Prediction and Analysis of Student Performance by Data Mining in WEKA

Report of Project submitted for the partial fulfillment of the requirements for the degree of **Bachelor of Technology**

In

Information Technology

Submitted by

AGNIK DEY

REGISTRATION NO – 141170110101
UNIVERSITY ROLL NO – 11700214006

ABHIRUP KHASNABIS

REGISTRATION NO – 141170110097
UNIVERSITY ROLL NO -11700214002

AJEET KUMAR

REGISTRATION NO – 141170110104
UNIVERSITY ROLL NO - 11700214009

Under the Guidance of **Mr. Sudarsan Biswas**

RCC Institute of Information Technology

Canal South Road, Beliaghata, Kolkata – 700015

[Affiliated to West Bengal University of Technology]

Acknowledgement

We would like to express our sincere gratitude to Mr. Sudarsan Biswas of the department of Information Technology, whose role as project guide was invaluable for the project. We are extremely thankful for the keen interest he took in advising us, for the books and reference materials provided for the moral support extended to us.

Last but not the least we convey our gratitude to all the teachers for providing us the technical skill that will always remain as our asset and to all non-teaching staff for the gracious hospitality they offered us.

Place: RCCIIT, Kolkata

Date:

……………..

AGNIK DEY

REGISTRATION NO – 141170110101
UNIVERSITY ROLL NO – 11700214006
B. TECH (IT) – 8TH SEMESTER, 2018

....

ABHIRUP KHASNABIS
REGISTRATION NO – 141170110097
UNIVERSITY ROLL NO -11700214002
B. TECH (IT) – 8TH SEMESTER, 2018

…………………………………………………….

AJEET KUMAR

REGISTRATION NO – 141170110104

UNIVERSITY ROLL NO - 11700214009
B. TECH (IT) – 8TH SEMESTER, 2018

RCC Institute of Information Technology

Certificate

This is to certify that the project report titled "Prediction and Analysis of student performance by Data Mining in WEKA" prepared under my supervision by Agnik Dey (Roll No.: 11700214006), Abhirup Khasnabis (Roll No.: 11700214002), Ajeet Kumar (Roll No.: 11700214009) of B. Tech. (IT) 8th Semester of 2018, be accepted in partial fulfillment for the degree of Bachelor of Technology in Information Technology.

It is to be understood that by this approval, the undersigned does not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn thereof, but approves the report only for the purpose for which it has been submitted.

....	....
Dr. Abhijit Das, Associate Professor & Head	Mr. Sudarsan Biswas, Assistant Professor

RCC Institute of Information Technology

Certificate of Acceptance

The report of the Project titled [Prediction and Analysis of student performance by Data Mining in WEKA] submitted by Agnik Dey (Roll No.: 11700214006), Abhirup Khasnabis (Roll No.: 11700214002), Ajeet Kumar (Roll No.: 11700214009) of B. Tech. (IT) 8th Semester of 2018 is hereby recommended to be accepted for the partial fulfillment of the requirements for B Tech (IT) degree in West Bengal University of Technology

Name of the Examiner	**Signature with Date**
1. ……………………………………….	…………………………………
2. ………………………………………..	…………………………………..
3. ………………………………………	……………………………………
4. ……………………………………….	………………………………….

TABLE OF CONTENTS

Abstract

Over the years, several statistical tools have been used to analyze and predict students' performance from different point of view. One of the biggest challenges for higher education
Today is to predict the paths of students through the educational process. Successful students' result prediction in early course stage depends on many factors. Data mining techniques could be used for this kind of job. Data mining techniques are widely used in educational field to find new hidden patterns from student's data. The hidden patterns that are discovered can be used to understand the problem arise in the educational field. Data Mining (DM), or Knowledge Discovery in Databases (KDD), is an approach to discover useful information from large amount of data. Data mining techniques apply various methods in order to discover and extract patterns from stored data Based on collected students' information, different data mining techniques need to be used. For the purpose of this project WEKA data mining software is used for the prediction of final student mark based on parameters in the given dataset. The dataset contains information about different students from one college course in the past semester. Student data from the last semester are used for test dataset.

1. Introduction

Nowadays, data mining is playing a vital role in educational institutions and one of the most important areas of research with the objective of finding meaningful information from the data stored in huge dataset. Educational data mining (EDM) is a very important research area which helpful to predict useful information from educational database to improve educational performance, better understanding and to have better assessment of the students learning process. Data Mining or knowledge discovery has become the area of growing significance because it helps in analyzing data from different perspectives and summarizing it into useful information.

- **What is Data Mining?**

Data Mining is defined as extracting information from huge sets of data. In other words, we can say that data mining is the procedure of mining knowledge from data.

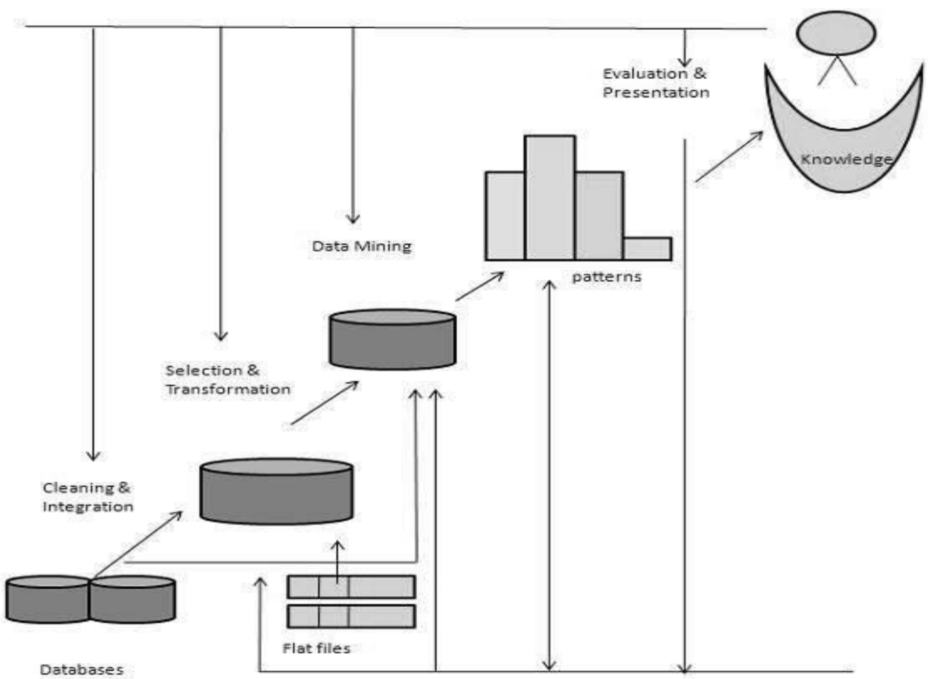
Data Mining could be a promising and flourishing frontier in analysis of data and additionally the result of analysis has many applications. Data Mining can also be referred as Knowledge Discovery from Data (KDD).This system functions as the machine-driven or convenient extraction of patterns representing knowledge implicitly keep or captured in huge databases, data warehouses, the Web, data repositories, and information streams. Data Mining is a multidisciplinary field, encompassing areas like information technology, machine learning, statistics, pattern recognition, data retrieval, neural networks, information based systems, artificial intelligence and data visualization.

The application of data mining is widely prevalent in education system. Educational data mining is an emerging field which can be effectively applied in the field of education. The educational data mining uses several ideas and concepts such as Association rule mining, classification and clustering. The knowledge that emerges can be used to better understand students' promotion rate, students' retention rate, students' transition rate and the students' success. The data mining system is pivotal and crucial to measure the students' performance improvement. The classification algorithms can be used to classify and analyze the students' data set in accurate manner. The students' academic performance is influenced by various factors like pa rents' education, locality, economic status, attendance, gender and result.

The main objective of the project is to use data mining methodologies to study and analyze the school students' performance. Data mining provides many tasks that could be used to study the students' performance. In this paper, the classification task is employed to gauge students' performance and deals with the accuracy, confusion matrices and the execution time taken by the various classification data mining algorithms

- **What is Knowledge Discovery Database (KDD)?**

Knowledge discovery in databases (KDD) is the process of discovering useful knowledge from a collection of data. This widely used data mining technique is a process that includes data preparation and selection, data cleansing, incorporating prior knowledge on data sets and interpreting accurate solutions from the observed results.
Here is a basic outline of KDD



1. 1 Application

Our project is on Educational Data Mining (EDM) field. It has several applications. The areas of EDM are-

- Analysis and visualization of data
- Providing feedback for supporting instructors
- Recommendations for students
- Predicting student performance
- Student modeling
- Detecting undesirable student behaviors
- Grouping students
- Social network analysis
- Developing concept maps
- Planning and scheduling

1. 2. Motivation

In India, there is largest no. of educational institutes, so it is second largest in the world after United States. There is more competition between all institutes for attracting students to get enrollment in their institutes so they focus on strength of students not quality of education at the time of enrollment. Today Admission process of institutes has become very critical. There are many problems at the time of admission in institutes because many students apply for courses but seats are limited, so there is no proper seat allocation of courses to the students so students are unable to get enroll in their interested courses. Some students have good marks but they get admission in other course (that is not according to their subjects) due to limited seats.

So there is a proper attention is needed in admission process. Every year huge amount of student data is recorded in database however this data is not put in proper form. There is a requirement of data mining that handle these challenges & overcome them. Then there is enough information for better planning, evaluation and decision making. Data mining will extract hidden information from student enrollment database, this information will be meaningful for institutes. Then a better & mined knowledge is present in database that can be use directly, there is no extra requirement. The motive behind in this paper is based on classification model for enrollment in higher educational courses using data mining techniques. This is useful for predicting the students that are interested to take admission in higher study course. By this study we will find some meaningful pattern that can be useful for institutes.

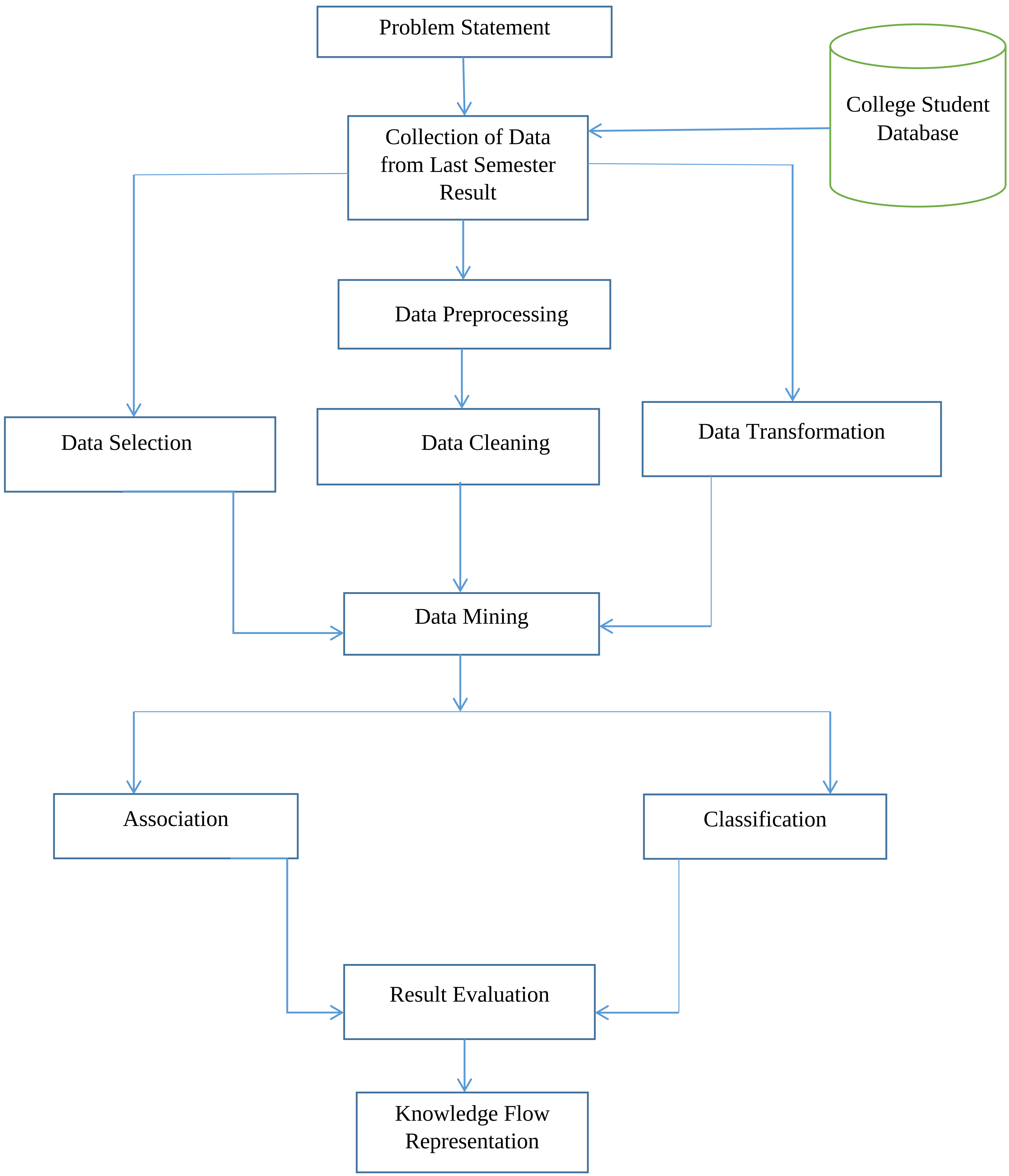
1. 3 Problem Definition

Data mining is widely used in educational field to find the problems arise in this field. Student performance is of great concern in the educational institutes where several factors may affect the performance. For prediction the three required components are: Parameters which affect the student performance, Data mining methods and third one is data mining tool. These Parameters may be psychological, personal, and environmental. We conduct this study to maintain the education quality of institute by minimizing the diverse affect of these factors on student's performance. In this Paper, Prediction of student Performance is done by applying Apriori classification techniques WEKA tool. By applying data mining techniques on student data we can obtain knowledge which describes the student performance. This knowledge will help to improve the education quality, student's performance and to decrease failure rate. All these will help to improve the quality of institute.

1. 4 Planning

The main objective of this work is to use data mining methodologies to student's performance in the semester. Data mining provides many tasks that could be used to study the student performance. Our work will be divided into two main parts- one is prediction by classification and another one is association rule mining by using the machine learning tool 'WEKA'. At first we will select our dataset and then perform preprocessing of it. After preprocess we will do classification over the dataset and perform prediction of result. Then we will apply association rule mining technique over the dataset and generate some rules which will be analyzed later. At last both result of prediction and association will be visualized by 'Knowledge Flow Representation'.

1.4.1 Work Flow Diagram



2. Background

Here is the background tool required for our project.

Required Software (WEKA) –

We have used a data mining software named as WEKA for this project. For the purposes of this study, we select WEKA (Waikato Environment for Knowledge Analysis) software that was developed at the University of Waikato in New Zealand. WEKA tool supports to a wider range of algorithms & very large data sets. The WEKA (pronounced Waykuh) workbench contains a collection of visualization tools & algorithms. WEKA is open source software issued under the GNU General Public License. It contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. The original non-java version of WEKA was a TCL/TK, but the recent java based version is WEKA 3(1997), is now used in many different application areas, in particular for education & research. WEKA's main user interface is Explorer. The Experimenter is also there by which we can compare WEKA's machine learning algorithms' performance. The Explorer interface has many panels by which we can access to main components of workbench. The Visualization tab allows visualizing a 2-D plot of the current working relation, it is very useful. In this study WEKA toolkit 3.8.1 is used for generating the association rules and prediction of result.

WEKA supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection. All of WEKA's techniques are predicated on the assumption that the data is available as a single flat file or relation, where each data point is described by a fixed number of attributes (normally, numeric or nominal attributes, but some other attribute types are also supported). WEKA provides access to SQL databases using Java Database Connectivity and can process the result returned by a database query. It is not capable of multi-relational data mining, but there is separate software for converting a collection of linked database tables into a single table that is suitable for processing using WEKA.

3. Literature Survey

Samrat Singh, Dr. Vikesh Kumar [1] .Data Mining is a powerful tool for academic performance. Educational Data Mining is concerned with developing new methods to discover knowledge from educational database and can used for decision making in educational system.

M. Goyal and R. Vohra [2] .Data analysis plays an important role for decision support irrespective of type of industry like any manufacturing unit and educations system. If data mining techniques such as clustering, decision tree and association are applied to higher education processes, it would help to improve students performance, their life cycle management, selection of courses, to measure their retention rate and the grant fund management of an institution.

Jason Brownlee [3]. After you have found a well performing machine learning model and tuned it, you must finalize your model so that you can make predictions on new data.

Neelam Naik & Seema Purohit [4] . The quality higher education is required for growth and development of country. Professional education is one of the pillars of higher education. Data mining techniques aim to discover hidden knowledge in existing educational data, predict future trends and use it for betterment of higher educational institutes as well as students.

Alaa M.El-Halees, Mohammed M. Abu Tair. [5] Educational data mining concerns with developing methods for discovering knowledge from data that come from educational domain. In this paper we used educational data mining to improve graduate students' performance, and overcome the problem of low grades of graduate students.

B.K. Bharadwaj and S. Pal [6] .Now-a-days the amount of data stored in educational database increasing rapidly. These databases contain hidden information for improvement of students' performance. The performance in higher education in India is a turning point in the academics for all students. This academic performance is influenced by many factors, therefore it is essential to develop predictive data mining model for students' performance so as to identify the difference between high learners and slow learners student. In the present investigation, an experimental methodology was adopted to generate a database.

Suchita Borkar, K. Rajeswari [7] .Education Data Mining is a promising discipline which has an imperative impact on predicting students' academic performance. In this paper, student's performance is evaluated using association rule mining algorithm. Research has been done on assessing student's performance based on various attributes. In our study important rules are generated to measure the correlation among various attributes which will help to improve the student's academic performance.

Randhir Singh, M.Tiwari, Neeraj Vimal [8]. Educational institutions are important parts of our society and playing a vital role for growth and development of nation and prediction of student's performance in educational environments is also important as well. Student's academic performance is based upon various factors like personal, social, psychological etc.

D.Magdalene Delighta Angeline [9].The objective of the educational institution that is producing good results in their academic exams can be achieved by using the data mining techniques which can be applied to predict the performance of the students and to impart the quality of education in the educational institutions. Data mining is used to extract meaningful information and to develop relationships among variables stored in large data set.

Mrs. M.S. Mythili, Dr. A.R.Mohamed Shanavas [10] . In recent years, the analysis and evaluation of students‟ performance and retaining the standard of education is a very important problem in all the educational institutions. The most important goal of the paper is to analyze and evaluate the school students‟ performance by applying data mining classification algorithms in WEKA tool.

S. Anupama Kumar and Dr. Vijayalakshmi M.N [11] .Educational data mining is used to study the data available in the educational field and bring out the hidden knowledge from it. Classification methods like decision trees, rule mining, Bayesian network etc can be applied on the educational data for predicting the students behavior, performance in examination etc

4. Design and Implementation

The followings are the step by step process of our project evaluation.

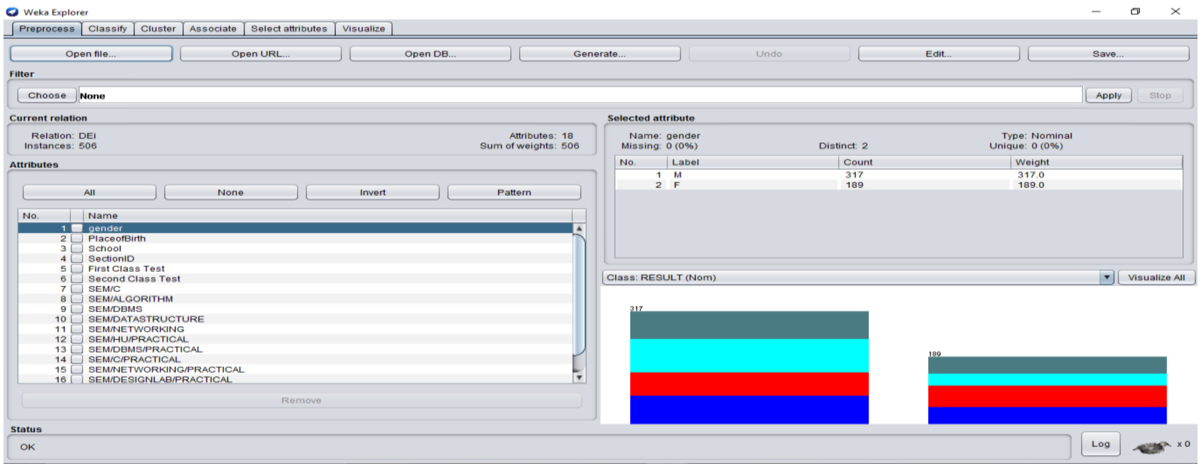
- **<u>Dataset and attribute selection</u>-** We have collected a dummy dataset contains the result of students of last semester. The dataset contains 507 instances and 18 attributes. It has some missing values also. The data file has to be in either in ‘CSV’ format or ‘ARFF’ format.

  Here is the sample of our dataset which is in ‘CSV’ format.

