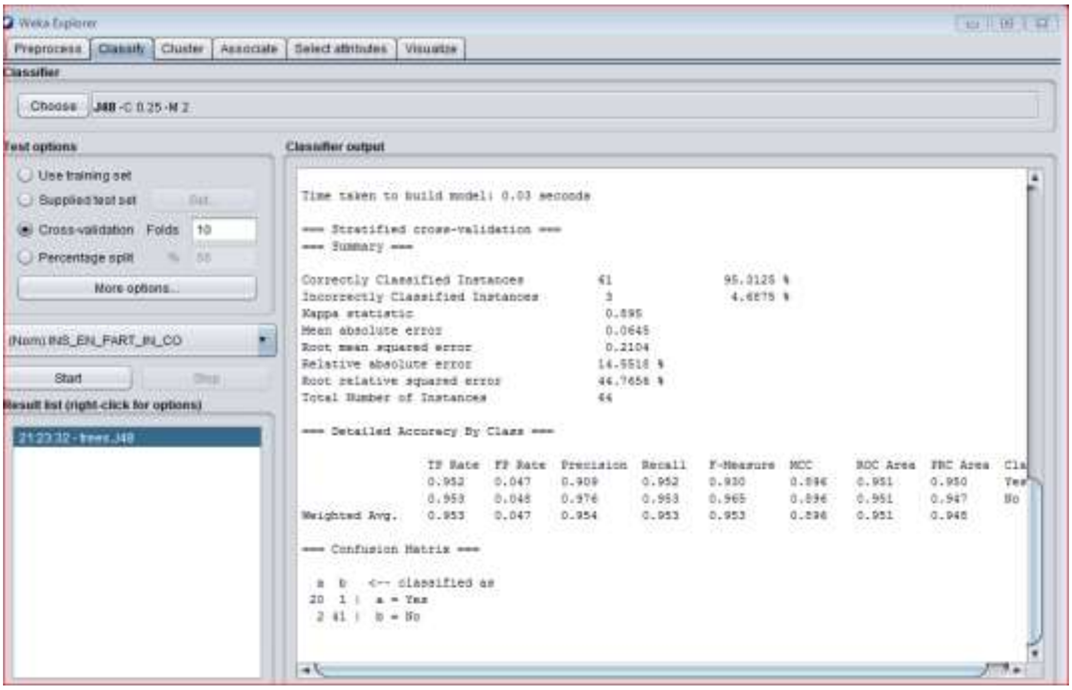
2.4.2. Building J48 Tree Model [12]

To get the usual illustration of the tree, the following must be done:

a) Clickthe correct push on the model type trees (J48) within the Result list frame and choose

b) The menu item visualizes tree size a replacement window with graphical Illustration of thetree.

c) Click with the correct push to the area during this screen, and within the popup menu choose the item appropriate screen. As in Figures 6 and 7.



Choose J48 -C 0.25 -M 2

Cross-validation Folds 10

(Nom) INS_EN_PART_IN_CO

21:23:32 - trees.J48

```
Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          61               95.3125 %
Incorrectly Classified Instances         3                4.6875 %
Kappa statistic                          0.895
Mean absolute error                      0.0645
Root mean squared error                  0.2104
Relative absolute error                 14.5518 %
Root relative squared error             44.7658 %
Total Number of Instances               64

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Cla
                 0.952    0.047    0.909      0.952    0.930      0.896    0.951     0.950     Yes
                 0.953    0.048    0.976      0.953    0.965      0.896    0.951     0.947     No
Weighted Avg.    0.953    0.047    0.954      0.953    0.953      0.896    0.951     0.948

=== Confusion Matrix ===

  a  b   <-- classified as
 20  1 |  a = Yes
  2 41 |  b = No
```