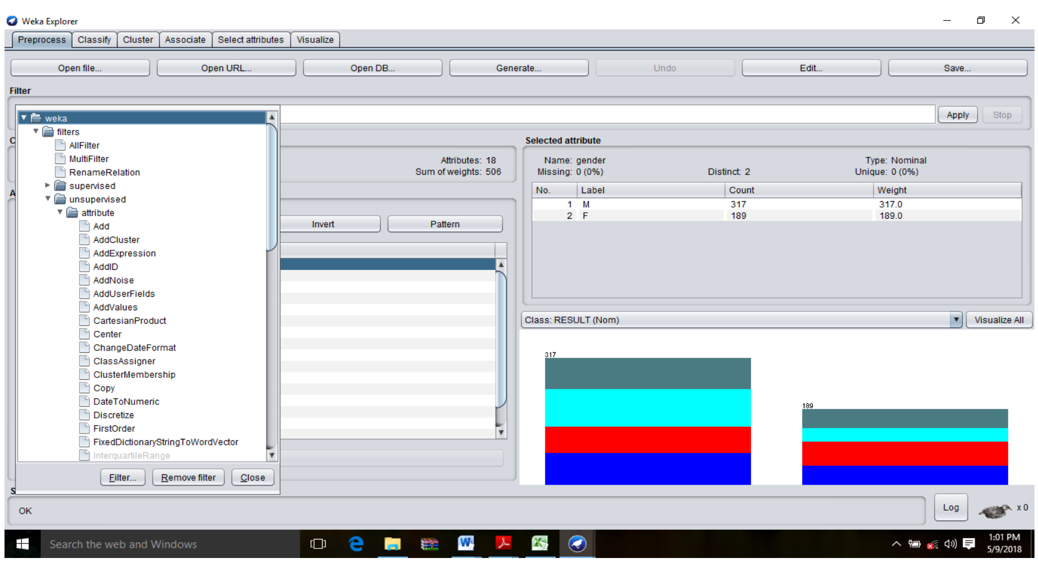
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	gender	PlaceofBi	School	SectionID	First Class	Second Cl	SEM/C	SEM/ALG(	SEM/DBM	SEM/DAT/	SEM/NET\	SEM/HU/F	SEM/DBM	SEM/C/PR	SEM/NET\	SEM/DESI	SEM/GRAI	RESULT	
2	M	Bengal	RCCIIT	A	Absence	Present	0-25	0-25	0-25	0-25	0-25		0-25	0-25	0-25	0-25	0-250	FAIL	
3	M	Bengal	RCCIIT	A	Absence	Present	26-50	26-50	26-50	26-50	26-50	26-50	26-50	26-50	26-50	26-50		GOOD	
4	M	Bengal	RCCIIT	A	Absence	Present	51-75	51-75	51-75	51-75	51-75	51-75	51-75	51-75	51-75	51-75	501-750	EXCELLENT	
5	M	Bengal	RCCIIT	A	Absence	Present						76-100	76-100	76-100	76-100	76-100	751-1000	OUTSTANDING	
6	M	Bengal	RCCIIT	A	Absence	Present						0-25	0-25	0-25	0-25	0-25	0-250	FAIL	
7	F	Bengal	RCCIIT	A	Absence	Present						26-50	26-50			26-50		GOOD	
8	M	Bengal	WBUT	A	Absence	Present						51-75	51-75			51-75	501-750	EXCELLENT	
9	M	Bengal	WBUT	A	Absence	Present						76-100	76-100			76-100	751-1000	OUTSTANDING	
10	F	Bengal	WBUT	A	Absence	Present	0-25	0-25	0-25	0-25	0-25	0-25	0-25	0-25	0-25	0-25	0-250	FAIL	
11	F	Bengal	WBUT	B	Absence	Present	26-50	26-50	26-50	26-50	26-50	26-50	26-50	26-50	26-50	26-50		GOOD	
12	M	Bengal	WBUT	A	Absence	Present	51-75			51-75	51-75	51-75	51-75	51-75	51-75	51-75	501-750	EXCELLENT	
13	M	Bengal	WBUT	B	Absence	Present	76-100			76-100	76-100	76-100	76-100	76-100	76-100	76-100	751-1000	OUTSTANDING	
14	M	Bengal	RCCIIT	A	Absence	Present	0-25			0-25	0-25	0-25	0-25	0-25	0-25	0-25	0-250	FAIL	
15	M	Bengal	WBUT	A	Absence	Present	26-50	26-50	26-50	26-50	26-50	26-50	26-50	26-50	26-50	26-50	251-500	GOOD	
16	F	Bengal	WBUT	A	Absence	Absence	51-75	51-75	51-75	51-75	51-75	51-75	51-75	51-75	51-75	51-75		EXCELLENT	
17	F	Bengal	WBUT	A	Absence	Present	76-100	76-100	76-100	76-100	76-100					76-100	751-1000	OUTSTANDING	
18	M	Bengal	WBUT	B	Absence	Present	0-25	0-25	0-25	0-25	0-25					0-25	0-250	FAIL	
19	M	Bengal	WBUT	A	Absence	Present	26-50	26-50	26-50	26-50	26-50					26-50		GOOD	
20	F	Bengal	WBUT	A	Absence	Absence	51-75	51-75	51-75	51-75	51-75	51-75	51-75	51-75	51-75	51-75	501-750	EXCELLENT	
21	M	Bengal	WBUT	B	Absence	Absence	76-100	76-100	76-100	76-100	76-100	76-100	76-100	76-100	76-100	76-100	751-1000	OUTSTANDING	
22	F	Bengal	WBUT	A	Absence	Present	0-25	0-25	0-25	0-25	0-25	0-25	0-25	0-25	0-25	0-25	0-250	FAIL	
23	F	Bengal	WBUT	B	Absence	Present	26-50	26-50	26-50	26-50	26-50	26-50	26-50	26-50	26-50	26-50		GOOD	
24	M	Bengal	WBUT	A	Absence	Present	51-75				51-75	51-75	51-75	51-75	51-75	51-75	501-750	EXCELLENT	
25	M	Bengal	WBUT	A	Absence	Present	76-100				76-100	76-100	76-100	76-100	76-100	76-100	751-1000	OUTSTANDING	

- **<u>Preprocessing</u>-**

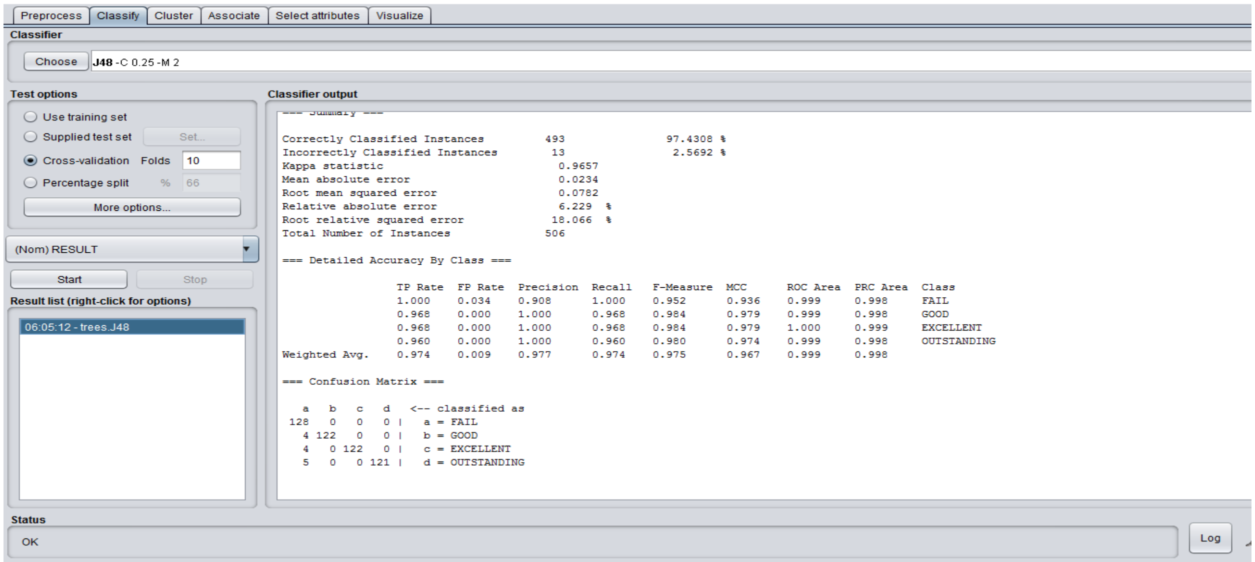
  Data Preprocessing is the first step of evaluation of this project. For our project we will choose WEKA Explorer interface. Here the source data file is selected from local machine. After loading the data in Explorer, we can refine the data by selecting different options which is known as ‘Data Cleaning’ and can also select or remove attributes as per our need. The following is the preprocessed of our dataset. Left hand side of the above screen shows detail of relation name, number of attributes and number of records. Right hand side gives details of attribute values, type, and number of distinct values. Specification of every attribute is displayed in the right bottom of the screen.



- **Filters -**The preprocess section allows filters to be defined that transform the data in various ways. The Filter box is used to set up the filters that are required. There are mainly two categories of filters-Supervised and Unsupervised. Here we will choose unsupervised category filters. In case if the dataset is contained with any numeric values we have to covert it nominal values(as Association in WEKA can only support nominal values) by using 'Numeric To Nominal' filter under attribute section of Unsupervised filters. Another one filter we will apply named as 'Replace Missing Values' which will replace all missing values of our dataset and will make the dataset able to perform 'Approximate Association Rule Generation' about which we will talk later on this paper.



- **<u>Classification -</u>** To predict nominal or numeric quantities we have classifiers in WEKA. For our prediction purpose we have to choose a classifier. We will select a standard classifier named as J48 for classification.



From the above example we can say J48 is a good classifier as it gives an accuracy of 97.43% because the percentage of correctly classified instances is often called accuracy or sample accuracy. The correctly and incorrectly classified instances show the percentage of test instances that were correctly and incorrectly classified. The raw numbers are shown in the confusion matrix, with a,b,c and d representing the class labels.

Here are some others factor in classifier output-

- **TP Rate**: rate of true positives (instances correctly classified as a given class)
- **FP Rate**: rate of false positives (instances falsely classified as a given class)
- **Precision**: proportion of instances that are truly of a class divided by the total instances classified as that class
- **Recall**: proportion of instances classified as a given class divided by the actual total in that class (equivalent to TP rate)
- **F-Measure**: A combined measure for precision and recall calculated as 2 * Precision * Recall / (Precision + Recall)

<u>Prediction of result</u> –First, the file with cases to predict needs to have the same structure that the file used to learn the model. The difference is that the value of the result attribute is "?" for all instances (question marks represent missing values in WEKA).

- **Train Dataset in ARFF**

```
@relation DEi
@attribute gender {M,F}
@attribute PlaceofBirth {Bengal,Mumbai,Delhi,Bihar,Pune}
@attribute School {RCCIIT,WBUT,IEM}
@attribute SectionID {A,B,C}
@attribute 'First Class Test' {Absence,Present}
@attribute 'Second Class Test' {Present,Absence}
@attribute SEM/C {0-25,26-50,51-75,76-100}
@attribute SEM/ALGORITHM {0-25,26-50,51-75,76-100}
@attribute SEM/DBMS {0-25,26-50,51-75,76-100,57-67}
@attribute SEM/DATASTRUCTURE {0-25,26-50,51-75,76-100,25-67}
@attribute SEM/NETWORKING {0-25,26-50,51-75,76-100}
@attribute SEM/HU/PRACTICAL {26-50,51-75,76-100,0-25}
@attribute SEM/DBMS/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/C/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/NETWORKING/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/DESIGNLAB/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/GRANDTOTAL {0-250,501-750,751-1000,251-500}
@attribute RESULT {FAIL,GOOD,EXCELLENT,OUTSTANDING}

@data
M,Bengal,RCCIIT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,?,0-25,0-25,0-25,0-25,0-250,FAIL
M,Bengal,RCCIIT,A,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,?,GOOD
M,Bengal,RCCIIT,A,Absence,Present,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,501-750,EXCELLENT
M,Bengal,RCCIIT,A,Absence,Present,?,?,?,?,?,76-100,76-100,76-100,76-100,76-100,751-1000,OUTSTANDING
M,Bengal,RCCIIT,A,Absence,Present,?,?,?,?,?,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
F,Bengal,RCCIIT,A,Absence,Present,?,?,?,?,?,26-50,26-50,?,?,26-50,?,GOOD
M,Bengal,WBUT,A,Absence,Present,?,?,?,?,?,51-75,51-75,?,?,51-75,501-750,EXCELLENT
M,Bengal,WBUT,A,Absence,Present,?,?,?,?,?,76-100,76-100,?,?,76-100,751-1000,OUTSTANDING
F,Bengal,WBUT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
F,Bengal,WBUT,B,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,?,GOOD
M,Bengal,WBUT,A,Absence,Present,51-75,?,?,51-75,51-75,51-75,51-75,51-75,51-75,51-75,501-750,EXCELLENT
```

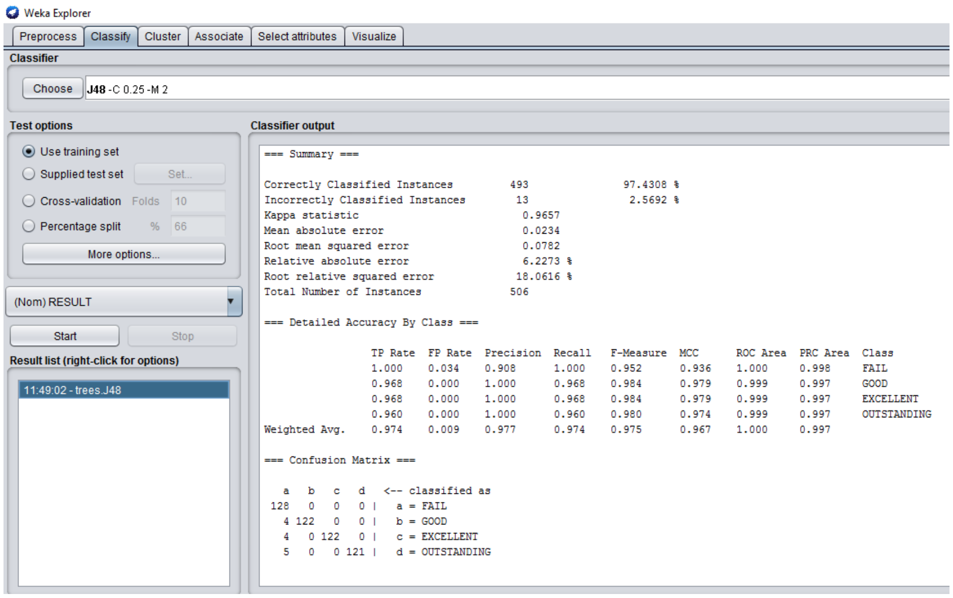
- **Test Dataset in ARFF**

```
@relation DE23i
@attribute gender {M,F}
@attribute PlaceofBirth {Bengal,Mumbai,Delhi,Bihar,Pune}
@attribute School {RCCIIT,WBUT,IEM}
@attribute SectionID {A,B,C}
@attribute 'First Class Test' {Absence,Present}
@attribute 'Second Class Test' {Present,Absence}
@attribute SEM/C {0-25,26-50,51-75,76-100}
@attribute SEM/ALGORITHM {0-25,26-50,51-75,76-100}
@attribute SEM/DBMS {0-25,26-50,51-75,76-100,57-67}
@attribute SEM/DATASTRUCTURE {0-25,26-50,51-75,76-100,25-67}
@attribute SEM/NETWORKING {0-25,26-50,51-75,76-100}
@attribute SEM/HU/PRACTICAL {26-50,51-75,76-100,0-25}
@attribute SEM/DBMS/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/C/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/NETWORKING/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/DESIGNLAB/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/GRANDTOTAL {0-250,501-750,751-1000,251-500}
@attribute RESULT {FAIL,GOOD,EXCELLENT,OUTSTANDING}

@data
M,Bengal,RCCIIT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,?,0-25,0-25,0-25,0-25,0-250,?
M,Bengal,RCCIIT,A,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,?,?
M,Bengal,RCCIIT,A,Absence,Present,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,501-750,?
M,Bengal,RCCIIT,A,Absence,Present,?,?,?,?,?,76-100,76-100,76-100,76-100,76-100,751-1000,?
M,Bengal,RCCIIT,A,Absence,Present,?,?,?,?,?,0-25,0-25,0-25,0-25,0-25,0-250,?
F,Bengal,RCCIIT,A,Absence,Present,?,?,?,?,?,26-50,26-50,?,?,26-50,?,?
M,Bengal,WBUT,A,Absence,Present,?,?,?,?,?,51-75,51-75,?,?,51-75,501-750,?
M,Bengal,WBUT,A,Absence,Present,?,?,?,?,?,76-100,76-100,?,?,76-100,751-1000,?
F,Bengal,WBUT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,?
F,Bengal,WBUT,B,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,?,?
M,Bengal,WBUT,A,Absence,Present,51-75,?,?,51-75,51-75,51-75,51-75,51-75,51-75,51-75,501-750,?
```

Manually Prediction in WEKA

i. First we have to load the dataset in WEKA Explorer and go to the classify tab. In classify tab make sure the test options should be 'Use Training Set' and focus should be on Result attribute only.

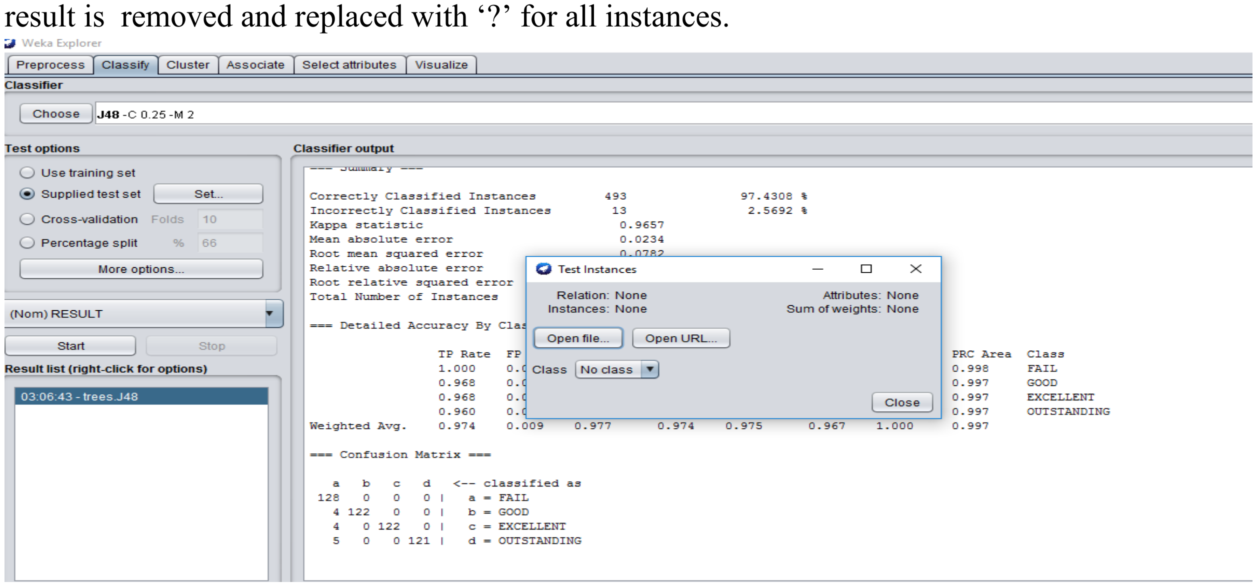
ii. Then we have to perform classification of Training set data by J48 Classifier.



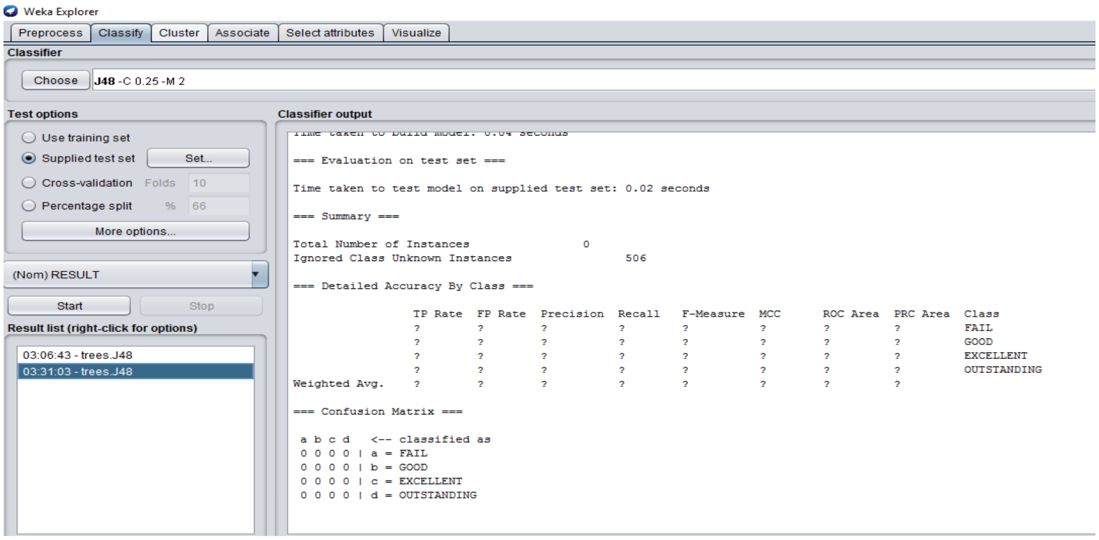
iii. Now we have to change the Test options into 'Supplied test set'.

iv. Now we have to Click on 'Supplied test set' by which 'set' tab will come.

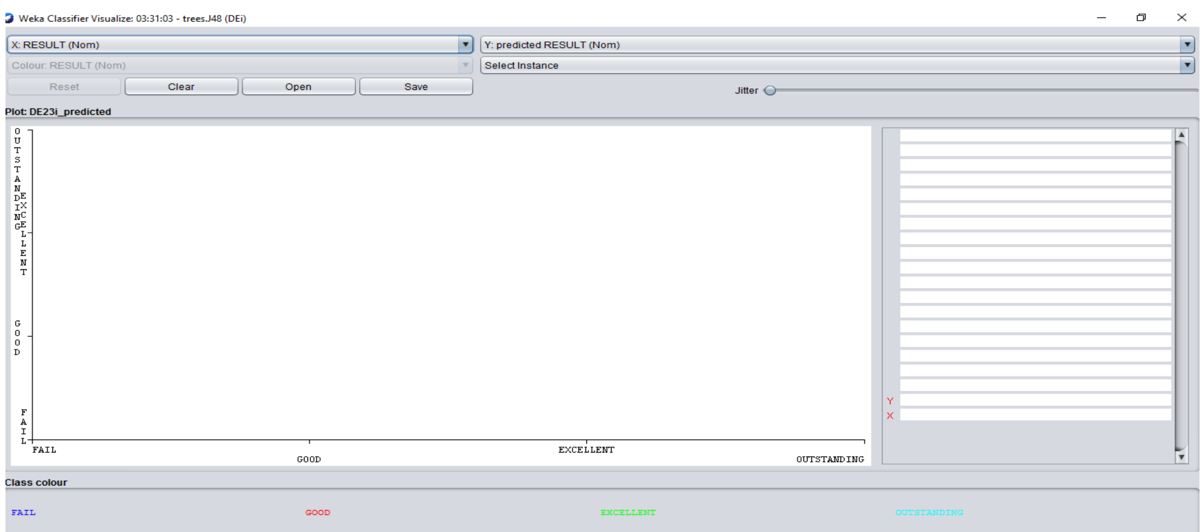
v. Now click on to 'set' tab and in Test instances select 'Open With ' to load the test set which is basically the same dataset of Training set except in the 'Result' attribute the values of result is removed and replaced with '?' for all instances.



vi. Now after loading the test data we have to perform the classification by J48 classifier on the test dataset.



vii. Now after classification of Test data in result list we have to select ‘Visualize Classifier errors .It will show a graph known as visualization of WEKA classifier. Then we have to save the visualization result which will be in ARFF format.



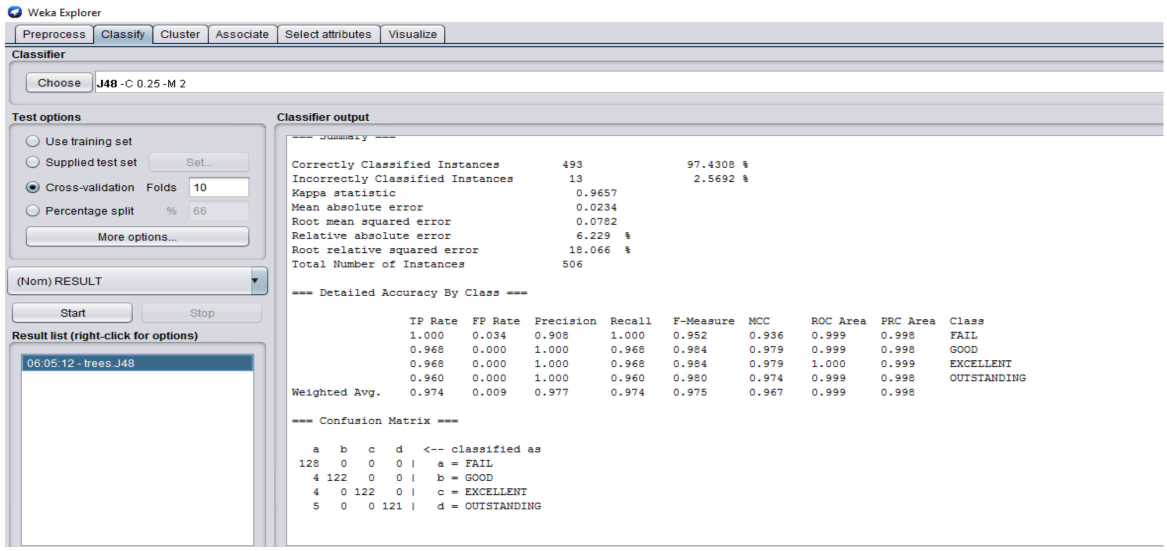
viii. Now if we open the saved result of ‘visualization of classifier errors’ in WEKA ARFF viewer we can see a new attribute named as ‘Predicted RESULT’ is generated in test dataset which maybe or may not be similar with the original result attribute of trained dataset. This known as prediction of WEKA where WEKA predicts the result of student performance which further can be studied for analysis purpose.

Here one more attribute is generated in test set known as ‘Prediction Margin’. The margin is defined as the difference between the probability predicted for the actual result and the highest probability predicted for the other results. One hypothesis as to the good performance of boosting algorithms is that they increase the margins on the training data and this gives better performance on test data.

18: prediction margin Numeric	19: predicted RESULT Nominal	20: RESULT Nominal
0.964427	FAIL	
-0.970356	GOOD	
-0.970356	EXCELLENT	
-0.972332	OUTSTANDING	
0.964427	FAIL	
-0.970356	GOOD	
-0.970356	EXCELLENT	
-0.972332	OUTSTANDING	
0.964427	FAIL	
-0.970356	GOOD	
-0.970356	EXCELLENT	
-0.972332	OUTSTANDING	
0.964427	FAIL	
-0.970356	GOOD	
-0.970356	EXCELLENT	
-0.972332	OUTSTANDING	
0.964427	FAIL	
-0.970356	GOOD	
-0.970356	EXCELLENT	
-0.972332	OUTSTANDING	
0.964427	FAIL	
-0.970356	GOOD	
-0.970356	EXCELLENT	
-0.972332	OUTSTANDING	
0.964427	FAIL	
0.003953	FAIL	
0.003953	FAIL	
0.003953	FAIL	
0.964427	FAIL	
-0.970356	GOOD	
-0.970356	EXCELLENT	
-0.972332	OUTSTANDING	
0.964427	FAIL	
-0.970356	GOOD	
-0.970356	EXCELLENT	
-0.972332	OUTSTANDING	
0.964427	FAIL	

Inbuilt WEKA Prediction

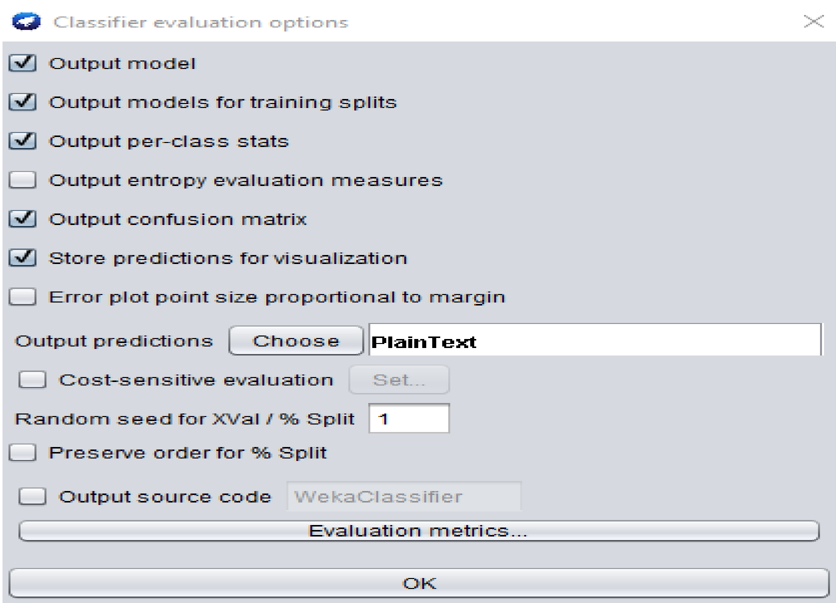
I. At first we have to load our dataset into WEKA Explorer.
II. After loading our dataset go to classify tab and start classification by J48 classifier. In classify tab Test options can be ‘Cross Validation’.



III. Then change the test option into 'Supplied Test set and load the same dataset as test file.

IV. After loading the test file in classify tab under test options select more options to go to the 'classifier evaluation options'.

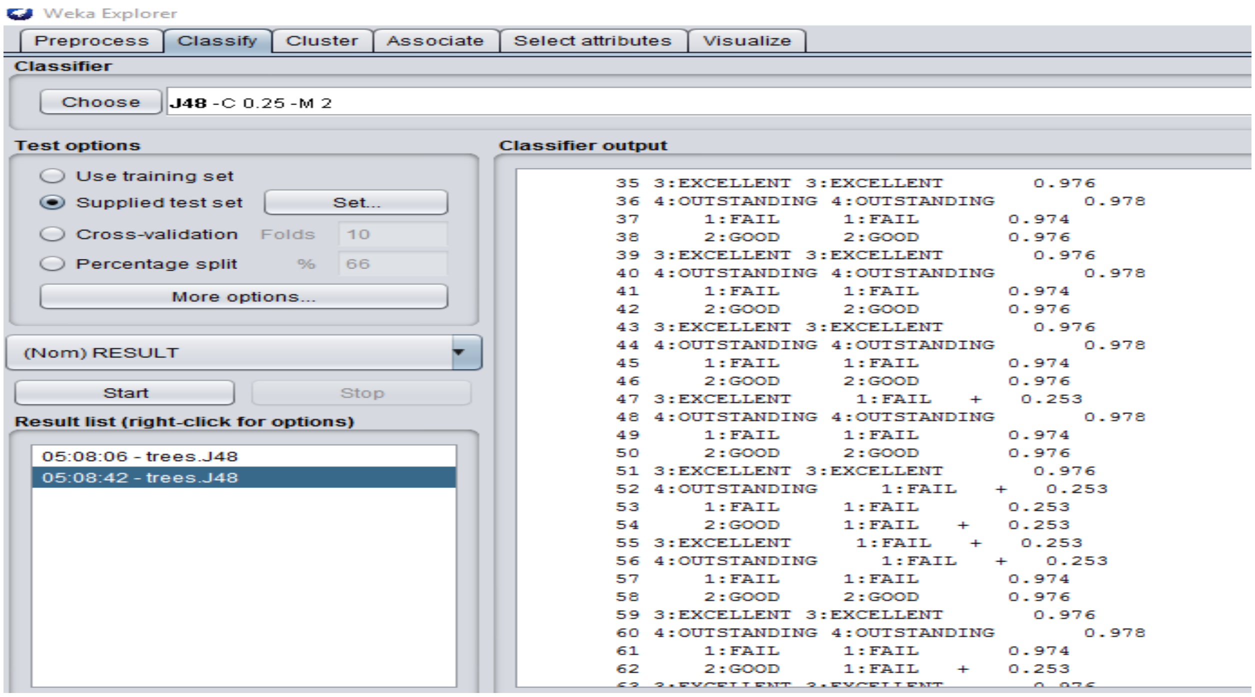
V. Now in classifier evaluation options select the output predictions and choose 'Plaintext' as prediction.



VI. Now perform the classification of test set by J48 Classifier. Now in classifier output it will be seen that WEKA performs predictions on test set. In the result the 'Predicted error' column contains predicted value of Result attribute which is the predicted result of original result of train data set. Thus WEKA performed prediction.

Here one more column is generated named as 'Prediction' which has some certain values for all instances. The 'Prediction is defined as the difference between the probability predicted for the actual result and the highest probability predicted for the other results. One hypothesis as to the good performance of boosting algorithms is that they increase the margins on the training data and this gives better performance on test data.

In the following picture for some instances the '+' sign signifies that WEKA prediction fails to match the actual result.



However from the two methods of prediction in WEKA ,they gives the same predicted result and difference between the probability predicted for the actual result and the highest probability predicted for the other results is same for both method. For the first method it is known as 'Prediction Margin 'but in second method it is known as 'Prediction'.

- **<u>Association Rule Mining</u> -**

  - **What is association mining?**

Finding frequent patterns, associations, correlations, or casual structures among set of items or objects in transaction databases, relational databases, and other information repositories

.

- **Apriori Algorithm –**

The apriori algorithm is an influential algorithm for mining frequent item sets for Boolean association rules.

Aprori uses a "bottom up" approach where frequent subsets are extended one time at a time (a step known as candidate generation and groups of candidates are tested against the data).

Apriori algorithm in Weka:

General Process

Association rule generation is usually split up into two separate steps:

1. First, minimum support is applied to find all frequent item sets in a database.

2. Second, these frequent item sets and the minimum confidence constraint are used to form rules.

While the second step is straight forward, the first step needs more attention.
Finding all frequent item sets in a database is difficult since it involves searching all possible item sets.

Support- The support for a rule X => Y is obtained by dividing the number of transactions Which satisfy the rule, N (X=>Y), by the total number of transactions, N

Support (X=>Y) =N (X=>Y) / N

The support is therefore the frequency of events for which both the LHS and RHS of the rule hold true. The higher the support the stronger the information that both type of events occur together.

Confidence- The confidence of the rule X => Y is obtained by dividing the number of Transactions which satisfy the rule N (X=>Y) by the number of transactions which contain the Body of the rule, X.

Confidence (X=>Y) = N (X=>Y) / N (X)

The confidence is the conditional probability of the RHS holding true given that the LHS Holds true. A high confidence that the LHS event leads to the RHS event implies causation or Statistical dependence.

Lift- The lift of the rule X => Y is the deviation of the support of the whole rule from the Support expected under independence given the supports of the LHS (X) and the RHS (Y).

Lift {X=>Y} = confidence (X=>Y) / support (Y)

= support (X=>Y) / support (X). support (Y)

Lift is an indication of the effect that knowledge that LHS holds true has on the probability of The RHS holding true. Hence Lift is a value that gives us information about the increase in Probability of the "then" (consequent RHS) given the "if" (antecedent LHS) part.

Lift is exactly 1: No effect (LHS and RHS independent). No relationship between Events.

Lift greater than 1: Positive effect (given that the LHS holds true, it is more likely that The Operational risk management RHS holds true). Positive dependence between events.

Lift is smaller than 1: Negative effect (when the LHS holds true, it is less likely that the RHS holds true). Negative dependence between events.

Leverage – proportion of additional examples covered by both the antecedent and the Consequent above those expected if the antecedent and consequent were independent of each Other, and finally.

$$\textbf{lev}(X \rightarrow Y) = \textbf{supp}(X,Y) - \textbf{sup}(X).\ \textbf{supp}(Y)$$

Conviction – a measure similar to Leverage that measures the departure from independence.

$$\textbf{conv}(X \rightarrow Y) = \textbf{supp}(X)(1-\textbf{supp}(Y)) / \textbf{supp}(X) - \textbf{supp}(X,Y)$$

Sample Theoretical example: Procedure of student performance analysis by rule generation method using Apriori algorithm in WEKA tools.

Let set,
Min-support = 0.1 or (10%)
Min-confidence=0.9 or (90%)
Take a student dataset-

T-ID/INSTANCES	ITEMSET/ATTRIBUTES			
	VIVA	**CT**	**ASSG**	**CLASS**
T-1	P	A	NG	FAIL
T-2	P	P	NG	PASS
T-3	A	P	NG	FAIL
T-4	A	A	G	FAIL
T-5	P	A	G	PASS

VIVA-Viva

CT-Class Test

ASSG-Assignment

P- Present

A-Absence

NG-Not Given

G-Given

Now find support count of each item set:
C1=

ITEMSET	SUPPORT
VIVA-P	3/5=0.6
VIVA-A	2/5=0.4
CT-P	2/5=0.4
CT-A	3/5=0.6
ASSG-G	2/5=0.4
ASSG-NG	3/5=0.6

Compare min support with each Item set support count.
L1= 6

ITEMSET	SUPPORT
VIVA-P	3/5=0.6
VIVA-A	2/5=0.4
CT-P	2/5=0.4
CT-A	3/5=0.6
ASSG-G	2/5=0.4
ASSG-NG	3/5=0.6

Generate pair to generate C2
C2=

Item set	Support-count
VIVA-P CT-P	1/5=0.2
VIVA-P CT-A	2/5=0.4
VIVA-P ASSG-G	1/5=0.2
VIVA-P ASSG-NG	2/5=0.4
VIVA-A CT-P	1/5=0.2
VIVA-A CT-A	1/5=0.2
VIVA-A ASSG-G	1/5=0.2
VIVA-A ASSG-NG	1/5=0.2
CT-P ASSG-G	0/5=0.0
CT-P ASSG-NG	2/5=0.4
CT-A ASSG-G	2/5=0.4
CT-A ASSG-NG	1/5=0.2

Now again compare C2 with min-support
L2= 12

Item set	Support-count
VIVA-P CT-P	1/5=0.2
VIVA-P CT-A	2/5=0.4
VIVA-P ASSG-G	1/5=0.2
VIVA-P ASSG-NG	2/5=0.4
VIVA-A CT-P	1/5=0.2
VIVA-A CT-A	1/5=0.2
VIVA-A ASSG-G	1/5=0.2
VIVA-A ASSG-NG	1/5=0.2
CT-P ASSG-NG	2/5=0.4
CT-A ASSG-G	2/5=0.4
CT-A ASSG-NG	1/5=0.2
VIVA-P CT-P	1/5=0.2

Generate pair to generate C3
C3=

Item set	Support-count
VIVA-P CT-P ASSG-G	0/5=0.0
VIVA-P CT-P ASSG-NG	1/5=0.2
VIVA-P CT-A ASSG-G	1/5=0.2
VIVA-P CT-A ASSG-NG	1/5=0.2
VIVA-A CT-P ASSG-G	0/5=0.0
VIVA-A CT-P ASSG-NG	1/5=0.2
VIVA-A CT-A ASSG-G	1/5=0.2
VIVA-A CT-A ASSG-NG	0/5=0.0

Now again compare C3 with min-support
L3= 5

Item set	Support-count
VIVA-P CT-P ASSG-NG	0.2
VIVA-P CT-A ASSG-G	0.2
VIVA-P CT-A ASSG-NG	0.2
VIVA-A CT-P ASSG-NG	0.2
VIVA-A CT-A ASSG-G	0.2

Now create association rules respect to minimum support (0.1) and confidence (90%).