Figure 6. The statistical of the J48 model

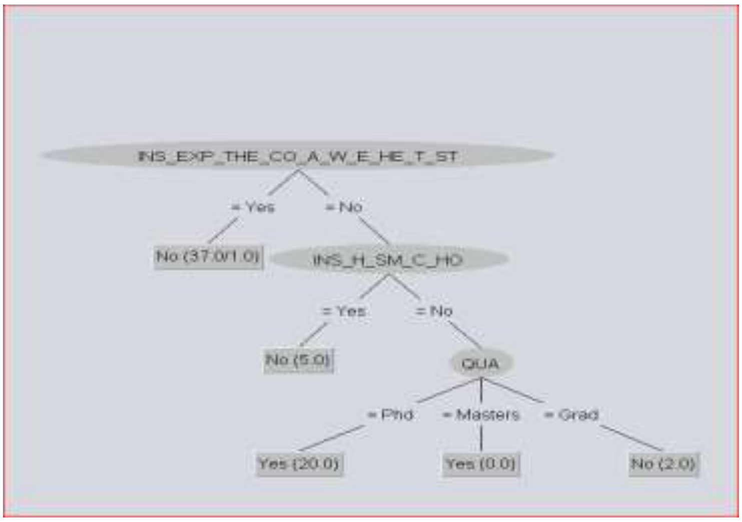
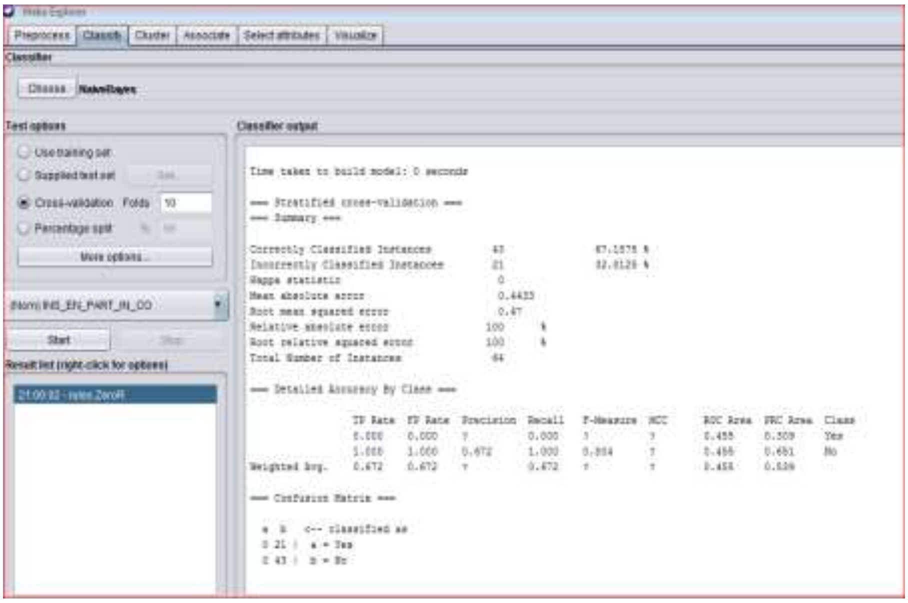


Figure 7. Visualize j48 tree

2.4.3. Building Naive Bayesian Model [12, 13]

As in previous models Naive Bayesian Model was built as shown in Figure 8.



NaiveBayes

○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) INS_EN_PART_IN_CO

21:00:02 - rules.ZeroR

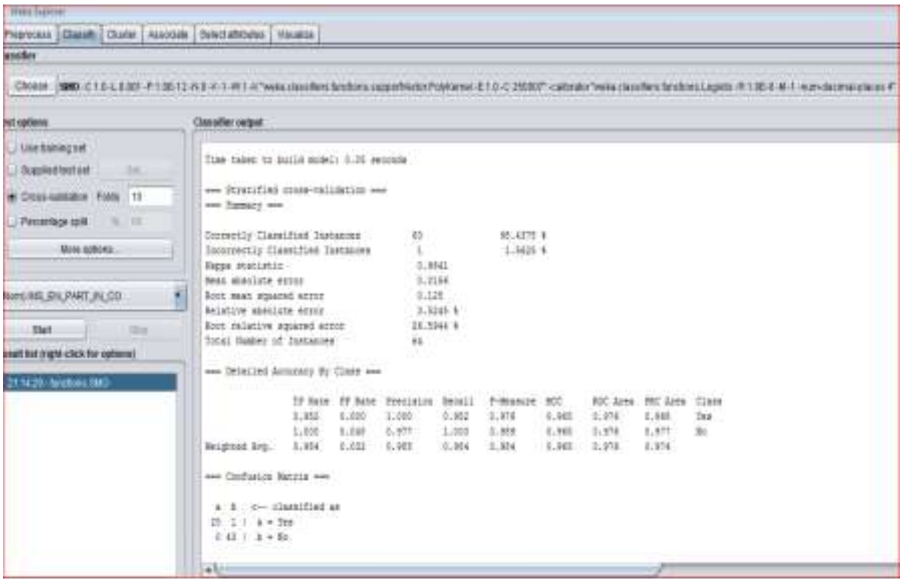
```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          43               67.1875 %
Incorrectly Classified Instances        21               32.8125 %
Kappa statistic                          0
Mean absolute error                      0.4433
Root mean squared error                  0.47
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances               64

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.000    0.000    ?          0.000    ?          ?        0.455     0.309     Yes
                 1.000    1.000    0.672      1.000    0.804      ?        0.455     0.651     No
Weighted Avg.    0.672    0.672    ?          0.672    ?          ?        0.455     0.539

=== Confusion Matrix ===

  a  b   <-- classified as
  0 21 |  a = Yes
  0 43 |  b = No
```

Figure 8. Navie bayesian model

2.4.4. Building Support Vector Machine Models [12-15]

The Weka software implements Sequential Minimal Optimization (SMO) algorithm for training a support vector classifier. Figure 9 show the classifier output for this model.