Association rule	Support	Confidence	Confidence in %
VIVA-P =>CT-P	0.2	0.2/0.6=0.34	34
CT-P=> VIVA-P	0.2	0.2/0.4=0.5	50
VIVA-P =>CT-A	0.4	0.4/0.6=0.67	67
CT-A=> VIVA-P	0.4	0.4/0.6=0.67	67
VIVA-P =>ASSG-G	0.2	0.2/0.6=0.34	34
ASSG-G=> VIVA-P	0.2	0.2/0.4=0.5	50
VIVA-P =>ASSG-NG	0.4	0.4/0.6=0.67	67
ASSG-NG=> VIVA-P	0.4	0.4/0.6=0.67	67
VIVA-A =>CT-P	0.2	0.2/0.4=0.5	50
CT-P=> VIVA-A	0.2	0.2/0.4=0.5	50
VIVA-A=> CT-A	0.2	0.2/0.4=0..5	50
CT-A=> VIVA-A	0.2	0.2/0.6=0.34	34
VIVA-A=> ASSG-G	0.2	0.2/0.4=0.5	50
ASSG-G=> VIVA-A	0.2	0.2/0.4=0.5	50
VIVA-A =>ASSG-NG	0.2	0.2/0.4=0.5	50
ASSG-NG=> VIVA-A	0.2	0.2/0.6=0.34	34
CT-P=> ASSG-NG	0.4	0.4/0.4=1.0	100
ASSG-NG=> CT1-P	0.4	0.4/0.6=0.67	67
CT-A =>ASSG-G	0.4	0.4/0.6=0.67	67
ASSG-G=> CT-A	0.4	0.4/0.4=1.0	100
CT-A =>ASSG-NG	0.2	0.2/0.6=0.34	34
ASSG-NG=> CT-A	0.2	0.2/0.6=0.34	34
VIVA-P CT-P => ASSG-NG	0.2	0.2/0.2=1	100
CT1-P ASSG-NG => VIVA-P	0.2	0.2/0.4=0.5	50
VIVA-P ASSG-NG => CT-P	0.2	0.2/0.4=0.5	50
VIVA-P CT-A => ASSG-NG	0.2	0.2/0.4=0.5	50

CT-A ASSG-NG => VIVA-P	0.2	0.2/0.2=1.0	100
VIVA-P ASSG-NG => CT-A	0.2	0.2/0.4=0.5	50
VIVA-P CT-A => ASSG-G	0.2	0.2/0.4=0.5	50
CT-A ASSG-G=> VIVA-P	0.2	0.2/0.4=0.5	50
ASSG-G VIVA-P=> CT-A	0.2	0.2/0.2=1.0	100
VIVA-A CT-P=> ASSG-NG	0.2	0.2/0.2=1.0	100
CT-P ASSG-NG=> VIVA-A	0.2	0.2/0.4=0.5	50
VIVA-A ASSG-NG =>CT-P	0.2	0.2/0.2=1.0	100
VIVA-A CT-A=> ASSG-G	0.2	0.2/0.2=1.0	100
CT-A ASSG-G=> VIVA-A	0.2	0.2/0.4=0.5	50
VIVA-A ASSG-G=> CT-A	0.2	0.2/0.2=1.0	100

Compare this with min-confidence=90%

Rules	**Support**	**Confidence**
CT-P=> ASSG-NG	0.4	100
ASSG-G=> CT-A	0.4	100
VIVA-P CT-P => ASSG-NG	0.2	100
CT-A ASSG-NG => VIVA-P	0.2	100
ASSG-G VIVA-P=> CT-A	0.2	100
VIVA-A CT-P=> ASSG-NG	0.2	100
VIVA-A ASSG-NG =>CT-P	0.2	100
VIVA-A CT-A=> ASSG-G	0.2	100
VIVA-A ASSG-G=> CT-A	0.2	100

Hence the final generated association rules are-

1. **CT-P=> ASSG-NG**
2. **ASSG-G=> CT-A**
3. **VIVA-P CT-P => ASSG-NG**
4. **CT-A ASSG-NG => VIVA-P**
5. **ASSG-G VIVA-P=> CT-A**
6. **VIVA-A CT-P=> ASSG-NG**
7. **VIVA-A ASSG-NG =>CT-P**
8. **VIVA-A CT-A=> ASSG-G**
9. **VIVA-A ASSG-G=> CT-A**

There are also a lot of uninteresting rules, like a number of redundant rules (rules with a generalization of relationships of several rules, **like rule 1 with rules 6 & 7, rule 2 with rules 8 & 9 and so on**). There are some similar rules (rules with the same element in antecedent and consequent but interchanged). And there are some random relationships (rules with random relations between variables). But there are also rules that show relevant information for educational purposes, which can be very useful for the teacher in decision making about the activities and detecting students with learning problems. Starting from this information, the teacher can pay more attention to these students because they are prone to failure. As a result, the teacher can motivate them in time to pass the course.

Hence the combination of one or more association rules for overall students' performance analysis are-

1. CT-P=> ASSG-NG 3. VIVA-P CT-P => ASSG-NG	**VIVA-P CT-P ASSG-NG=>CLASS-PASS (conf-0.2/0.2=1.0)**
4. CT-A ASSG-NG => VIVA-P	**VIVA-P CT-A ASSG-NG=>CLASS-FAIL (conf-0.2/0.2=1.0)**
2. ASSG-G=> CT-A 5. ASSG-G VIVA-P=> CT-A	**VIVA-P CT-A ASSG-G=>CLASS-PASS (conf-0.2/0.2=1.0)**
1. CT-P=> ASSG-NG 6. VIVA-A CT-P=> ASSG-NG 7. VIVA-A ASSG-NG =>CT-P	**VIVA-A CT-P ASSG-NG=>CLASS-FAIL (conf-0.2/0.2=1.0)**
2. ASSG-G=> CT-A **8.** VIVA-A CT-A=> ASSG-G 9. VIVA-A ASSG-G=> CT-A	**VIVA-A CT-A ASSG-G=>CLASS-FAIL (conf-0.2/0.2=1.0)**

In here, we can see that large rules (rule no-9, 7, etc.) can contained almost all attributes which has generated by small rules (rule no-2, 1, etc.).

So, Teacher can easily check the possibility of result/class/status of overall students by looking large rules from huge no of generated association rules in weka associator.

Practical work on Apriori in Weka tools:

Since the data mining software used to generate association rules accepts data only in arff format, the researcher first converted the data on Ms Excel file into comma separated text format and then to arff format. Data in arff format of nominal form is then given to Weka associate, then select Apriori for association rule mining.

For our test we shall consider 506 students data with respect to different type of nominal attributes. The ARFF file presented bellow contains information regarding each student's performance.

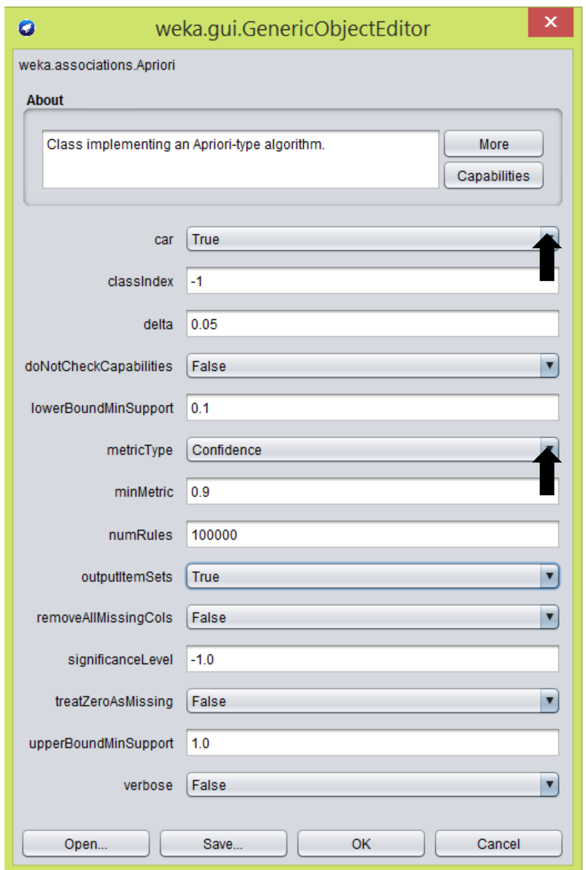
```
@relation DEi

@attribute gender {M,F}
@attribute PlaceofBirth {Bengal,Mumbai,Delhi,Bihar,Pune}
@attribute School {RCCIIT,WBUT,IEM}
@attribute SectionID {A,B,C}
@attribute 'First Class Test' {Absence,Present}
@attribute 'Second Class Test' {Present,Absence}
@attribute SEM/C {0-25,26-50,51-75,76-100}
@attribute SEM/ALGORITHM {0-25,26-50,51-75,76-100}
@attribute SEM/DBMS {0-25,26-50,51-75,76-100,57-67}
@attribute SEM/DATASTRUCTURE {0-25,26-50,51-75,76-100,25-67}
@attribute SEM/NETWORKING {0-25,26-50,51-75,76-100}
@attribute SEM/HU/PRACTICAL {26-50,51-75,76-100,0-25}
@attribute SEM/DBMS/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/C/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/NETWORKING/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/DESIGNLAB/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/GRANDTOTAL {0-250,501-750,751-1000,251-500}
@attribute RESULT {FAIL,GOOD,EXCELLENT,OUTSTANDING}

@data
M,Bengal,RCCIIT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,?,0-25,0-25,0-25,0-25,0-250,FAIL
M,Bengal,RCCIIT,A,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,?,GOOD
M,Bengal,RCCIIT,A,Absence,Present,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,501-750,EXCELLENT
M,Bengal,RCCIIT,A,Absence,Present,?,?,?,?,?,76-100,76-100,76-100,76-100,76-100,751-1000,OUTSTANDING
M,Bengal,RCCIIT,A,Absence,Present,?,?,?,?,?,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
F,Bengal,RCCIIT,A,Absence,Present,?,?,?,?,?,26-50,26-50,?,?,26-50,?,GOOD
M,Bengal,WBUT,A,Absence,Present,?,?,?,?,?,51-75,51-75,?,?,51-75,501-750,EXCELLENT
M,Bengal,WBUT,A,Absence,Present,?,?,?,?,?,76-100,76-100,?,?,76-100,751-1000,OUTSTANDING
F,Bengal,WBUT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
F,Bengal,WBUT,B,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,?,GOOD
M,Bengal,WBUT,A,Absence,Present,51-75,?,?,51-75,51-75,51-75,51-75,51-75,51-75,51-75,501-750,EXCELLENT
M,Bengal,WBUT,B,Absence,Present,76-100,?,?,76-100,76-100,76-100,76-100,76-100,76-100,76-100,751-1000,OUTSTANDING
M,Bengal,RCCIIT,A,Absence,Present,0-25,?,?,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
M,Bengal,WBUT,A,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,251-500,GOOD
F,Bengal,WBUT,A,Absence,Absence,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,?,EXCELLENT
F,Bengal,WBUT,A,Absence,Present,76-100,76-100,76-100,76-100,76-100,?,?,?,?,76-100,751-1000,OUTSTANDING
M,Bengal,WBUT,B,Absence,Present,0-25,0-25,0-25,0-25,0-25,?,?,?,?,0-25,0-250,FAIL
M,Bengal,WBUT,A,Absence,Present,26-50,26-50,26-50,26-50,26-50,?,?,?,?,26-50,?,GOOD
F,Bengal,WBUT,A,Absence,Absence,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,501-750,EXCELLENT
M,Bengal,WBUT,B,Absence,Absence,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,751-1000,OUTSTANDING
F,Bengal,WBUT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
F,Bengal,WBUT,B,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,?,GOOD
```



```
M,Bengal,WBUT,A,Absence,Present,51-75,?,?,?,51-75,51-75,51-75,51-75,51-75,51-75,501-750,EXCELLENT
M,Bengal,WBUT,A,Absence,Present,76-100,?,?,?,76-100,76-100,76-100,76-100,76-100,76-100,751-1000,OUTSTANDING
M,Bengal,WBUT,B,Absence,Present,0-25,?,?,?,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
M,Bengal,WBUT,A,Absence,Present,26-50,?,?,?,26-50,26-50,26-50,?,?,?,251-500,GOOD
M,Bengal,WBUT,B,Absence,Present,51-75,51-75,51-75,51-75,51-75,51-75,51-75,?,?,?,501-750,EXCELLENT
M,Bengal,WBUT,A,Absence,Present,76-100,76-100,76-100,76-100,76-100,76-100,76-100,?,?,?,751-1000,OUTSTANDING
M,Bengal,WBUT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
F,Bengal,WBUT,A,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,251-500,GOOD
F,Bengal,WBUT,A,Absence,Present,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,?,EXCELLENT
M,Bengal,WBUT,A,Absence,Present,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,751-1000,OUTSTANDING
F,Bengal,RCCIIT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
M,Bengal,RCCIIT,A,Absence,Present,?,?,?,?,?,?,?,?,?,26-50,251-500,GOOD
M,Bengal,WBUT,B,Absence,Present,?,?,?,?,?,?,?,?,?,51-75,?,EXCELLENT
M,Bengal,WBUT,A,Absence,Present,?,?,?,?,?,?,?,?,?,76-100,?,OUTSTANDING
M,Bengal,WBUT,B,Absence,Absence,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
F,Mumbai,WBUT,A,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,251-500,GOOD
M,Bengal,WBUT,A,Absence,Present,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,501-750,EXCELLENT
F,Delhi,WBUT,B,Absence,Present,76-100,76-100,76-100,76-100,76-100,?,?,?,?,76-100,?,OUTSTANDING
F,Bengal,WBUT,A,Absence,Present,?,?,0-25,0-25,0-25,?,?,?,?,0-25,0-250,FAIL
M,Bengal,WBUT,A,Absence,Present,?,?,26-50,26-50,26-50,?,?,?,?,26-50,251-500,GOOD
M,Bengal,IEM,A,Absence,Present,?,?,51-75,51-75,51-75,?,?,?,?,51-75,?,EXCELLENT
F,Bengal,IEM,A,Absence,Present,?,?,76-100,76-100,76-100,?,?,?,?,76-100,751-1000,OUTSTANDING
F,Bengal,IEM,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
M,Bengal,WBUT,A,Absence,Present,26-50,26-50,26-50,?,26-50,26-50,26-50,26-50,26-50,26-50,?,GOOD
M,Bengal,RCCIIT,A,Absence,Present,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,?,501-750,EXCELLENT
F,Bengal,IEM,A,Absence,Absence,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,751-1000,OUTSTANDING
F,Bengal,IEM,A,Absence,Absence,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
F,Bengal,IEM,A,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,?,GOOD
F,Bengal,IEM,A,Absence,Present,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,501-750,EXCELLENT
M,Bengal,IEM,A,Absence,Present,?,76-100,76-100,?,?,?,?,?,?,?,751-1000,OUTSTANDING
F,Bengal,IEM,A,Absence,Present,0-25,0-25,0-25,?,?,?,?,?,?,?,?,FAIL
F,Bengal,IEM,A,Absence,Absence,26-50,26-50,26-50,?,?,?,?,?,?,?,251-500,GOOD
M,Bengal,IEM,A,Absence,Absence,51-75,51-75,51-75,?,?,?,?,?,?,?,?,EXCELLENT
M,Bengal,WBUT,A,Absence,Present,76-100,76-100,76-100,?,?,?,?,?,?,?,751-1000,OUTSTANDING
M,Bihar,WBUT,B,Absence,Present,0-25,?,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
M,Bengal,WBUT,A,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,251-500,GOOD
M,Bengal,WBUT,A,Absence,Present,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,?,EXCELLENT
F,Bihar,WBUT,A,Absence,Present,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,751-1000,OUTSTANDING
F,Bengal,RCCIIT,A,Absence,Absence,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
M,Bengal,IEM,A,Absence,Present,?,?,?,?,?,26-50,26-50,?,?,?,?,GOOD
M,Pune,IEM,A,Absence,Absence,?,?,?,?,?,51-75,51-75,51-75,51-75,51-75,501-750,EXCELLENT
M,Bengal,IEM,A,Absence,Present,?,?,?,?,?,76-100,76-100,76-100,76-100,76-100,?,OUTSTANDING
M,Bengal,IEM,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
```

………………………………………………………………… Up to 506 instances along with 18 attributes.

Using the Apriori Algorithm we want to find the association rules that have **min Support=0.1(10%)** and **minimum confidence=0.9(90%).** We will do this using WEKA GUI. After we launch the WEKA application and open the ***DEi.arff*** file, we move to the **Associate** tab and we set up the following configuration:



In here, we can set minimum support= **0.1**, because this can generate more frequent item set. If we set minimum support= **0.2** or more, then this can remove many attributes, but minimum no of attributes is not sufficient to give a proper decision.

But, minimum confidence=**0.9**, can set higher because this boundary can give less amount of rules.

Result of Apriori Algorithm

=== Run information ===

Scheme: weka.associations.Apriori -N 100000 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -A -c -1

Relation: DEi

Instances: 506

Attributes: 18

gender

PlaceofBirth

School

SectionID

First Class Test

Second Class Test

SEM/C

SEM/ALGORITHM

SEM/DBMS

SEM/DATASTRUCTURE

SEM/NETWORKING

SEM/HU/PRACTICAL

SEM/DBMS/PRACTICAL

SEM/C/PRACTICAL

SEM/NETWORKING/PRACTICAL

SEM/DESIGNLAB/PRACTICAL

SEM/GRANDTOTAL

RESULT

=== Associator model (full training set) ===

Apriori

=======

Minimum support: 0.1 (51 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 73

Size of set of large itemsets L(2): 515

Size of set of large itemsets L(3): 2243

Size of set of large itemsets L(4): 6535

Size of set of large itemsets L(5): 13740

Size of set of large itemsets L(6): 21510

Size of set of large itemsets L(7): 25455

Size of set of large itemsets L(8): 22960

Size of set of large itemsets L(9): 15791

Size of set of large itemsets L(10): 8214

Size of set of large itemsets L(11): 3173

Size of set of large itemsets L(12): 881

Size of set of large itemsets L(13): 166

Size of set of large itemsets L(14): 19

Size of set of large itemsets L(15): 1

```
Best rules found:

  1. SEM/DESIGNLAB/PRACTICAL=0-25 125 ==> RESULT=FAIL 125    conf:(1)
  2. SEM/NETWORKING=0-25 124 ==> RESULT=FAIL 124    conf:(1)
  3. SEM/GRANDTOTAL=0-250 124 ==> RESULT=FAIL 124    conf:(1)
  4. SEM/DESIGNLAB/PRACTICAL=0-25 SEM/GRANDTOTAL=0-250 124 ==> RESULT=FAIL 124    conf:(1)
  5. SEM/C=0-25 123 ==> RESULT=FAIL 123    conf:(1)
  6. SEM/DBMS=0-25 123 ==> RESULT=FAIL 123    conf:(1)
  7. SEM/DATASTRUCTURE=0-25 123 ==> RESULT=FAIL 123    conf:(1)
  8. SEM/DBMS/PRACTICAL=0-25 123 ==> RESULT=FAIL 123    conf:(1)
  9. SEM/DATASTRUCTURE=0-25 SEM/NETWORKING=0-25 123 ==> RESULT=FAIL 123    conf:(1)
 10. SEM/NETWORKING=0-25 SEM/DESIGNLAB/PRACTICAL=0-25 123 ==> RESULT=FAIL 123    conf:(1)
 11. SEM/NETWORKING=0-25 SEM/GRANDTOTAL=0-250 123 ==> RESULT=FAIL 123    conf:(1)
 12. SEM/NETWORKING=0-25 SEM/DESIGNLAB/PRACTICAL=0-25 SEM/GRANDTOTAL=0-250 123 ==> RESULT=FAIL 123    conf:(1)
 13. SEM/C/PRACTICAL=0-25 122 ==> RESULT=FAIL 122    conf:(1)
 14. SEM/NETWORKING/PRACTICAL=0-25 122 ==> RESULT=FAIL 122    conf:(1)
 15. SEM/DESIGNLAB/PRACTICAL=26-50 122 ==> RESULT=GOOD 122    conf:(1)
 16. SEM/DESIGNLAB/PRACTICAL=51-75 122 ==> RESULT=EXCELLENT 122    conf:(1)
 17. SEM/C=0-25 SEM/NETWORKING=0-25 122 ==> RESULT=FAIL 122    conf:(1)
 18. SEM/DBMS=0-25 SEM/DATASTRUCTURE=0-25 122 ==> RESULT=FAIL 122    conf:(1)
 19. SEM/DBMS=0-25 SEM/NETWORKING=0-25 122 ==> RESULT=FAIL 122    conf:(1)
 20. SEM/DATASTRUCTURE=0-25 SEM/DESIGNLAB/PRACTICAL=0-25 122 ==> RESULT=FAIL 122    conf:(1)
 21. SEM/DATASTRUCTURE=0-25 SEM/GRANDTOTAL=0-250 122 ==> RESULT=FAIL 122    conf:(1)
 22. SEM/C/PRACTICAL=0-25 SEM/NETWORKING/PRACTICAL=0-25 122 ==> RESULT=FAIL 122    conf:(1)
 23. SEM/DBMS=0-25 SEM/DATASTRUCTURE=0-25 SEM/NETWORKING=0-25 122 ==> RESULT=FAIL 122    conf:(1)
 24. SEM/DATASTRUCTURE=0-25 SEM/NETWORKING=0-25 SEM/DESIGNLAB/PRACTICAL=0-25 122 ==> RESULT=FAIL 122    conf:(1)
 25. SEM/DATASTRUCTURE=0-25 SEM/NETWORKING=0-25 SEM/GRANDTOTAL=0-250 122 ==> RESULT=FAIL 122    conf:(1)
 26. SEM/DATASTRUCTURE=0-25 SEM/DESIGNLAB/PRACTICAL=0-25 SEM/GRANDTOTAL=0-250 122 ==> RESULT=FAIL 122    conf:(1)
 27. SEM/DATASTRUCTURE=0-25 SEM/NETWORKING=0-25 SEM/DESIGNLAB/PRACTICAL=0-25 SEM/GRANDTOTAL=0-250 122 ==> RESULT=FAIL 122    conf:(1)
 28. SEM/C=51-75 121 ==> RESULT=EXCELLENT 121    conf:(1)
 29. SEM/ALGORITHM=0-25 121 ==> RESULT=FAIL 121    conf:(1)
 30. SEM/HU/PRACTICAL=0-25 121 ==> RESULT=FAIL 121    conf:(1)
 31. SEM/DESIGNLAB/PRACTICAL=76-100 121 ==> RESULT=OUTSTANDING 121    conf:(1)
 32. SEM/GRANDTOTAL=751-1000 121 ==> RESULT=OUTSTANDING 121    conf:(1)
 33. SEM/C=0-25 SEM/DBMS=0-25 121 ==> RESULT=FAIL 121    conf:(1)
 34. SEM/C=0-25 SEM/DATASTRUCTURE=0-25 121 ==> RESULT=FAIL 121    conf:(1)
 35. SEM/C=0-25 SEM/DESIGNLAB/PRACTICAL=0-25 121 ==> RESULT=FAIL 121    conf:(1)
 36. SEM/C=0-25 SEM/GRANDTOTAL=0-250 121 ==> RESULT=FAIL 121    conf:(1)
 37. SEM/ALGORITHM=0-25 SEM/DBMS=0-25 121 ==> RESULT=FAIL 121    conf:(1)
 38. SEM/DBMS=0-25 SEM/DESIGNLAB/PRACTICAL=0-25 121 ==> RESULT=FAIL 121    conf:(1)
 39. SEM/DBMS=0-25 SEM/GRANDTOTAL=0-250 121 ==> RESULT=FAIL 121    conf:(1)
 40. SEM/NETWORKING=0-25 SEM/DBMS/PRACTICAL=0-25 121 ==> RESULT=FAIL 121    conf:(1)
 41. SEM/HU/PRACTICAL=0-25 SEM/DBMS/PRACTICAL=0-25 121 ==> RESULT=FAIL 121    conf:(1)
 42. SEM/DBMS/PRACTICAL=0-25 SEM/C/PRACTICAL=0-25 121 ==> RESULT=FAIL 121    conf:(1)
 43. SEM/DBMS/PRACTICAL=0-25 SEM/NETWORKING/PRACTICAL=0-25 121 ==> RESULT=FAIL 121    conf:(1)
```

```
63932. First Class Test=Absence Second Class Test=Present SEM/DBMS=76-100 SEM/NETWORKING=76-100 SEM/DBMS/PRACTICAL=76-100 SEM/C/PRACTICAL=76-100 SEM/NETWORKING/PRACTICAL=76-100 SEM/DESIGNLAB/PRACTICAL=76-100 SEM/GRANDTOTAL=751-1000 67 ==> RESULT=
63933. First Class Test=Absence Second Class Test=Present SEM/DBMS=76-100 SEM/HU/PRACTICAL=76-100 SEM/DBMS/PRACTICAL=76-100 SEM/C/PRACTICAL=76-100 SEM/NETWORKING/PRACTICAL=76-100 SEM/DESIGNLAB/PRACTICAL=76-100 SEM/GRANDTOTAL=751-1000 67 ==> RESUL
63934. gender=M PlaceofBirth=Bengal First Class Test=Absence SEM/C=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 67 ==> RESULT=EXCELLENT 67
63935. gender=M PlaceofBirth=Bengal First Class Test=Absence SEM/C=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 67 ==> RESULT=EXCELLE
63936. gender=M PlaceofBirth=Bengal Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 67 ==> RESULT=EXCELLENT 67    conf:(1
63937. gender=M PlaceofBirth=Bengal Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 67 ==> RESULT=EXCELLENT 67
63938. gender=M PlaceofBirth=Bengal Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 67 ==> RESULT=EXCELLENT 67    c
63939. gender=M PlaceofBirth=Bengal Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 67 ==> RESULT=EXCELLENT 67
63940. gender=M PlaceofBirth=Bengal Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 67 ==> RESULT=EXCELLENT 67
63941. gender=M PlaceofBirth=Bengal Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 67 ==> RESULT=EXCELLENT 67
63942. gender=M PlaceofBirth=Bengal Second Class Test=Present SEM/C=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 67 ==> RESULT=EXCELLENT 67
63943. gender=M PlaceofBirth=Bengal Second Class Test=Present SEM/C=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 67 ==> RESULT=EXCE
63944. gender=M PlaceofBirth=Bengal Second Class Test=Present SEM/C=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 67 ==> RESULT=EXCELLE
63945. gender=M PlaceofBirth=Bengal Second Class Test=Present SEM/C=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 67 ==> RESULT=EXCEL
63946. gender=M PlaceofBirth=Bengal Second Class Test=Present SEM/C=51-75 SEM/DATASTRUCTURE=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 67 ==> RESULT=EXC
63947. gender=M PlaceofBirth=Bengal Second Class Test=Present SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 67 ==> RESULT=EXCE
63948. gender=M PlaceofBirth=Bengal Second Class Test=Present SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 67 ==> R
63949. gender=M PlaceofBirth=Bengal SEM/C=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXCELLENT 67
63950. gender=M PlaceofBirth=Bengal SEM/C=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXCELLENT
63951. gender=M First Class Test=Absence SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXCELLENT 67
63952. gender=M First Class Test=Absence SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXCELLENT 67
63953. gender=M First Class Test=Absence SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXCELLE
63954. gender=M First Class Test=Absence SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXCELLENT 67
63955. gender=M First Class Test=Absence SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXCEL
63956. gender=M First Class Test=Absence SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXCELLEN
63957. gender=M First Class Test=Absence SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXCELLENT 67
63958. gender=M First Class Test=Absence SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXC
63959. gender=M First Class Test=Absence SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/HU/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXCELL
63960. gender=M First Class Test=Absence SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXCE
63961. gender=M First Class Test=Absence SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXCELLENT 67
63962. gender=M First Class Test=Absence SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXCELL
63963. gender=M First Class Test=Absence SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXCELLENT
63964. gender=M First Class Test=Absence SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/NETWORKING=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXCELLE
63965. gender=M First Class Test=Absence SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXCEL
63966. gender=M First Class Test=Absence SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESULT=EXC
63967. gender=M First Class Test=Absence SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> R
63968. gender=M First Class Test=Absence SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RESU
63969. gender=M First Class Test=Absence SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RE
63970. gender=M First Class Test=Absence SEM/ALGORITHM=51-75 SEM/DATASTRUCTURE=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==>
63971. gender=M First Class Test=Absence SEM/ALGORITHM=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 67 ==> RES
63972. gender=M Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 67 ==> RESULT=EXCELLENT 67    conf
63973. gender=M Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 67 ==> RESULT=EXCELLENT 67    conf:(1
63974. gender=M Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 67 ==> RESULT=EXCELLENT 67
63975. gender=M Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 67 ==> RESULT=EXCELLENT 67    conf:
63976. gender=M Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 67 ==> RESULT=EXCELLENT 67
```

……

```
99958. gender=M PlaceofBirth=Bengal First Class Test=Absence Second Class Test=Present SEM/DBMS=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56
99959. gender=M PlaceofBirth=Bengal First Class Test=Absence Second Class Test=Present SEM/DBMS=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLE
99960. gender=M PlaceofBirth=Bengal First Class Test=Absence Second Class Test=Present SEM/DBMS=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT
99961. gender=M PlaceofBirth=Bengal First Class Test=Absence Second Class Test=Present SEM/DBMS=51-75 SEM/NETWORKING=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLEN
99962. gender=M PlaceofBirth=Bengal First Class Test=Absence Second Class Test=Present SEM/DBMS=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELL
99963. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/C/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    conf:(1)
99964. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    con
99965. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/HU/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    conf:(1)
99966. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/HU/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    c
99967. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    conf:(1)
99968. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56
99969. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    cc
99970. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    conf:(1)
99971. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    conf
99972. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/NETWORKING=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    conf:(1)
99973. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/NETWORKING=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    cc
99974. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/NETWORKING=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    conf:
99975. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    conf:(1)
99976. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56
99977. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/HU/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    con
99978. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    c
99979. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    conf:(1)
99980. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56
99981. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    conf:(1
99982. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56
99983. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    c
99984. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    conf:
99985. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56
99986. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/HU/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56
99987. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56
99988. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/DBMS=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    conf:(1)
99989. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/DBMS=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56
99990. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/DBMS=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    cc
99991. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/DBMS=51-75 SEM/NETWORKING=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56
99992. gender=M First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/DBMS=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56
99993. gender=M First Class Test=Absence Second Class Test=Present SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56    c
99994. gender=M First Class Test=Absence Second Class Test=Present SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLEN
99995. gender=M First Class Test=Absence Second Class Test=Present SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56
99996. gender=M First Class Test=Absence Second Class Test=Present SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELL
99997. gender=M First Class Test=Absence Second Class Test=Present SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT
99998. gender=M First Class Test=Absence Second Class Test=Present SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLENT 56
99999. gender=M First Class Test=Absence Second Class Test=Present SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCE
100000. gender=M First Class Test=Absence Second Class Test=Present SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/HU/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 56 ==> RESULT=EXCELLE
```

Meaning of the parameters mentioned above

-N(required number of rules output)	100000
-C (the minimum confidence of a rule)	0.9
D (delta at which the minimum support is decreased at each iteration)	0.05
-U (upper bound for minimum support)	1.0
-M (the lower bound for the minimum support)	0.1

In selecting rules for discussion, the researcher focused on the rules generated from the Superset of frequent or large item set consisting of the highest size of large item set. The Following are rules selected for discussion from experiment:

Best rules found:

83367.

PlaceofBirth=Bengal First Class Test=Absence Second Class Test=Present SEM/C=0-25 SEM/ALGORITHM=0-25 SEM/DATASTRUCTURE=0-25 SEM/NETWORKING=0-25 SEM/HU/PRACTICAL=0-25 SEM/DBMS/PRACTICAL=0-25 SEM/C/PRACTICAL=0-25 SEM/NETWORKING/PRACTICAL=0-25 SEM/DESIGNLAB/PRACTICAL=0-25 SEM/GRANDTOTAL=0-250 61 ==> RESULT=**FAIL** 61 conf:(1)

83368.

PlaceofBirth=Bengal First Class Test=Absence Second Class Test=Present SEM/C=0-25 SEM/DBMS=0-25 SEM/DATASTRUCTURE=0-25 SEM/NETWORKING=0-25 SEM/HU/PRACTICAL=0-25 SEM/DBMS/PRACTICAL=0-25 SEM/C/PRACTICAL=0-25 SEM/NETWORKING/PRACTICAL=0-25 SEM/DESIGNLAB/PRACTICAL=0-25 SEM/GRANDTOTAL=0-250 61 ==> RESULT=**FAIL** 61 conf:(1)

83369.

PlaceofBirth=Bengal First Class Test=Absence Second Class Test=Present SEM/C=0-25 SEM/ALGORITHM=0-25 SEM/DBMS=0-25 SEM/DATASTRUCTURE=0-25 SEM/NETWORKING=0-25 SEM/HU/PRACTICAL=0-25 SEM/DBMS/PRACTICAL=0-25 SEM/C/PRACTICAL=0-25 SEM/NETWORKING/PRACTICAL=0-25 SEM/DESIGNLAB/PRACTICAL=0-25 SEM/GRANDTOTAL=0-250 61 ==> RESULT=**FAIL** 61 conf:(1)

The meaning of those 3 large rules from 100000 rules is, those student/group of student who have **PlaceofBirth=Bengal, First Class Test=Absence, Second Class Test=Present** & In semester getting the marks in **SEM/C from 0-25, SEM/ALGORITHM from 0-25, SEM/DBMS from 0-25, SEM/DATASTRUCTURE from 0-25, SEM/NETWORKING from 0-25, SEM/HU/PRACTICAL from 0-25, SEM/DBMS/PRACTICAL from 0-25, SEM/C/PRACTICAL from 0-25, SEM/NETWORKING/PRACTICAL from 0-25, SEM/DESIGNLAB/PRACTICAL from 0-25, SEM/GRANDTOTAL from 0-250** then possibility of those group of students will also fall in **FAIL** class .

The support for this rule can be computed by dividing the figure on the right-hand side Of the rule **61** by the total number of instances considered in generating association Rules, **506**. This rule has a support of **12%**. The number **61** on the right-hand-side of the Rule indicates the number of items covered by its antecedent. The confidence is also Computed by dividing the figure on the left-hand-side of the rule by the figure on the Right-hand-side of the rule **(61/61=1).**

50120.

PlaceofBirth=Bengal First Class Test=Absence SEM/C=26-50 SEM/ALGORITHM=26-50 SEM/DBMS=26-50 SEM/DATASTRUCTURE=26-50 SEM/NETWORKING=26-50 SEM/HU/PRACTICAL=26-50 SEM/DBMS/PRACTICAL=26-50 SEM/C/PRACTICAL=26-50 SEM/NETWORKING/PRACTICAL=26-50 SEM/DESIGNLAB/PRACTICAL=26-50 SEM/GRANDTOTAL=251-500 73 ==> RESULT=**GOOD** 73 conf:(1)

50121.

First Class Test=Absence Second Class Test=Present SEM/C=26-50 SEM/ALGORITHM=26-50 SEM/DBMS=26-50 SEM/DATASTRUCTURE=26-50 SEM/NETWORKING=26-50 SEM/HU/PRACTICAL=26-50 SEM/DBMS/PRACTICAL=26-50 SEM/C/PRACTICAL=26-50 SEM/NETWORKING/PRACTICAL=26-50 SEM/DESIGNLAB/PRACTICAL=26-50 SEM/GRANDTOTAL=251-500 73 ==> RESULT=**GOOD** 73 conf:(1)

The meaning of those 2 large rules from 100000 rules is, those student/group of student who have **PlaceofBirth=Bengal, First Class Test=Absence, Second Class Test=Present** & In semester getting the marks in **SEM/C from 26-50, SEM/ALGORITHM from 26-50, SEM/DBMS from 26-50, SEM/DATASTRUCTURE from 26-50, SEM/NETWORKING from 26-50, SEM/HU/PRACTICAL from 26-50, SEM/DBMS/PRACTICAL from 26-50, SEM/C/PRACTICAL from 26-50, SEM/NETWORKING/PRACTICAL from 26-50, SEM/DESIGNLAB/PRACTICAL from 26-50, SEM/GRANDTOTAL from 251-500** then possibility of those group of students will also fall in **GOOD** class .

The support for this rule can be computed by dividing the figure on the right-hand side Of the rule **73** by the total number of instances considered in generating association Rules, **506**. This rule has a support of **14%**. The number **73** on the right-hand-side of the Rule indicates the number of items covered by its antecedent. The confidence is also Computed by dividing the figure on the left-hand-side of the rule by the figure on the Right-hand- side of the rule **(73/73=1).**

94478.

gender=M PlaceofBirth=Bengal First Class Test=Absence Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 58 ==> RESULT=**EXCELLENT** 58 conf:(1)

94479.

gender=M PlaceofBirth=Bengal Second Class Test=Present SEM/C=51-75 SEM/ALGORITHM=51-75 SEM/DBMS=51-75 SEM/DATASTRUCTURE=51-75 SEM/NETWORKING=51-75 SEM/HU/PRACTICAL=51-75 SEM/DBMS/PRACTICAL=51-75 SEM/C/PRACTICAL=51-75 SEM/NETWORKING/PRACTICAL=51-75 SEM/DESIGNLAB/PRACTICAL=51-75 SEM/GRANDTOTAL=501-750 58 ==> RESULT=**EXCELLENT** 58 conf:(1)

The meaning of those 2 large rules from 100000 rules is, those student/group of student who have **gender=M, PlaceofBirth=Bengal, First Class Test=Absence, Second Class Test=Present** & In semester getting the marks in **SEM/C from 51-75, SEM/ALGORITHM from 51-75, SEM/DBMS from 51-75, SEM/DATASTRUCTURE from 51-75, SEM/NETWORKING from 51-75, SEM/HU/PRACTICAL from 51-75, SEM/DBMS/PRACTICAL from 51-75, SEM/C/PRACTICAL from 51-75, SEM/NETWORKING/PRACTICAL from 51-75, SEM/DESIGNLAB/PRACTICAL from 51-75, SEM/GRANDTOTAL from 501-750** then possibility of those group of students will also fall in **EXCELLENT** class.

The support for this rule can be computed by dividing the figure on the right-hand side Of the rule **58** by the total number of instances considered in generating association Rules, **506**. This rule has a support of **11%**. The number **58** on the right-hand-side of the Rule indicates the number of items covered by its antecedent. The confidence is also Computed by dividing the figure on the left-hand-side of the rule by the figure on the Right- hand-side of the rule **(58/58=1).**

97712.

gender=M First Class Test=Absence SEM/C=76-100 SEM/ALGORITHM=76-100 SEM/DBMS=76-100 SEM/DATASTRUCTURE=76-100 SEM/NETWORKING=76-100 SEM/HU/PRACTICAL=76-100 SEM/DBMS/PRACTICAL=76-100 SEM/C/PRACTICAL=76-100 SEM/NETWORKING/PRACTICAL=76-100 SEM/DESIGNLAB/PRACTICAL=76-100 57 ==> RESULT=**OUTSTANDING** 57 conf:(1)

97730.

PlaceofBirth=Bengal First Class Test=Absence Second Class Test=Present SEM/C=76-100 SEM/ALGORITHM=76-100 SEM/DBMS=76-100 SEM/NETWORKING=76-100 SEM/HU/PRACTICAL=76-100 SEM/DBMS/PRACTICAL=76-100 SEM/C/PRACTICAL=76-100 SEM/NETWORKING/PRACTICAL=76-100 SEM/DESIGNLAB/PRACTICAL=76-100 SEM/GRANDTOTAL=751-1000 57 ==> RESULT=**OUTSTANDING** 57 conf:(1)

The meaning of those 2 large rules from 100000 rules is, those student/group of student who have **gender=M, PlaceofBirth=Bengal, First Class Test=Absence, Second Class Test=Present** & In semester getting the marks in **SEM/C from 76-100, SEM/ALGORITHM from 76-100, SEM/DBMS from 76-100, SEM/DATASTRUCTURE from 76-100, SEM/NETWORKING from 76-100, SEM/HU/PRACTICAL from 76-100, SEM/DBMS/PRACTICAL from 76-100, SEM/C/PRACTICAL from 76-100, SEM/NETWORKING/PRACTICAL from 76-100, SEM/DESIGNLAB/PRACTICAL from 76-100, SEM/GRANDTOTAL from 751-1000** then possibility of those group of students will also fall in **OUTSTANDING** class.

The support for this rule can be computed by dividing the figure on the right-hand side Of the rule **57** by the total number of instances considered in generating association Rules, **506**. This rule has a support of **11%**. The number **57** on the right-hand-side of the Rule indicates the number of items covered by its antecedent. The confidence is also Computed by dividing the figure on the left-hand-side of the rule by the figure on the Right-hand- side of the rule **(57/57=1).**

There are also a lot of uninteresting rules, like a number of redundant rules (rules with a Generalization of relationships of several rules, like rule **83369** with rules **83367** and **83368**). There are some similar rules (rules with the same element in antecedent and consequent but interchanged). And there are some random relationships (rules with random relations between variables).But there are also rules that show relevant information for educational purposes, which can be very useful for the teacher in decision making about the activities and detecting students with learning problems. Starting from this information, the teacher can pay more attention to these students because they are prone to failure.

Useful Concepts:

Interestingness measures of rules in weka:

For the dataset, association rules of the form X -> Y, where the frequent item-sets are generated using methods Aproiri techniques. The item-sets X and Y are called antecedent and consequent of the rule respectively. Generation of association rules (AR) is generally controlled by the two measures or metrics Called support and confidence, Some important are given below.

1. P(X)= count of total no of tuples at antecedent
2. P(Y)= count of total no of tuples at consequent
3. P(XY)=P(X U Y)=P(X,Y)=P(X→Y)= total no of tuples that contain both X and Y

Now, In this Student Performance dataset, we can calculate the interestingness as per as Weka results for every generating association rules.
But here, We only calculate for one rule which was generated by weka.

```
Best rules found:

 1. School=RCCIIT Second Class Test=Present 167 ==> First Class Test=Absence 164    <conf:(0.98)> lift:(1.21) lev:(0.06) [28] conv:(7.92)
```

In here,

itemset (School=RCCIIT Second Class Test=Present)= **P(X)**= 167
itemset (First Class Test=Absence) = **P(Y)**= 410
itemset (School=RCCIIT Second Class Test=Present First Class Test=Absence)= **P(XY)**=**P(X U Y)**= 164

To select interesting rules from the set of all possible rules, constraints on various measures of significance and interest can be used. The best-known constraints are minimum thresholds on support and confidence.

Support:

The support for a rule X => Y is obtained by dividing the number of transactions which satisfy the rule, N {X=>Y}, by the total number of transactions N.

The support supp(*X*) or supp(Y) of an itemset X or Y is defined as the proportion of transactions in the data set which contain the itemset.

Support {X=>Y} =N {X=>Y} / N

supp(X)= no. of transactions which contain the itemset X / total no. of transactions

supp(Y)= no. of transactions which contain the itemset Y / total no. of transactions

In the example database, the itemset {School=RCCIIT Second Class Test=Present First Class Test=Absence} has a support of 164/506= 0.324 since it occurs in 32% of all transactions. To be even more explicit we can point out that 164 is the number of transactions from the database which contain the itemset { School=RCCIIT Second Class Test=Present First Class Test=Absence } while 506 represents the total number of transactions.

Coverage: [supp(School=RCCIIT Second Class Test=Present)]=**supp(X)**=167 / 506=0.33

Prevalence: [supp(First Class Test=Absence)]= **supp(Y)**= 410 / 506=0.81

Confidence:

The confidence of a rule is defined:

Conf(X→Y) = Supp(X U Y)/Supp(X)

For the rule { School=RCCIIT Second Class Test=Present }=>{ First Class Test=Absence } we have the following confidence:

supp({School=RCCIIT Second Class Test=Present First Class Test=Absence }) / supp({School=RCCIIT Second Class Test=Present }) = 0.324 / 0.33 = 0.98
This means that for 98% of the transactions containing milk and bread the rule is correct.
Confidence can be interpreted as an estimate of the probability $P(Y \mid X)$, the probability of finding the RHS of the rule in transactions under the condition that these transactions also contain the LHS.