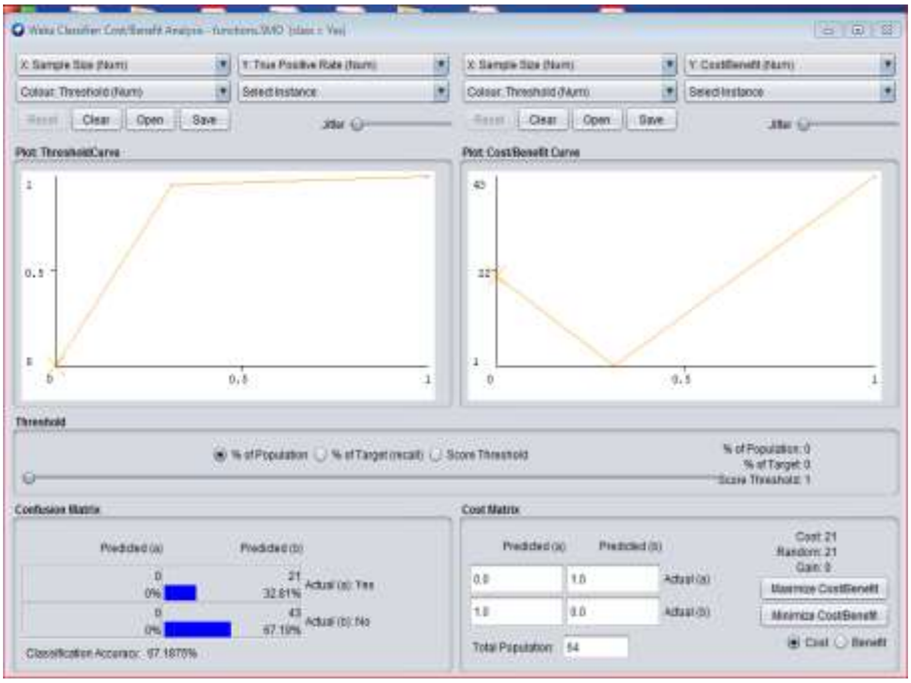
The accuracy of (Correctly Classifieds Instances) of this model is extremely high ninety eight.4375%. This truth clearly indicates that the accuracy can't be used for assessing the utility of classification models designed exploitation unbalanced datasets. For this purpose an honest selection is to use the Kappa statistic, that is =0.964 for this case. Kappa statistic is the academic degree analog of the constant of correlation. It's worth is zero for the shortage of any relation and approaches to (1) for terribly sturdy applied math relation between the category label and attributes of instances, Another helpful applied math characteristic is “ROC Area”, that the worth =0.976 means that sensible mythical monster curves may be build and therefore the cost/benefit analysis will simply be performed. As in Figure 10.



SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4"

```
Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          63               98.4375 %
Incorrectly Classified Instances         1                1.5625 %
Kappa statistic                          0.9641
Mean absolute error                      0.0156
Root mean squared error                  0.125
Relative absolute error                  3.5245 %
Root relative squared error             26.5946 %
Total Number of Instances               64

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.952    0.000    1.000      0.952    0.976      0.965    0.976     0.968     Yes
                 1.000    0.048    0.977      1.000    0.989      0.965    0.976     0.977     No
Weighted Avg.    0.984    0.032    0.985      0.984    0.984      0.965    0.976     0.974

=== Confusion Matrix ===

  a  b   <-- classified as
 20  1 |  a = Yes
  0 43 |  b = No
```

Figure 9. The classifier output for SMO



X: Sample Size (Num) | Y: True Positive Rate (Num) | X: Sample Size (Num) | Y: Cost/Benefit (Num)
Colour: Threshold (Num) | Select instance | Colour: Threshold (Num) | Select instance

● % of Population ○ % of Target (recall) ○ Score Threshold

% of Population: 0
% of Target: 0
Score Threshold: 1

Predicted (a)	Predicted (b)	
0 0%	21 32.81%	Actual (a): Yes
0 0%	43 67.19%	Actual (b): No