Lift:

The lift of a rule is defined as:

$$\text{Lift}(X \rightarrow Y) = \frac{\text{supp}(X \cup Y)}{\text{supp}(Y) * \text{sup}(X)}$$

The rule { School=RCCIIT Second Class Test=Present }=>{ First Class Test=Absence } has the following lift:

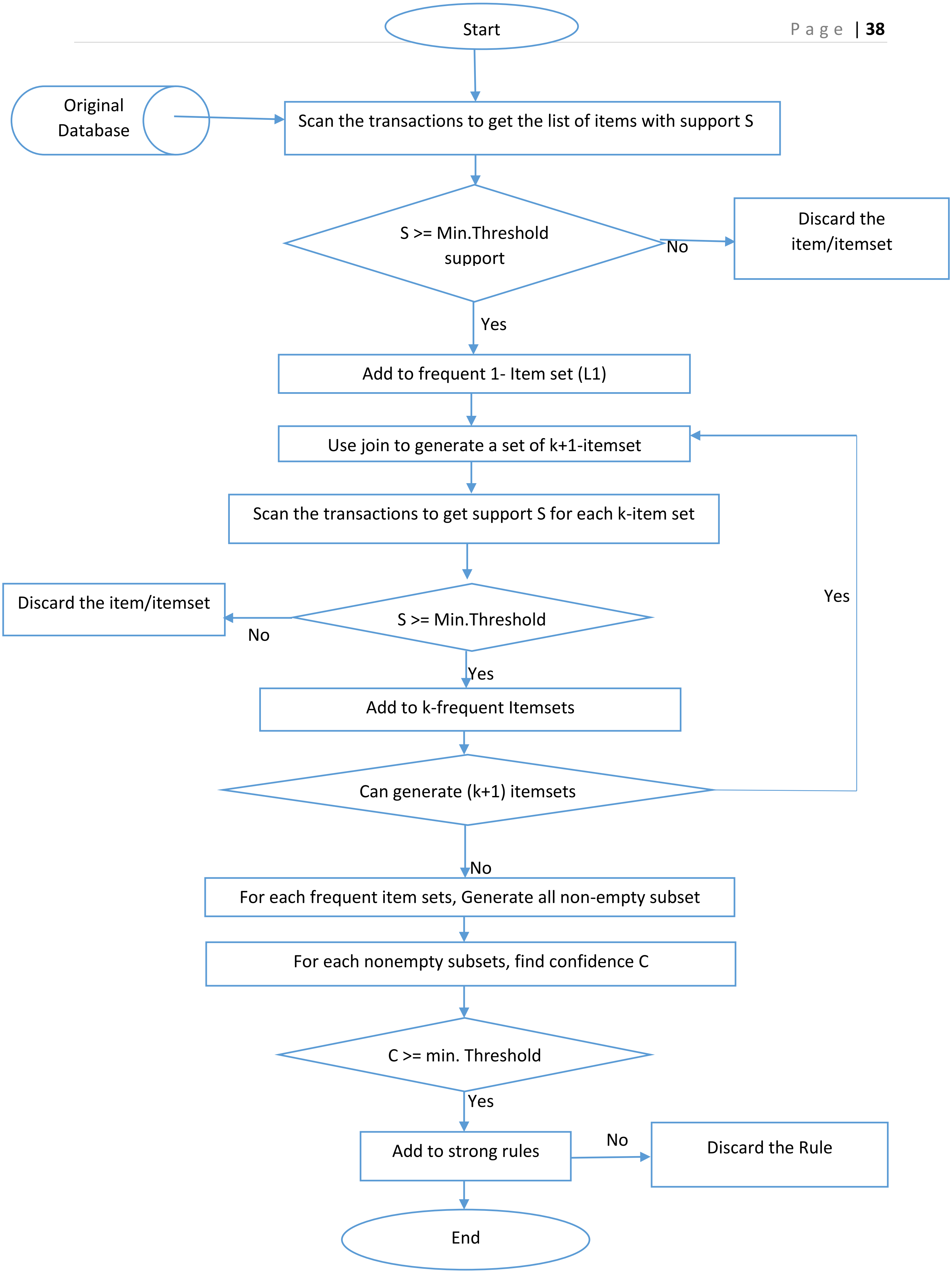
supp({School=RCCIIT Second Class Test=Present First Class Test=Absence }) / supp({First Class Test=Absence }) x supp({School=RCCIIT Second Class Test=Present })= 0.324/0.81 x 0.33= 1.21

Leverage:

Leverage is the proportion of additional elements covered by both the premise and consequence above the expected if independent.

$$\text{lev}(X \rightarrow Y) = \text{supp}(X \cup Y) - \text{sup}(X).\ \text{supp}(Y)$$

lev({ School=RCCIIT Second Class Test=Present }=>{ First Class Test=Absence }) = supp(School=RCCIIT Second Class Test=Present First Class Test=Absence) − sup(School=RCCIIT Second Class Test=Present). supp(First Class Test=Absence) = 0.324-(0.81x0.33)= 0.06


Start
Original Database
Scan the transactions to get the list of items with support S
S >= Min.Threshold support
No
Discard the item/itemset
Yes
Add to frequent 1- Item set (L1)
Use join to generate a set of k+1-itemset
Scan the transactions to get support S for each k-item set
Discard the item/itemset
S >= Min.Threshold
No
Yes
Yes
Add to k-frequent Itemsets
Can generate (k+1) itemsets
No
For each frequent item sets, Generate all non-empty subset
For each nonempty subsets, find confidence C
C >= min. Threshold
Yes
Add to strong rules
No
Discard the Rule
End

Approximate Association Rule Mining

The goal of this research is to develop an association rule algorithm that accepts partial support from data. By generating these "approximate" rules, data can contribute to the discovery despite the presence of noisy or missing values.

The approximate association rule algorithm, called ~AR, is built upon the Apriori algorithm and uses two main steps to handle missing and noisy data. First, missing values are replaced with a probability distribution over possible values represented by existing data. Second, all data contributes probabilistically to candidate patterns. Patterns which receive a sufficient amount of full or partial support are kept and expanded. To demonstrate the capabilities of ~AR, we incorporate the algorithm into the Weka implementation of Apriori. Results are shown on several sample databases.

our approach to approximate association rule mining is embodied in the ~AR algorithm. The ~AR algorithm represents an enhancement of the Apriori algorithm included as part of the Weka of data mining tools [Weka]. The Weka algorithms, including the basic Apriori algorithm, are written in Java and include a uniform interface. The first step of the ~AR algorithm is to impute missing values. Each missing value is replaced by a probability distribution. In order to adopt this approach, we make the assumption that fields are named or ordered consistently between data entries. This probability distribution represents the likelihood of possible values for the missing data calculated using frequency counts from the entries that do contain data for the corresponding field.

For example, consider a database that contains the following transactions, where"**?**" represents a missing value.

A, B, C
E,F,E
?, B, E
A,B,F

The missing value is replaced by a probability distribution calculated using the existing data. In this case, the probability that the value is "A" is P(A) = 0.67, and the probability the value is "E"is P(E)= 0.33. The second step of the ~AR algorithm is to discover the association rules. The main difference between ~AR and the Apriori algorithm is in the calculation of support for a candidate item set. In the Apriori algorithm, a transaction supports a pattern if the transaction includes precise matches for all of the items in the candidate item set. In contrast, ~AR allows transactions to partially support a candidate pattern. Two types of inexact match may occur. In the first case, the transaction entry exists but does not match the corresponding entry in the candidate item set. In this case, the support is incremented according to the similarity of the two values. In this case of nominal attributes, the difference is maximal and support is not incremented. In the case of numeric values, the support is incremented by the absolute value of the difference between the value, divided by the maximum possible value for the given item.
Consider a candidate itemset containing four items:

C=A,B,C,D

A database transaction may exist that fully matches the candidate itemsct:

T1 = A, B,C, D

In this example, support for candidate C is incremented by 1/4+1/4 +1/4+1/4=1

Using the second example, the transaction does not completely match the candidate item set:

T2 =A,E,C,D

Support for candidate C is incremented based on transaction T2 by 1/4 + 0 + 1/4+ 1/4 = 3/4.

The second type of inexact match considers a missing value which has been replaced by a probability distribution, and is considered for possible support of a candidate item set. The probability that the missing value corresponds to the value in the candidate item set For example, if the transaction is:

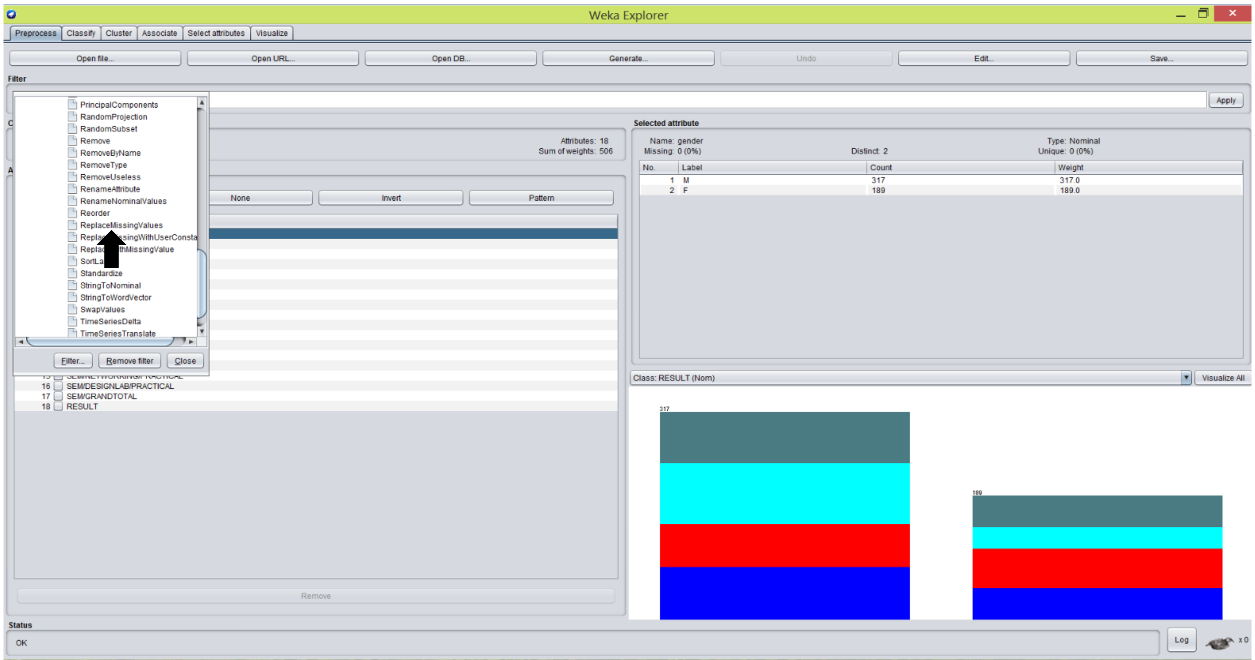
T3 = A, B,?, D and P(C =1/4), P(E = 1//4),P(F = 1/2)

Support for candidate item set C is incremented by 1/4+ (1/4*1/2) +1/4 = 7/8.

One danger with this approach is that every transaction can potentially support every candidate item set. To prevent transactions from supporting patterns that differ greatly from the transaction, a minimum match threshold is set by the user. If the support provided by any transaction falls below the threshold, then the transaction does not contribute any support to the candidate pattern.

In the Weka implementation of the ~AR algorithm, a minimum support threshold of 90% is initially specified. Multiple iterations of the discovery algorithm are executed until at least N item sets are discovered with the user specified minimum confidence, or until the user-specified minimum support level is reached. The ~AR algorithm is composed of 3 steps, First, all of the transactions are read from a database stored in the ARFF format. Second, item sets are generated that meet the support and confidence thresholds. Finally, all possible rules are generated from the large item sets.

```
@relation DEi

@attribute gender {M,F}
@attribute PlaceofBirth {Bengal,Mumbai,Delhi,Bihar,Pune}
@attribute School {RCCIIT,WBUT,IEM}
@attribute SectionID {A,B,C}
@attribute 'First Class Test' {Absence,Present}
@attribute 'Second Class Test' {Present,Absence}
@attribute SEM/C {0-25,26-50,51-75,76-100}
@attribute SEM/ALGORITHM {0-25,26-50,51-75,76-100}
@attribute SEM/DBMS {0-25,26-50,51-75,76-100,57-67}
@attribute SEM/DATASTRUCTURE {0-25,26-50,51-75,76-100,25-67}
@attribute SEM/NETWORKING {0-25,26-50,51-75,76-100}
@attribute SEM/HU/PRACTICAL {26-50,51-75,76-100,0-25}
@attribute SEM/DBMS/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/C/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/NETWORKING/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/DESIGNLAB/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/GRANDTOTAL {0-250,501-750,751-1000,251-500}
@attribute RESULT {FAIL,GOOD,EXCELLENT,OUTSTANDING}

@data
M,Bengal,RCCIIT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,?,0-25,0-25,0-25,0-25,0-250,FAIL
M,Bengal,RCCIIT,A,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,?,GOOD
M,Bengal,RCCIIT,A,Absence,Present,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,501-750,EXCELLENT
M,Bengal,RCCIIT,A,Absence,Present,?,?,?,?,?,76-100,76-100,76-100,76-100,76-100,751-1000,OUTSTANDING
M,Bengal,RCCIIT,A,Absence,Present,?,?,?,?,?,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
F,Bengal,RCCIIT,A,Absence,Present,?,?,?,?,?,26-50,26-50,?,?,26-50,?,GOOD
M,Bengal,WBUT,A,Absence,Present,?,?,?,?,?,51-75,51-75,?,?,51-75,501-750,EXCELLENT
M,Bengal,WBUT,A,Absence,Present,?,?,?,?,?,76-100,76-100,?,?,76-100,751-1000,OUTSTANDING
F,Bengal,WBUT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
F,Bengal,WBUT,B,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,?,GOOD
M,Bengal,WBUT,A,Absence,Present,51-75,?,?,51-75,51-75,51-75,51-75,51-75,51-75,51-75,501-750,EXCELLENT
M,Bengal,WBUT,B,Absence,Present,76-100,?,?,76-100,76-100,76-100,76-100,76-100,76-100,76-100,751-1000,OUTSTANDING
M,Bengal,RCCIIT,A,Absence,Present,0-25,?,?,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
M,Bengal,WBUT,A,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,251-500,GOOD
F,Bengal,WBUT,A,Absence,Absence,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,?,EXCELLENT
F,Bengal,WBUT,A,Absence,Present,76-100,76-100,76-100,76-100,76-100,?,?,?,?,76-100,751-1000,OUTSTANDING
M,Bengal,WBUT,B,Absence,Present,0-25,0-25,0-25,0-25,0-25,?,?,?,?,0-25,0-250,FAIL
M,Bengal,WBUT,A,Absence,Present,26-50,26-50,26-50,26-50,26-50,?,?,?,?,26-50,?,GOOD
F,Bengal,WBUT,A,Absence,Absence,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,501-750,EXCELLENT
M,Bengal,WBUT,B,Absence,Absence,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,751-1000,OUTSTANDING
F,Bengal,WBUT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
F,Bengal,WBUT,B,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,?,GOOD
```



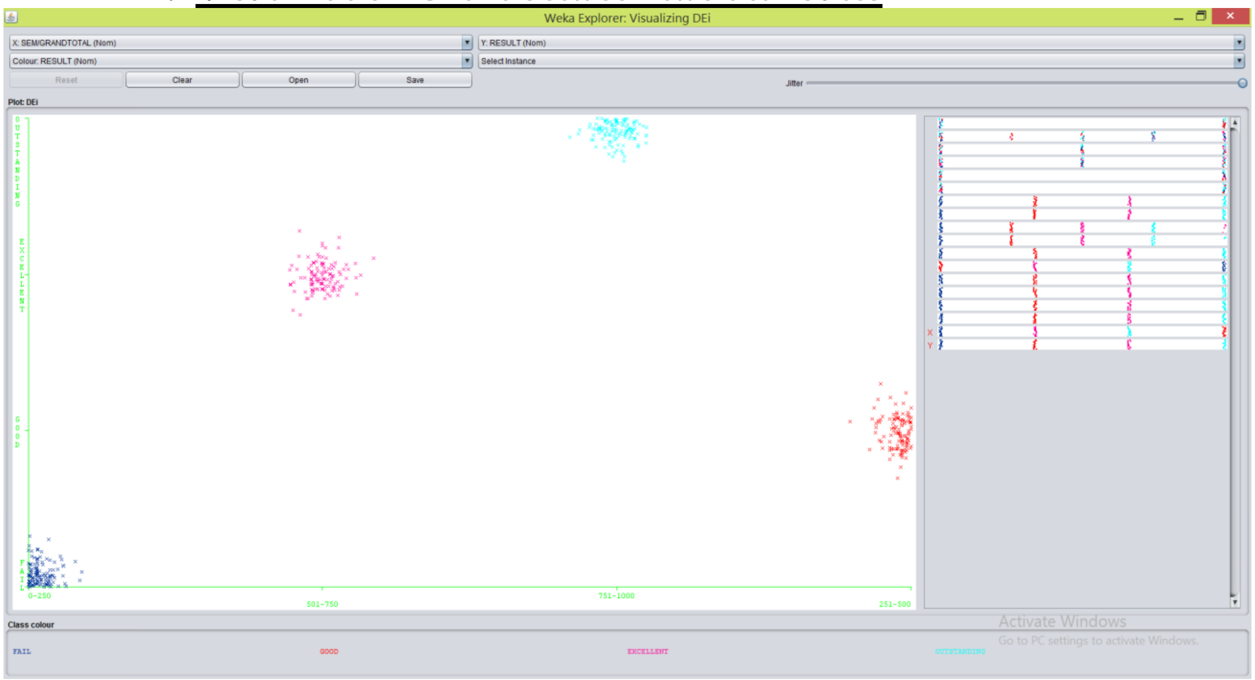
After apply filter “**replace missing values**”, generated a modified .arff data formet, which can be used for Approximate association rule generation and in here, more approximate no of missing rules will be generated as same as normal rule generation.

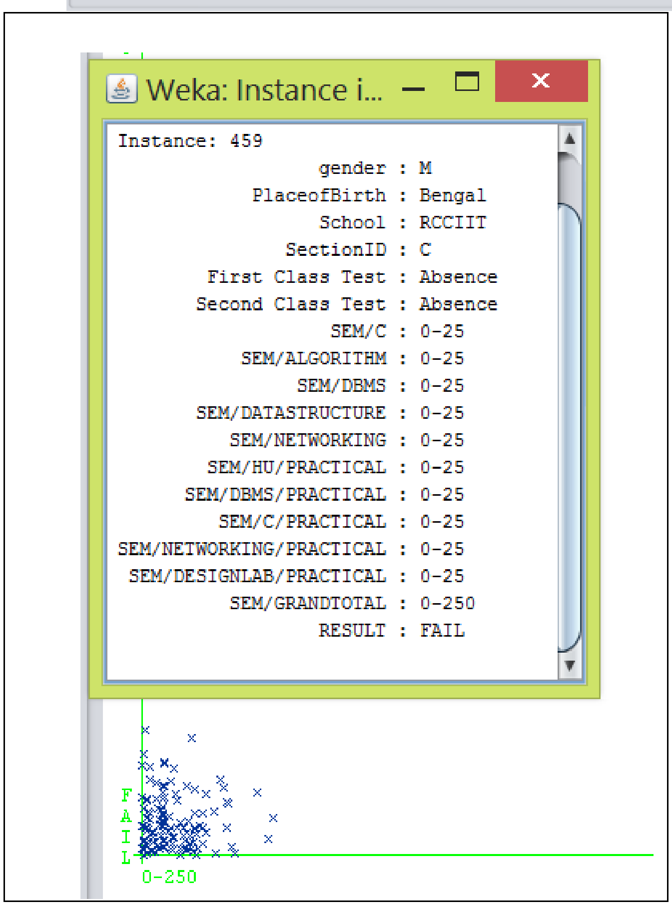
Modified data format is shown below,

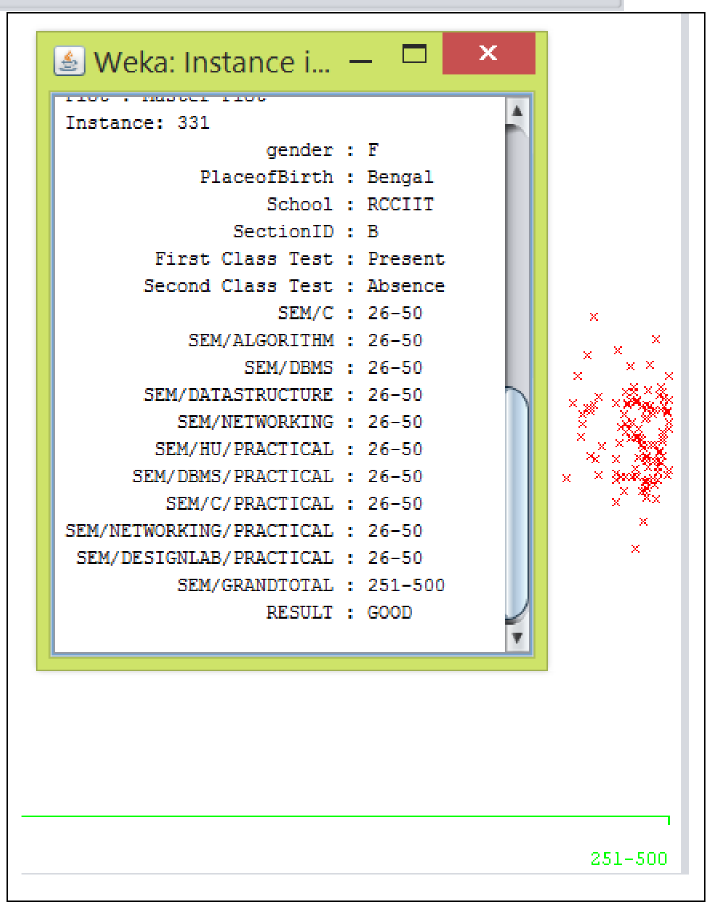
```
@relation DEi-weka.filters.unsupervised.attribute.ReplaceMissingValues

@attribute gender {M,F}
@attribute PlaceofBirth {Bengal,Mumbai,Delhi,Bihar,Pune}
@attribute School {RCCIIT,WBUT,IEM}
@attribute SectionID {A,B,C}
@attribute 'First Class Test' {Absence,Present}
@attribute 'Second Class Test' {Present,Absence}
@attribute SEM/C {0-25,26-50,51-75,76-100}
@attribute SEM/ALGORITHM {0-25,26-50,51-75,76-100}
@attribute SEM/DBMS {0-25,26-50,51-75,76-100,57-67}
@attribute SEM/DATASTRUCTURE {0-25,26-50,51-75,76-100,25-67}
@attribute SEM/NETWORKING {0-25,26-50,51-75,76-100}
@attribute SEM/HU/PRACTICAL {26-50,51-75,76-100,0-25}
@attribute SEM/DBMS/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/C/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/NETWORKING/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/DESIGNLAB/PRACTICAL {0-25,26-50,51-75,76-100}
@attribute SEM/GRANDTOTAL {0-250,501-750,751-1000,251-500}
@attribute RESULT {FAIL,GOOD,EXCELLENT,OUTSTANDING}

@data
M,Bengal,RCCIIT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
M,Bengal,RCCIIT,A,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,0-250,GOOD
M,Bengal,RCCIIT,A,Absence,Present,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,501-750,EXCELLENT
M,Bengal,RCCIIT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,76-100,76-100,76-100,76-100,76-100,751-1000,OUTSTANDING
M,Bengal,RCCIIT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
F,Bengal,RCCIIT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,26-50,26-50,0-25,0-25,26-50,0-250,GOOD
M,Bengal,WBUT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,51-75,51-75,0-25,0-25,51-75,501-750,EXCELLENT
M,Bengal,WBUT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,76-100,76-100,0-25,0-25,76-100,751-1000,OUTSTANDING
F,Bengal,WBUT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
F,Bengal,WBUT,B,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,0-250,GOOD
M,Bengal,WBUT,A,Absence,Present,51-75,0-25,0-25,51-75,51-75,51-75,51-75,51-75,51-75,51-75,501-750,EXCELLENT
M,Bengal,WBUT,B,Absence,Present,76-100,0-25,0-25,76-100,76-100,76-100,76-100,76-100,76-100,76-100,751-1000,OUTSTANDING
M,Bengal,RCCIIT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
M,Bengal,WBUT,A,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,251-500,GOOD
F,Bengal,WBUT,A,Absence,Absence,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,0-250,EXCELLENT
F,Bengal,WBUT,A,Absence,Present,76-100,76-100,76-100,76-100,76-100,0-25,0-25,0-25,0-25,76-100,751-1000,OUTSTANDING
M,Bengal,WBUT,B,Absence,Present,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
M,Bengal,WBUT,A,Absence,Present,26-50,26-50,26-50,26-50,26-50,0-25,0-25,0-25,0-25,26-50,0-250,GOOD
F,Bengal,WBUT,A,Absence,Absence,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,501-750,EXCELLENT
M,Bengal,WBUT,B,Absence,Absence,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,76-100,751-1000,OUTSTANDING
F,Bengal,WBUT,A,Absence,Present,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
F,Bengal,WBUT,B,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,0-250,GOOD
```

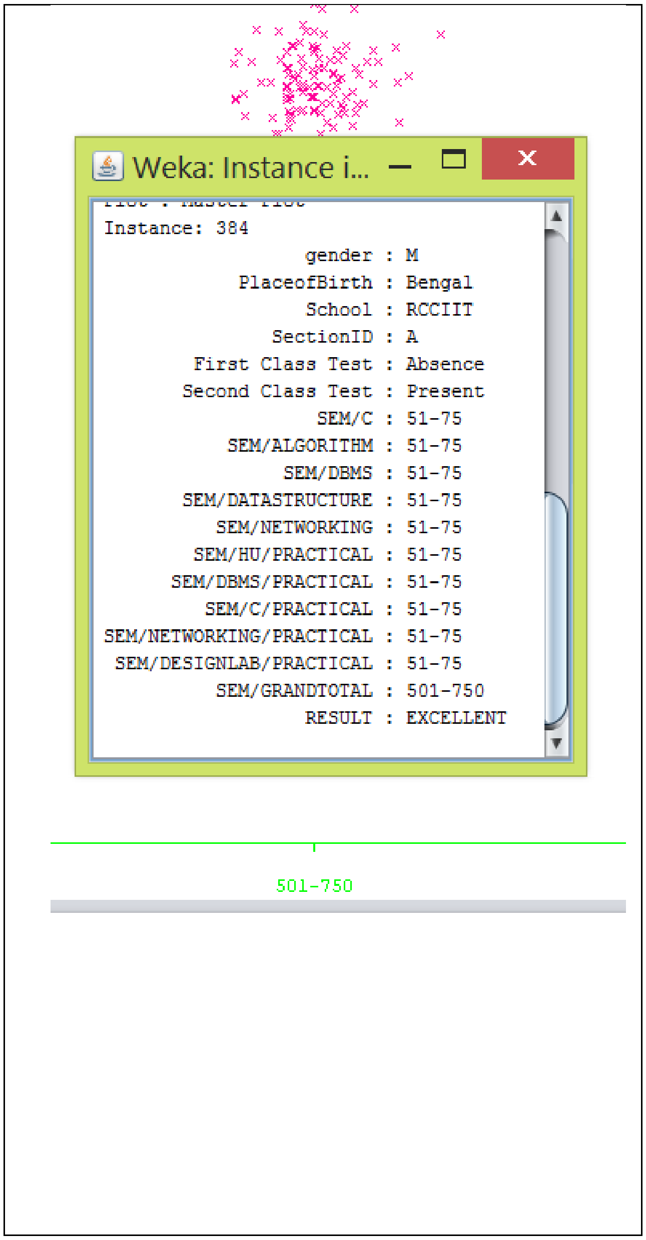
Visualization:

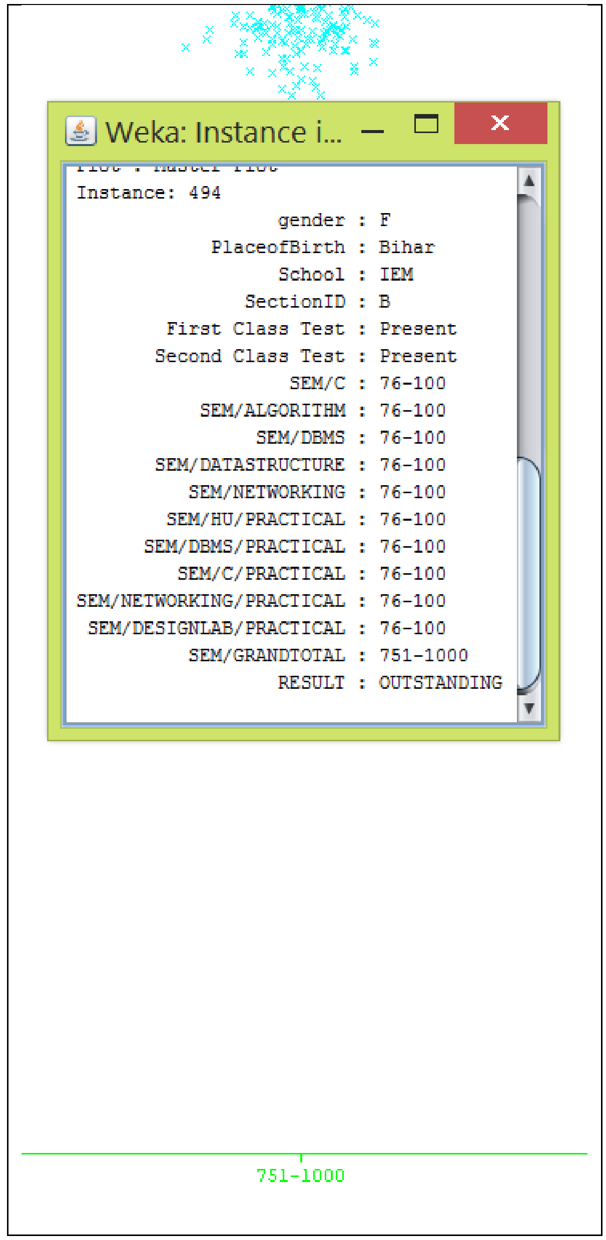
1. Visualization Chart between two attributes







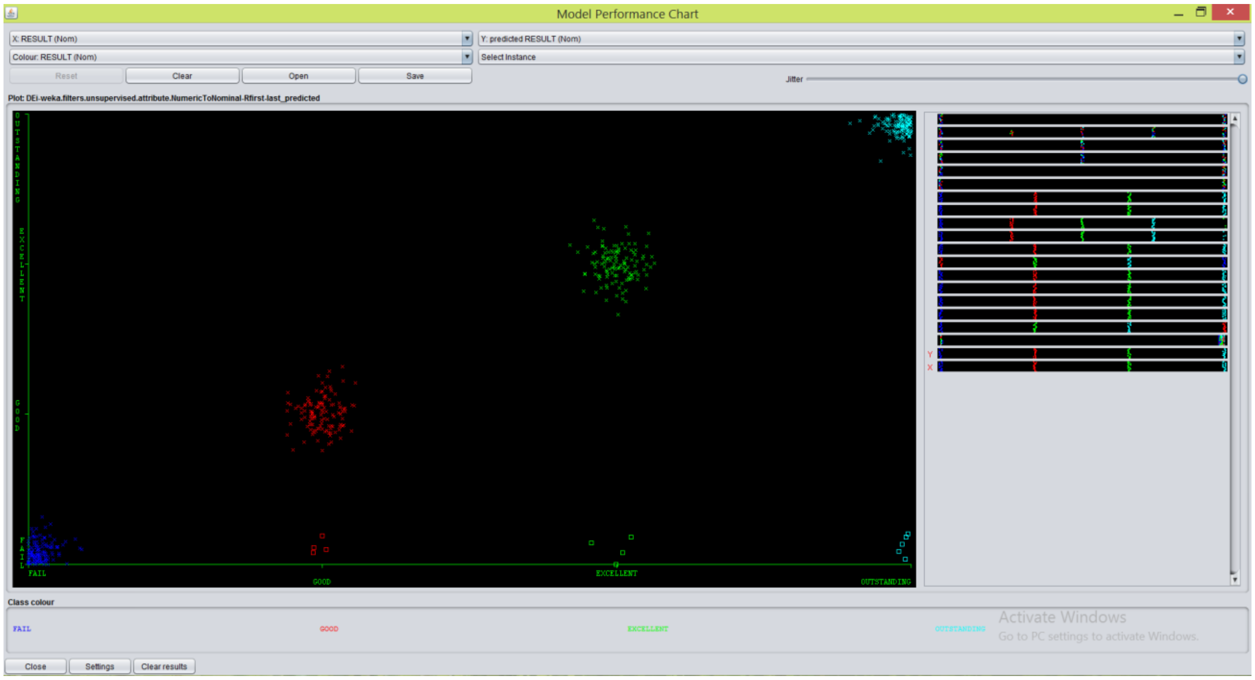


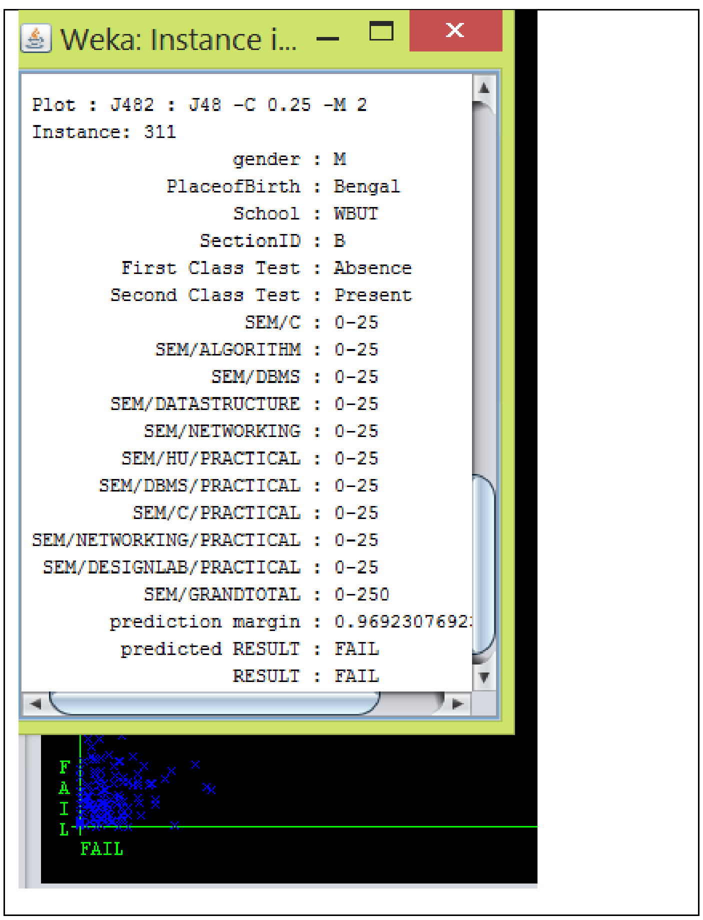


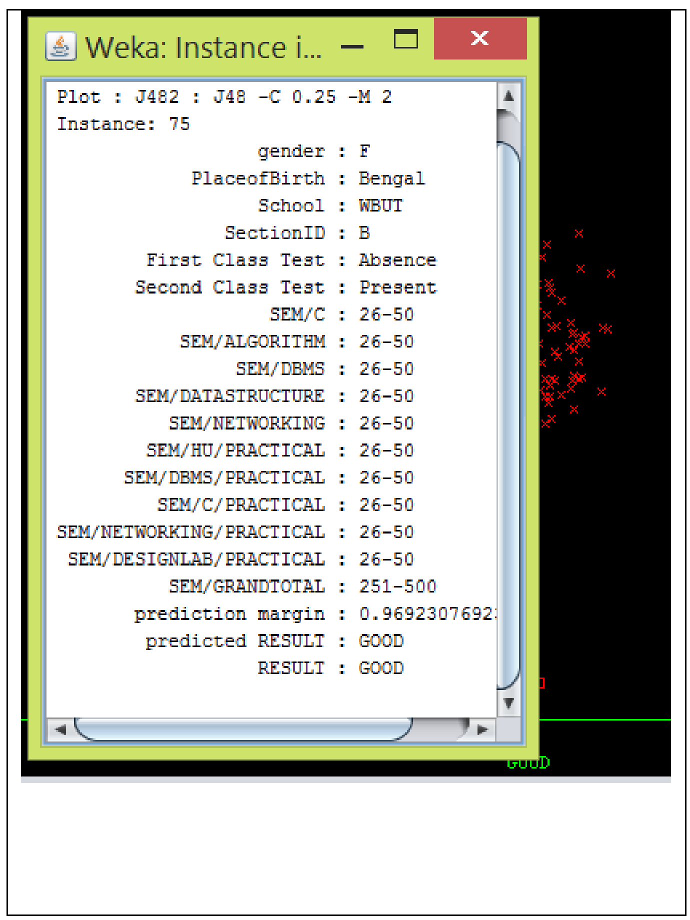
This visualization portion based on two attributes, when SEM/GRANDTOTAL along with X axis corresponding to RESULT along with Y axis.

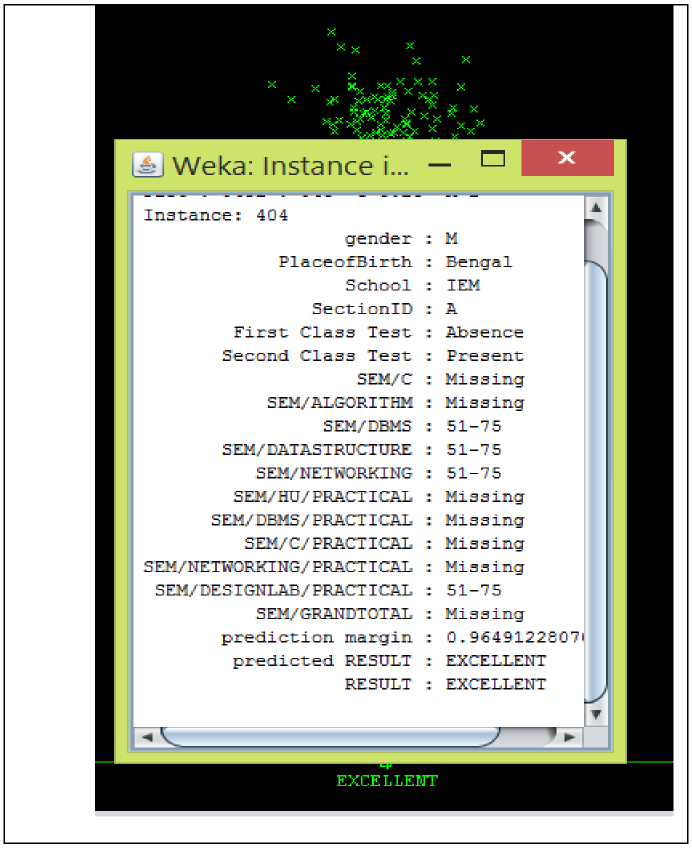
But, there can be more no of visualization portion based on other two attributes. But we have shown only one visualization chart.

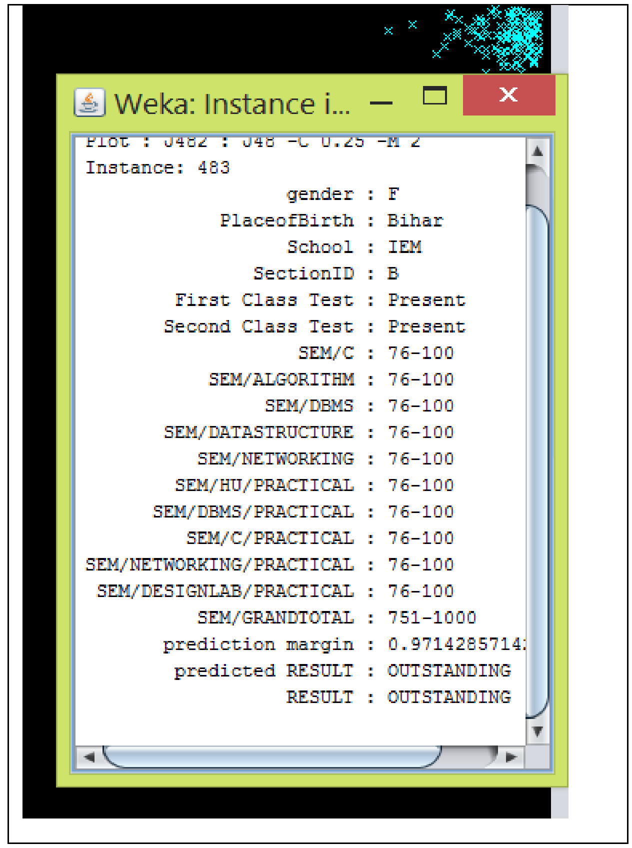
2. Visualize Classifier errors chart between predicted result and actual result