Classification Accuracy: 67.1875%

Predicted (a)	Predicted (b)	
0.0	1.0	Actual (a)
1.0	0.0	Actual (b)

Total Population: 64

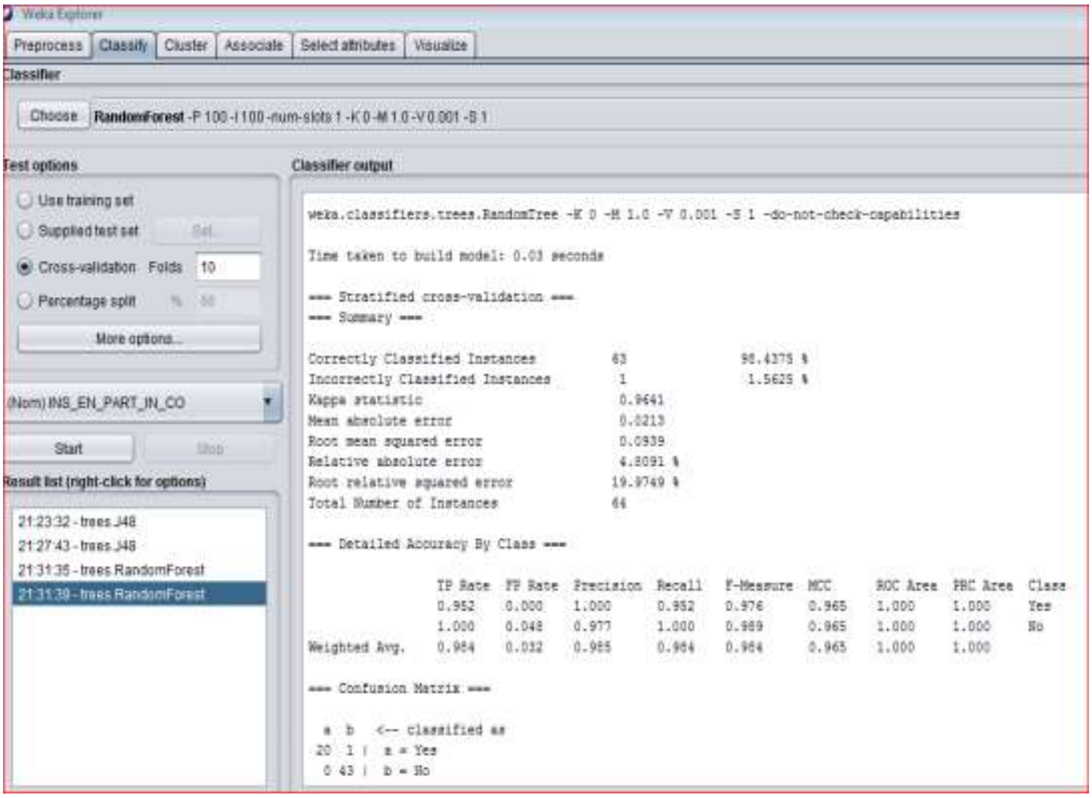
Cost: 21
Random: 21
Gain: 0

● Cost ○ Benefit

Figure 10. Cost/benefit analysis for SMO

2.4.5. Building the Random Forest Models [12-15]

The classifier output used for "Random Forest" algorithm be shown in Figure 11 below the accuracy of this model is extremely high= ninety eight.4375%. This reality clearly indicates the accuracy cannot be used to assessing the worth of classification models designed victimization unbalanced datasets. For this purpose an honest alternative is to use the “Kappa statistic”, that is =0.964 for this case [6]. Its price is =0.9641 it's terribly robust applied math relation between the category label and attributes of instances, Another helpful applied math characteristic is “ROC Area”, that the worth =1.000 means that sensible mythical creature curves are built and therefore the cost/benefit analysis will simply be performed. As in Figure 12.