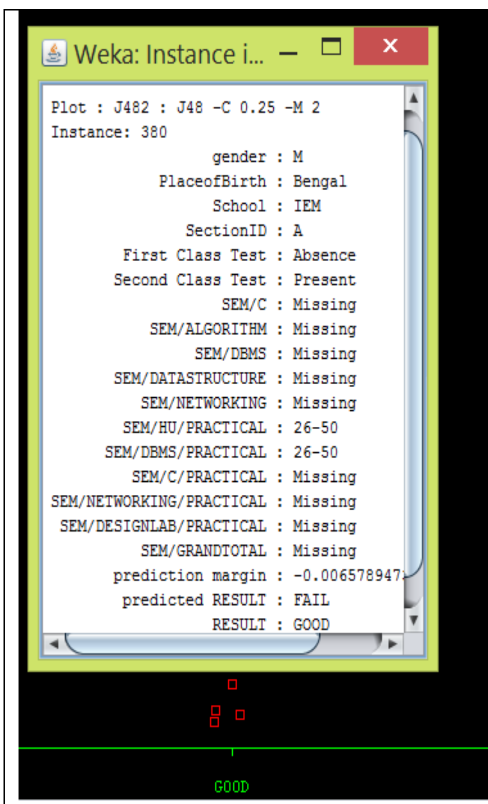


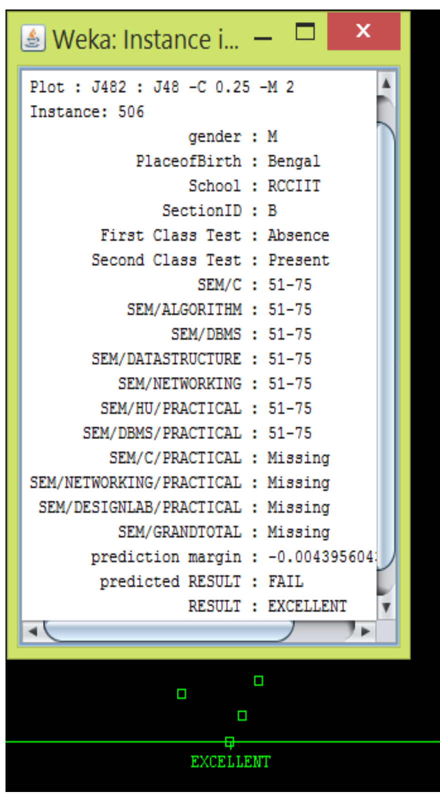


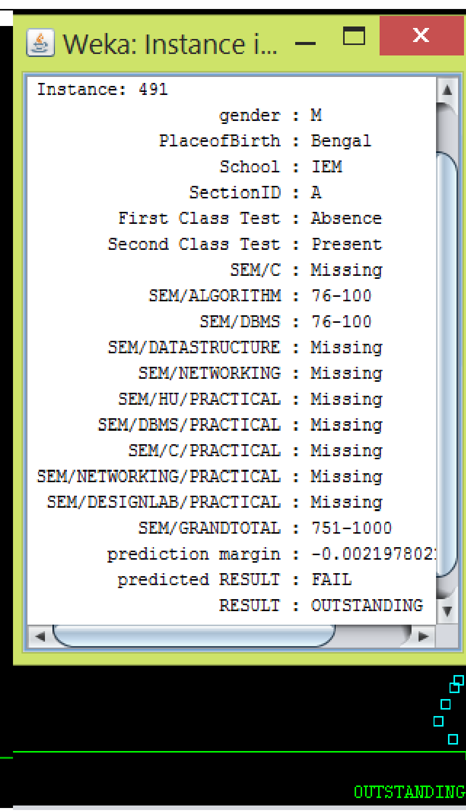



Weka: Instance i...
Instance: 404
gender : M
PlaceofBirth : Bengal
School : IEM
SectionID : A
First Class Test : Absence
Second Class Test : Present
SEM/C : Missing
SEM/ALGORITHM : Missing
SEM/DBMS : 51-75
SEM/DATASTRUCTURE : 51-75
SEM/NETWORKING : 51-75
SEM/HU/PRACTICAL : Missing
SEM/DBMS/PRACTICAL : Missing
SEM/C/PRACTICAL : Missing
SEM/NETWORKING/PRACTICAL : Missing
SEM/DESIGNLAB/PRACTICAL : 51-75
SEM/GRANDTOTAL : Missing
prediction margin : 0.9649122807
predicted RESULT : EXCELLENT
RESULT : EXCELLENT
EXCELLENT


Weka: Instance i...
Instance: 483
gender : F
PlaceofBirth : Bihar
School : IEM
SectionID : B
First Class Test : Present
Second Class Test : Present
SEM/C : 76-100
SEM/ALGORITHM : 76-100
SEM/DBMS : 76-100
SEM/DATASTRUCTURE : 76-100
SEM/NETWORKING : 76-100
SEM/HU/PRACTICAL : 76-100
SEM/DBMS/PRACTICAL : 76-100
SEM/C/PRACTICAL : 76-100
SEM/NETWORKING/PRACTICAL : 76-100
SEM/DESIGNLAB/PRACTICAL : 76-100
SEM/GRANDTOTAL : 751-1000
prediction margin : 0.9714285714
predicted RESULT : OUTSTANDING
RESULT : OUTSTANDING


Weka: Instance i...
Plot : J482 : J48 -C 0.25 -M 2
Instance: 380
gender : M
PlaceofBirth : Bengal
School : IEM
SectionID : A
First Class Test : Absence
Second Class Test : Present
SEM/C : Missing
SEM/ALGORITHM : Missing
SEM/DBMS : Missing
SEM/DATASTRUCTURE : Missing
SEM/NETWORKING : Missing
SEM/HU/PRACTICAL : 26-50
SEM/DBMS/PRACTICAL : 26-50
SEM/C/PRACTICAL : Missing
SEM/NETWORKING/PRACTICAL : Missing
SEM/DESIGNLAB/PRACTICAL : Missing
SEM/GRANDTOTAL : Missing
prediction margin : -0.006578947
predicted RESULT : FAIL
RESULT : GOOD
GOOD


Weka: Instance i...
Plot : J482 : J48 -C 0.25 -M 2
Instance: 506
gender : M
PlaceofBirth : Bengal
School : RCCIIT
SectionID : B
First Class Test : Absence
Second Class Test : Present
SEM/C : 51-75
SEM/ALGORITHM : 51-75
SEM/DBMS : 51-75
SEM/DATASTRUCTURE : 51-75
SEM/NETWORKING : 51-75
SEM/HU/PRACTICAL : 51-75
SEM/DBMS/PRACTICAL : 51-75
SEM/C/PRACTICAL : Missing
SEM/NETWORKING/PRACTICAL : Missing
SEM/DESIGNLAB/PRACTICAL : Missing
SEM/GRANDTOTAL : Missing
prediction margin : -0.004395604
predicted RESULT : FAIL
RESULT : EXCELLENT
EXCELLENT


Weka: Instance i...
Instance: 491
gender : M
PlaceofBirth : Bengal
School : IEM
SectionID : A
First Class Test : Absence
Second Class Test : Present
SEM/C : Missing
SEM/ALGORITHM : 76-100
SEM/DBMS : 76-100
SEM/DATASTRUCTURE : Missing
SEM/NETWORKING : Missing
SEM/HU/PRACTICAL : Missing
SEM/DBMS/PRACTICAL : Missing
SEM/C/PRACTICAL : Missing
SEM/NETWORKING/PRACTICAL : Missing
SEM/DESIGNLAB/PRACTICAL : Missing
SEM/GRANDTOTAL : 751-1000
prediction margin : -0.002197802
predicted RESULT : FAIL
RESULT : OUTSTANDING
OUTSTANDING

4.1 Result Analysis

Knowledge Discovery Database (KDD)

The KDD (Knowledge Discovery in Databases) paradigm is a step by step process for finding interesting patterns in large amounts of data. Data mining is one step in the process. The algorithms' potential as good analytical tools for performance evaluation is shown by looking at results from a computer performance dataset.
It is much easier to store data than it is to make sense of it. Being able to find relationships in large amounts of stored data can lead to enhanced analysis strategies in fields such as educational, marketing, computer performance analysis, and data analysis in general. The problem addressed by KDD is to find patterns in these massive datasets. Traditionally data has been analyzed manually, but there are human limits. Large databases offer too much data to analyze in the traditional manner. The focus of this paper is to first summarize exactly what the KDD process is.

Procedure of prediction and analysis of Students performance using KDD:

1. After completing all part test (preprocess, classification, filter, association and visualization), we are going to show the final accumulate structure of student performance by using KDD process.
2. If you choose ARFF file in your experiment then select ARFF loader or, you choose csv file in your experiment then select on csv loader .we take csv file.
3. Click on csv loader and paste it on screen, then pass the dataset to next position.
4. For transfer numerical value to nominal value into csv file, use the intermediate filter “numeric to nominal” .Then pass the dataset to next position.
5. To classify the file need some intermediate evaluation-
    1. Class assigner (to assign the class)
    2. Cross validation fold maker(to show
    3. Training Set Maker (to train the dataset for prediction)

    -by passing the dataset both of these three parts.
6. After that, need to choose a standard classifier to classify & prediction result of the given dataset by test set and training set. Take classifier like- J48.
7. Then connect the classifier with some intermediate evaluation

    I. classifier performance evaluator (use for getting some important parameter result)
    II. prediction appender (use for getting predicted result)

    - Both connect by batch classifier from J48 classifier.

To show the result use text viewer ,by connect with text ,means to get parameters result that means, confusion matrix , accuracy ,TP rate, FP rate, precision ,recall and so on & also predicted result along with actual result.

8. (i) To show the generating graph or image by the classifier, need a graph viewer by passing graphs signal.

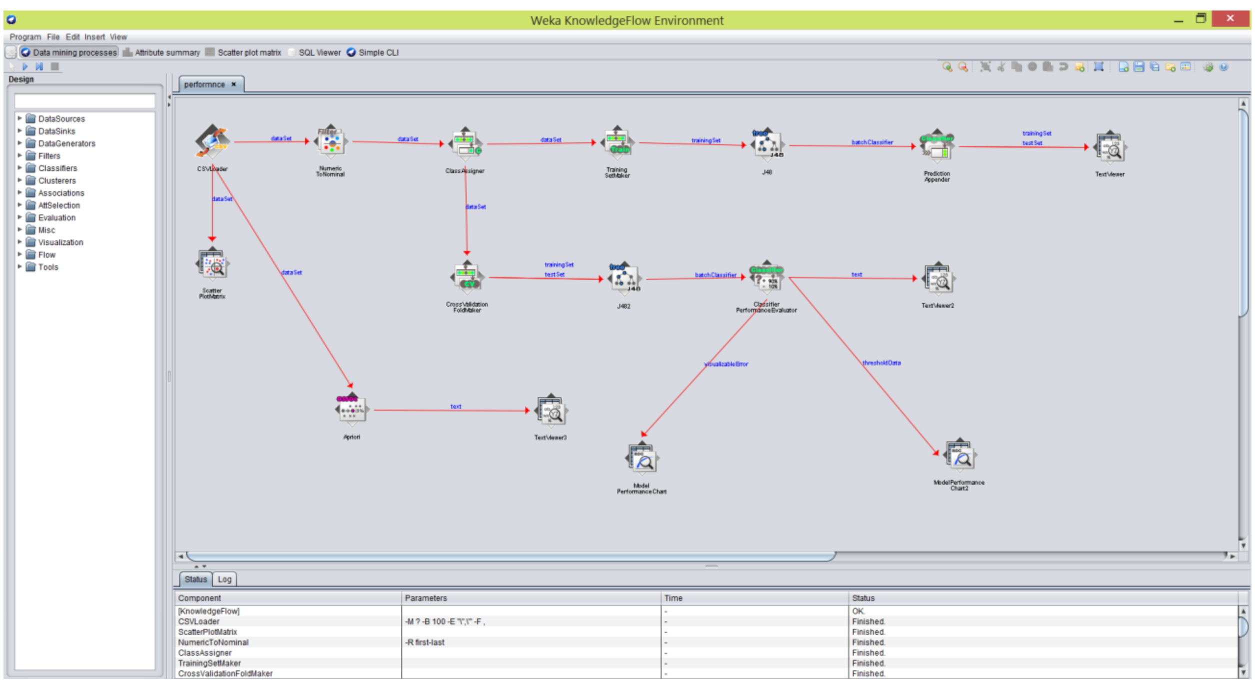
(ii) There will be needed a visualization tool which is model performance chart by passing the threshold data for getting some chart between classifier parameters and by using visualizable error signal for getting some chart between error points between attributes(like- actual result vs predicted result).

9. After completing Classification Stage, will go to Association Stage for generating association rules by using Apriori algorithm, passing the dataset from loader to see the result of rules, need a text viewer for showing the output by using text signal.
10. (i) At last, for visualization of the dataset need to choose Scatter Plot Matrix tool by passing the dataset from loader.

(ii)After completing diagrams, need to load the data in the csv loader portion or tools and Click on run at the left top portion then look on bottom status portion for checking success or errors point

[All the signal can be found by click on those input tools are getting from the left side of the WEKA knowledge Environment {version 3.8.1}]

Finally we create a Knowledge Discovery Database of Student performance based and we got a KDD as same result as weka classifier, associater, visualizer or others result. This KDD diagram is running successfully and got a pattern.
KDD diagram is shown below-

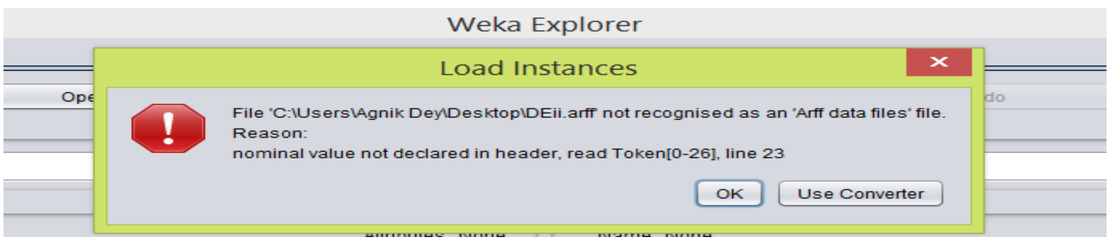


4.2 WEKA Limitation

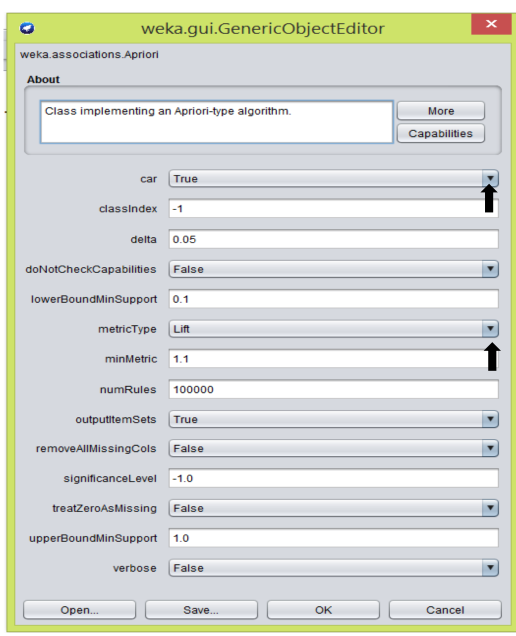
There have some limitation in WEKA. Those are explained below-

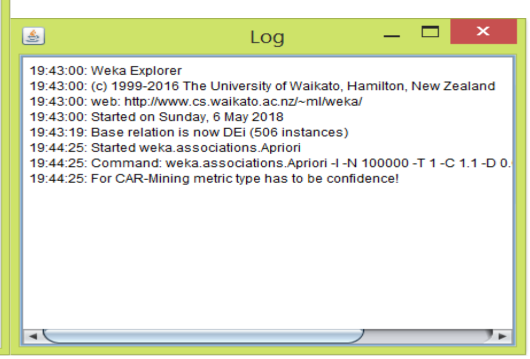
1. In WEKA, when we have to declare any item set values in an attribute portion, then only those items will be used in creating of data format. If we are not given those similar item set, then WEKA show an error pop-up message because WEKA does not support any undeclared numerical or string value.

```
1  @relation DEi
2
3  @attribute gender {M,F}
4  @attribute PlaceofBirth {Bengal,Mumbai,Delhi,Bihar,Pune}
5  @attribute School {RCCIIT,WBUT,IEM}
6  @attribute SectionID {A,B,C}
7  @attribute 'First Class Test' {Absence,Present}
8  @attribute 'Second Class Test' {Present,Absence}
9  @attribute SEM/C {0-25,26-50,51-75,76-100}
10 @attribute SEM/ALGORITHM {0-25,26-50,51-75,76-100}
11 @attribute SEM/DBMS {0-25,26-50,51-75,76-100,57-67}
12 @attribute SEM/DATASTRUCTURE {0-25,26-50,51-75,76-100,25-67}
13 @attribute SEM/NETWORKING {0-25,26-50,51-75,76-100}
14 @attribute SEM/HU/PRACTICAL {26-50,51-75,76-100,0-25}
15 @attribute SEM/DBMS/PRACTICAL {0-25,26-50,51-75,76-100}
16 @attribute SEM/C/PRACTICAL {0-25,26-50,51-75,76-100}
17 @attribute SEM/NETWORKING/PRACTICAL {0-25,26-50,51-75,76-100}
18 @attribute SEM/DESIGNLAB/PRACTICAL {0-25,26-50,51-75,76-100}
19 @attribute SEM/GRANDTOTAL {0-250,501-750,751-1000,251-500}
20 @attribute RESULT {FAIL,GOOD,EXCELLENT,OUTSTANDING}
21
22 @data
23 M,Bengal,RCCIIT,A,Absence,Present,0-25,0-26,0-25,0-25,0-25,?,0-25,0-25,0-25,0-25,0-250,FAIL
24 M,Bengal,RCCIIT,A,Absence,Present,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,26-50,?,PASS
25 M,Bengal,RCCIIT,A,Absence,Present,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,51-75,501-750,EXCELLENT
26 M,Bengal,RCCIIT,A,Absence,Present,?,?,?,?,?,76-100,76-100,76-100,76-100,76-100,751-1000,OUTSTANDING
27 M,Bengal,RCCIIT,A,Absence,Present,?,?,?,?,?,0-25,0-25,0-25,0-25,0-25,0-250,FAIL
28 F,Bengal,RCCIIT,A,Absence,Present,?,?,?,?,?,26-50,26-50,?,?,26-50,?,GOOD
29 M,Bengal,WBUT,A,Absence,Present,?,?,?,?,?,51-75,51-75,?,?,51-75,501-750,EXCELLENT
30 M,Bengal,WBUT,A,Absence,Present,?,?,?,?,?,76-100,76-100,?,?,76-100,751-1000,OUTSTANDING
```



2. In WEKA associate, class cannot be generated by using lift/others without confidence.





3. WEKA cannot give result in sorting order format that means it creates all rules which is above minimum support and minimum confidence looking from L(2) to L(n) by first in process. So, it is not possible to separate largest rules or meaningful rule exactly. That's why, we have to find exact meaningful rule by checking all possible largest rule from huge amount of rules.

4. WEKA cannot generate individually any student's performance, we can find only frequently possibility of overall students' performance. From here, teachers can judge, from next time what type of students will be going to get good remarks in his absence.

5. WEKA cannot generate some other type of interestingness measurement result (like- Certainty Factor, Relative Risk, Cosine, Information Gain, $\emptyset-$Coefficient or etc.) along with a particular rules. For getting this, we want to calculate result manually by using some formula.

5 . Conclusion and future work

This paper presents data mining in education environment that identifies students' failure patterns using association rule mining technique. The identified patterns are analyzed to offer a helpful and constructive recommendations to the academic planners in higher institutions of learning to enhance their decision making process. Association rule mining has been applied to Education systems for analysis of student result. In this research, the association rule mining technique is used to find hidden patterns and evaluate students' performance and trends. Apriori algorithm is used for finding associations among attributes.

The students' academic performance was evaluated based on academic and personal data collected from college's last semester result. After that J48 classification algorithms were used. The data mining tool used in the experiment was WEKA 3.8.2. Based on the accuracy and the classification errors one may conclude that the J48 Classification method was the most suited algorithm for the dataset. The Apriori algorithm was applied to the dataset using WEKA to find analysis of overall student performance by some of the best rules. The data may be extended to collect some of the extra-curricular aspects and technical skills of the students and mined with different classification algorithms to predict the student performance.

In future work the authors also interested in working in future on data of students assessments for each course trying to know what kind of student succeed on what kind of courses. It may define what kinds of courses are adapted for every student's model who shares the same characteristics. It may also provide various multidimensional summary reports and redefine pedagogical learning paths.

6 .References/Bibliography

[1] Samrat Singh, Dr. Vikesh Kumar , "Performance Analysis of Engineering Students for Recruitment Using Classification Data Mining Techniques ",IJCSET February 2013.

[2] M. Goyal and R. Vohra, “Applications of Data Mining in Higher Education”, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue2, No 1, March 2012.

[3] Jason Brownlee ,"How to Save Your Machine Learning Model and Make Predictions in Weka", August 3, 2016.

[4] Neelam Naik & Seema Purohit, “Prediction of Final Result and Placement of Students using Classification Algorithm”International Journal of Computer Applications (0975 – 8887) Volume 56– No.12, October 2012

[5] Alaa M.El-Halees,Mohammed M. Abu Tair, “Mining Educational Data to Improve Students’Performance: A Case Study”,International Journal of Information and Communication Technology Research, 2012.

[6] B.K. Bharadwaj and S. Pal,“Data Mining: A prediction for performance improvement using classification”, International Journal of Computer Science and Information Security (IJCSIS), Vol. 9, No. 4, pp. 136-140, 2011.

[7] Suchita Borkar, K. Rajeswari, "Predicting Students Academic Performance Using Education Data Mining ", IJCSMC,Vol. 2, Issue. 7, July 2013, pg.273– 279.

[8] Randhir Singh, M.Tiwari, Neeraj Vimal,"An Empirical Study of Applications of Data Mining Techniques for Predicting Student Performance in Higher Education", 2013.

[9] D.Magdalene Delighta Angeline,"Association Rule Generation for Student Performance Analysis using Apriori Algorithm",The SIJ Transactions on Computer Science Engineering & its Applications (CSEA), Vol. 1, No. 1, March-April 2013

[10] Mrs. M.S. Mythili, Dr. A.R.Mohamed Shanavas,"An Analysis of students’ performance using classification algorithms ",ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 1, Ver .III (Jan. 2014), PP 63-69

[11] S. Anupama Kumar and Dr. Vijayalakshmi M.N “Implication of classification Techniques in Predicting Student’s Recital” International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.1, No.5, September 2011.