Classifier

RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options

○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) INS_EN_PART_IN_CO

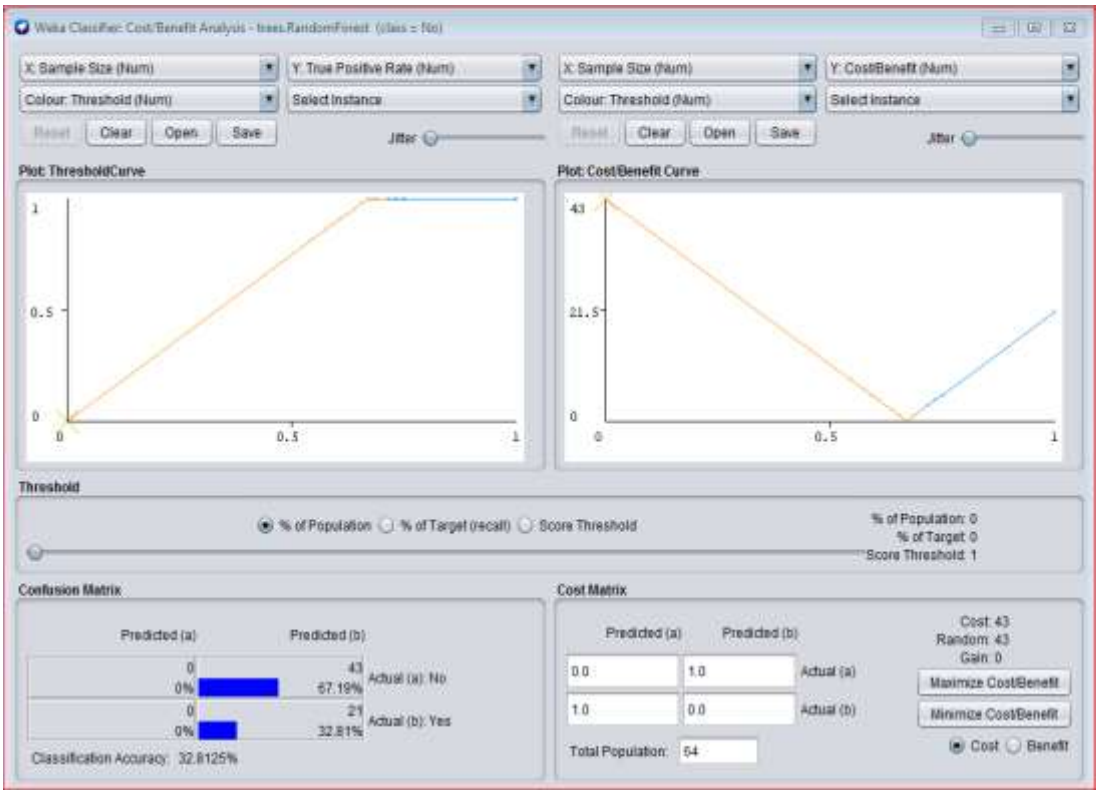
Result list (right-click for options)

21:23:32 - trees.J48
21:27:43 - trees.J48
21:31:35 - trees.RandomForest
21:31:39 - trees.RandomForest

Classifier output

```
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          63               98.4375 %
Incorrectly Classified Instances         1                1.5625 %
Kappa statistic                          0.9641
Mean absolute error                      0.0213
Root mean squared error                  0.0939
Relative absolute error                  4.8091 %
Root relative squared error             19.9749 %
Total Number of Instances               64

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.952    0.000    1.000      0.952    0.976      0.965    1.000     1.000     Yes
                 1.000    0.048    0.977      1.000    0.989      0.965    1.000     1.000     No
Weighted Avg.    0.984    0.032    0.985      0.984    0.984      0.965    1.000     1.000

=== Confusion Matrix ===

  a  b   <-- classified as
 20  1 |  a = Yes
  0 43 |  b = No
```

Figure 11. Classifier Output for Random Forest



X: Sample Size (Num) | Y: True Positive Rate (Num) | Colour: Threshold (Num) | Select instance

X: Sample Size (Num) | Y: CostBenefit (Num) | Colour: Threshold (Num) | Select instance

Threshold

● % of Population ○ % of Target (recall) ○ Score Threshold

% of Population: 0
% of Target: 0
Score Threshold: 1

Confusion Matrix

Predicted (a)	Predicted (b)	
0 0%	43 67.19%	Actual (a): No
0 0%	21 32.81%	Actual (b): Yes

Classification Accuracy: 32.8125%

Cost Matrix

Predicted (a)	Predicted (b)	
0.0	1.0	Actual (a)
1.0	0.0	Actual (b)

Total Population: 64

Cost: 43
Random: 43
Gain: 0

● Cost ○ Benefit

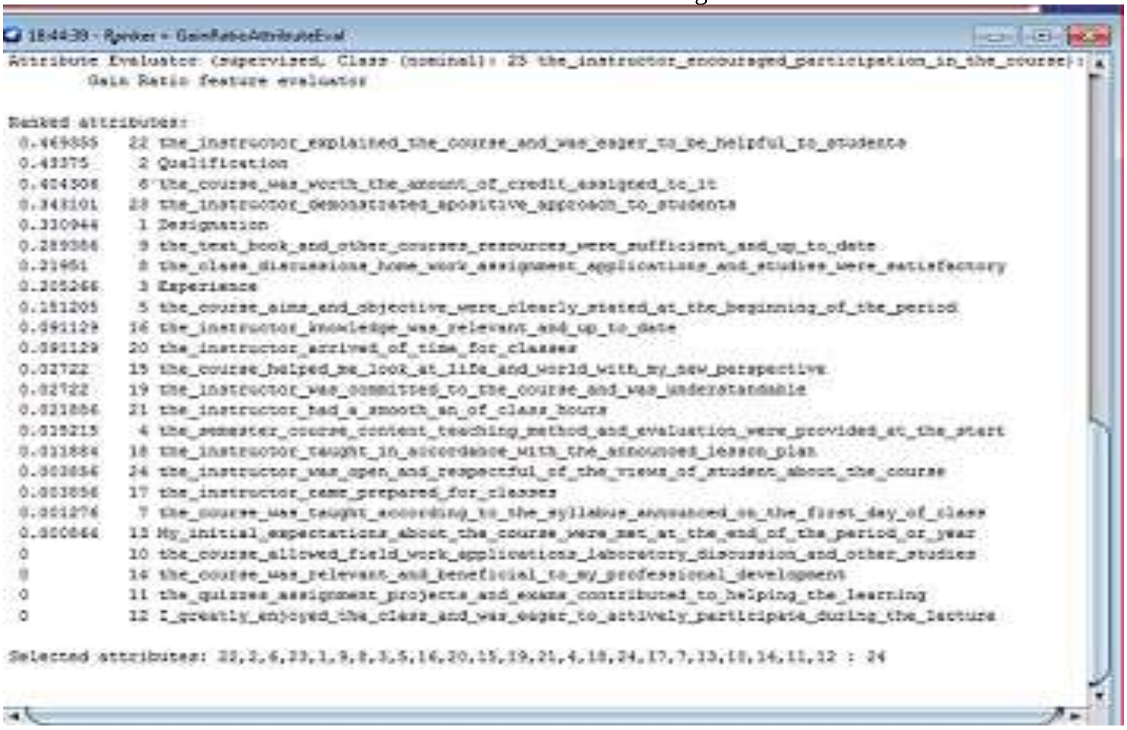
Figure 12. Cost/benefit Random Forest

3. RESULTS AND DISCUSSION

3.1. Attribute Ranking [16-20]

Weka explorer can evaluate the attributes of the data by using the following steps:
Select attributes→GainRatoAttribute→start→showEval→Rank Attribute
The list of attributes and their value are appearing from higher to lower, in Table 2 show that.

Table 2. Attribute Ranking



```
Attribute Evaluator (supervised, Class (nominal): 25 the_instructor_encouraged_participation_in_the_course):
        Gain Ratio feature evaluator

Ranked attributes:
 0.469855    22 the_instructor_explained_the_course_and_was_eager_to_be_helpful_to_students
 0.43375      2 Qualification
 0.424306     6 the_course_was_worth_the_amount_of_credit_assigned_to_it
 0.343101    23 the_instructor_demonstrated_apositive_approach_to_students
 0.330944     1 Designation
 0.289386     9 the_text_book_and_other_courses_resources_were_sufficient_and_up_to_date
 0.21461      8 the_class_discussions_home_work_assignment_applications_and_studies_were_satisfactory
 0.205266     3 Experience
 0.151205     5 the_course_aims_and_objective_were_clearly_stated_at_the_beginning_of_the_period
 0.091129    16 the_instructor_knowledge_was_relevant_and_up_to_date
 0.091129    20 the_instructor_arrived_of_time_for_classes
 0.02722     15 the_course_helped_me_look_at_life_and_world_with_my_new_perspective
 0.02722     19 the_instructor_was_committed_to_the_course_and_was_understandable
 0.021886    21 the_instructor_had_a_smooth_an_of_class_hours
 0.015215     4 the_semester_course_content_teaching_method_and_evaluation_were_provided_at_the_start
 0.011884    18 the_instructor_taught_in_accordance_with_the_announced_lesson_plan
 0.003856    24 the_instructor_was_open_and_respectful_of_the_views_of_student_about_the_course
 0.003856    17 the_instructor_came_prepared_for_classes
 0.001276     7 the_course_was_taught_according_to_the_syllabus_announced_on_the_first_day_of_class
 0.000666    13 My_initial_expectations_about_the_course_were_met_at_the_end_of_the_period_or_year
 0           10 the_course_allowed_field_work_applications_laboratory_discussion_and_other_studies
 0           14 the_course_was_relevant_and_beneficial_to_my_professional_development
 0           11 the_quizzes_assignment_projects_and_exams_contributed_to_helping_the_learning
 0           12 I_greatly_enjoyed_the_class_and_was_eager_to_actively_participate_during_the_lecture

Selected attributes: 22,2,6,23,1,9,8,3,5,16,20,15,19,21,4,18,24,17,7,13,10,14,11,12 : 24
```

3.2. Models Comparison [21, 22]

The performances of the 5 models were evaluated primarily based on the standards as illustrated In Table 3.

a) Prediction accuracy: The share of properly classified instances is usually referred to as accuracy of a model.
b) Time is taken to create the model.
c) Error rate.

Table 3. Comparison Analysis on The Models

Metric	ZeroR	J48	NaïveBayesian	SMO	RandomForest
Time To Build The Model	0	0.03	0	0.05	0.03
Correctly Classified Instances	67.187%	95.312%	92.187%	98.437%	98.435%
In Correctly Classified Instances	32.812%	4.687%	7.812%	1.562%	1.562%
Kappa Statistics	0	0.895	0.833	0.964	0.964
Mean Absolute Error	0.443	0.0645	0.0791	0.0156	0.0213
Root Mean Square Error	0.47	0.210	0.279	0.125	0.0939
Relative Absolute Error	100%	14.55 %	17.838%	3.524%	4.809 %
Root Relative Square Error	100%	44.765%	59.428%	26.594%	

3.3. Performance of the Models

Table 4 show the performance of the (5) algorithms: **[11, 23-25]**

TP=true positives": a variety of examples": Predicted positive that are literally positive.

FP=false positives": a variety of examples: "Expected positive that are literally negative.

TN=true negatives": a variety of Examples ": predicted negative that are literally negative.

FN=false negatives": a variety of Examples: "Expected negative that are literally positive.

Weka (3.8.2) Confusion Matrix: The quantity of properly classified instances is that the total of diagonals within the matrix; all others area unit incorrectly classified.

x y<-- classified as, actual x=0 TP FN
Actual y=1 FN TP
TP=TP+FN / Recall
Precision=TP/TP+FP
Accuracy=TP+TN /TP+TN+FP+FN

Table 4. Performance of The Model

Algorithm	TP Rate	FP Rate	Precision	Recall	F-Measure	Roc Area
ZeroR	0.672	0.672	?	0.672	?	0.455
J48 Tree	0.953	0.047	0.954	0.953	0.953	0.951
Naive Bayesian	0.922	0.038	0.937	0.922	0.924	0.927
SMO	0.984	0.032	0.985	0.984	0.984	0.976
RandomForest	0.984	0.032	0.985	0.984	0.984	1.000

4. CONCLUSION

From the result of comparison of the five algorithms as in Tables 4 and 5 it conclude that Algorithms SMO and Random forest predicts higher than alternative algorithms since their accuracy is that the highest and have lowest average error compared to others algorithms on functioning on performance, several attributes are tested, and found that a few of them are effective on the performance prediction. "The teacher clarification and was wanted to be useful to students", was the strongest attribute and then the result plays a vital role within the performance of academics. More a lot of removing the worst hierarchal attributes (10, 11, 12, and 14), that have a lower impact on the dataset can increase the algorithms performance accuracies.

REFERENCES

[1] Shafiq Aslam,Imran Ashraf, “Data Mining Algorithms and Their Applications in Education Data Mining”, *International Journal of Advance Research in Computer Science and Management Studies*, Vol 2, Issue 7, pg. 50-56, July 2014.

[2] Asanbe M.O., Osofisan A.O., William W.F. “Teachers’ Performance Evaluation in Higher Educational Institution using Data Mining Technique”, *International Journal of Applied Information Systems (IJAIS)* – ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 10 – No.7, March 2016.

[3] Renuka Agrawal1, Jyoti Singh, A.S. Zadgoankar, “Summative Assessment for Performance Evaluation of a Faculty Using Data Mining Techniques,”*International Journal of Advanced Research in Computer and Communication Engineering* ISO 3297:2007 Certified Vol. 5, Issue 10, October 2016.

[4] Randa Kh. Hemaid1, Alaa M. El-Halees, “Improving Teacher Performance using Data Mining”, *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 4, Issue 2, February 2015.

[5] Ahmed Mohamed Ahmeda, Ahmet Rizanerc, Ali Hakan Ulusoyc, *“Using Data Mining to Predict Instructor Performance”,* 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS, Vienna, Austria.29-30 August 2016.

[6] A. Mohamed Shahiria,W. Husaina , N. Abdul Rashida, *“A Review on Predicting Student’s Performance using Data Mining Techniques”* Procedia Computer Science 72 ,414 – 422, ELSEVIER 2015. Available online at www.sciencedirect.com

[7] Ms.A.Pavithra, Mr.S.Dhanaraj, “Prediction Accuracy on Academic Performance of Students Using Different Data Mining Algorithms with Influencing Factors”, *IJSRCSAMS* Vol 7, Issue 5, 2018.

[8] Farid Jauhari, Ahmad Afif Supianto , “Building Student’s Performance Decision Tree Classifier Using Boosting Algorithm”, *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*Vol. 14, No. 3, pp. 1298-1304, June 2019.

[9] U. Bin Mat and N. Buniyamin, “Using Neuro-Fuzzy Technique to Classify and Predict Electrical Engineering StudentsAchievement Upon Graduation Based On Mathematics Competency,” *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*., vol. 5, no. 3, pp. 684–690, 2017.

[10] S. Hussain, N. A. Dahan, F. M. Ba-Alwib, and N. Ribata, “Educational data mining and analysis of students academic performance using WEKA,” *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*., vol. 9, no. 2, pp. 447–459, 2018.

[11] Saouabi Mohamed, Abdullah Ezzati “A data mining process using classification techniques for employability prediction”, *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)* Vol. 14, No. 2, pp. 1025-1029, May 2019,

[12] Tutorialspoint.com, *“Data Mining Tutorial Simply Easy Learning,”* 2-11-2014

[13] The WEKA Workbench Eibe Frank, Mark A. Hall, and Ian H. Witten Online Appendix for. "Data Mining: Practical Machine Learning Tools and Techniques,"*Morgan Kaufmann*, Fourth Edition, 2016

[14] Miss. Sharayu N. Bonde , Dr. D. K. Kirange, *"Survey on Evaluation of Student's Performance in Educational Data Mining,"*Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018) IEEE Xplore Compliant - Part Number: CFP18BAC-ART; ISBN:978-1-5386-1974. 2018.

[15] Anil Kumar Tiwari, G. Ramakrishna, Lokesh Kumar Sharma, Sunil Kumar Kashyap. "Academic performance prediction algorithm based on fuzzy data mining," *International Journal of Artificial Intelegence (IJ-AI)* Vol. 8, No. 1,:pp 26 – 32, March 2019.

[16] Weka – "*Data Mining Machine Learning Software*". Available at: http://www.cs.waikato.ac.nz/ml/weka/, (Accessed 20 April 2014).

[17] Yadav, S.K., Bhardwaj B.K., and Pal, S. ",Data Mining Applications: A Comparative Study for Predicting Student's Performance", *International Journal of Innovative Technology and Creative Engineering (IJITCE),*2011, pp. 13-19.

[18] Asma Lamani, Brahim Erraha, Malika Elkyal, Abdallah Sair, "Data mining techniques application for prediction in OLAP cube", *International Journal of Electrical and Computer Engineering (IJECE)* Vol. 9, No. 3, June 2019, pp. 2094~2102

[19] Abdulsalam Sulaiman O., Babatunde Akinbowale N.,& Babatunde Ronke S.," Comparative Analysis of Decision Tree Algorithms for Predicting Undergraduate Students' Performance in Computer Programming", *A Multidisciplinary Journal Publication of the Faculty of Science, Adeleke University*, Ede, Nigeria,2015 Vol 2.

[20] Adillah Dayana Ahmad Dali, Nurul Aswa Omar, Aida Mustapha,"Data Mining Approach to Herbs Classification", *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, Vol. 12, No. 2, November 2018 : 570 – 576

[21] Kandasamy P, Balamurali., "Performance analysis of classifier models to predict diabetes mellitus*", Journal of Elesevier,* vol 47, pp.45-51, 2014.

[22] Parneet Kaura, Manpreet Singhb ,Gurpreet Singh Josanc *"Classification and Prediction based Data Mining Algorithms to Predict Slow Learners in Education Sector",* Science Direct Procedia Computer Science 57; 500-508,2015 (ICRTC- 2015).

[23] Hlaudi Daniel Masethe, Mosima Anna Masethe., *"Prediction of heart disease using classification algorithms",* Proceedings of the world congress on engineering and computer science, San Francisco, USA, pp. 185-191, 2014.

[24] Aqliima Aziz, Cik Feresa Mohd Foozy, Palaniappan Shamala, Zurinah Suradi ," YouTube Spam Comment Detection Using Support Vector Machine and K–Nearest Neighbor", *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, Vol. 12, No. 2, November 2018 : 607 – 611

[25] Ratna Patil, et al., "A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Diabetes", *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 5, pp. 3966-3975, 2018.