# Performance Evaluation and Data Analysis of Student Dataset using different Data Mining Algorithms

# Introduction

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information. It also refers to *extracting or “mining” knowledge from large amounts of data*. Many other terms carry a similar or slightly different meaning to data mining, such as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology and data dredging.

## Data Analysis and Knowledge Discovery

Data mining treated as synonym for another popular used term Knowledge Discovery from Data, or KDD. Data mining is a step in the knowledge discovery process; it is a process of discovering interesting knowledge from large amounts of data stored in databases, data warehouses, or other information repositories.

Data mining functionalities are used to specify the kind of patterns to be found in

data mining tasks.

In general, data mining tasks can be classified into two categories:

Descriptive and predictive. **Descriptive mining tasks characterize the general properties**

**of the data in the database**. In order to make predictions Predictive mining tasks perform inference on the current data

By data mining system we can generate thousands of patterns or rules but only small fraction of the patterns potentially generated would actually be of interest to any given user. A pattern is interested if it is (1) *easily understood* by humans, (2) *valid* on new or test data with some degree of *certainty*, (3) potentially *useful*, and (4) *novel*. A pattern is also interesting if it validates a hypothesis that the user *sought to confirm*. An interesting pattern represents knowledge[. [1]](#_References)

Here data mining is being used for analysis of Student Dataset as it would enable us to determine relationship among various attributes e.g. student age, citizenship, nationality, student university, Student Major and student level.

# Algorithms used for in-depth analysis of data

In current dataset Student data is having a large amount of Categorical Data; it is having 17 attributes; Date Of Birth, Age, Gender, Admission year, nation, nation-type, citizenship, residency, college, student\_major, Diploma description, cert\_average, cert date, student level, school and school city.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

In order to analyze and for extracting of useful information I would analyze data by various following data mining techniques and algorithms.

## Data Preprocessing

Today's databases are highly susceptible to noisy, missing, and inconsistent data due to their typically huge size (often several gigabytes or more) and their likely origin from multiple, heterogeneous sources.

There are a number of data preprocessing techniques. *Data cleaning* can be applied to remove noise and correct inconsistencies in the data. *Data cleaning* (or *data cleansing*) routines attempt **to fill in missing values**, smooth out noise while identifying outliers, and correct inconsistencies in the data. So here I would be applying this technique on my student dataset and all missing values would be replaced by attribute average. For example in my dataset 30th record has no value under Nation, Nation-type and School City. After applying data cleaning method for replacing missing values school city would get average value e.g. sharjah as it is mean or average value. In section3 this function input and output is shown.

## Data Visualization

Its major part of data mining for analyzing and understanding data behavior. Here visualization and knowledge representation techniques are used to present the mined knowledge to the user.

## Mining Frequent Patterns

### What is frequent item-set mining?

Frequent pattern mining searches for recurring relationships in a given data set**.** Frequent item set mining leads to the discovery of associations and correlations among items in large transactional or relational data sets.

A typical example of frequent item set mining is market basket analysis. This process analyzes customer buying habits by finding associations between the different items that customers place in their “shopping baskets”. The discovery of such associations can help retailers develop marketing strategies by gaining insight into which items are frequently purchased together by customers. For instance, ***if customers are buying milk, how likely are they to also buy bread (and what kind of bread) on the same trip to the supermarket***? Such information can lead to increased sales by helping retailers do selective marketing and plan their shelf space.

### Why to use Frequent Item-set Mining

Similarly,in given data set if I want to analyze Graduate student's behavior who belongs to college "Art, Humanities and Social sci" then what is the most student's citizenship, gender, Cert-Avg(High score) etc w.r.t their college.

I would like to get such information by using frequent item-set generation that what is likelihood value of graduate students with respect to their residency, citizenship, diploma description and gender as well.

## Classification

Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. Such analysis can help provide us with a better understanding of the data at large. I need analysis of student dataset that which school producing mostly graduated male/female students per year.

Data classification is a two-step process, as shown for the Student data in Figure 2.1. (The data are simplified for illustrative purposes. In reality, we may expect many more attributes to be considered.) In the first step, a classifier is built describing a predetermined set of data classes or concepts. This is the learning step (or training phase), where a classification algorithm builds the classifier by analyzing or “learning from” a training set made up of database tuples and their associated class labels. A tuple, ***X***, is represented by an *n*-dimensional attribute vector, ***X*** = (*x*1, *x*2, : : : , *xn*), depicting *n* measurements made on the tuple from *n* database attributes, respectively, *A*1, *A*2, : : : , *An*.1 Each tuple, ***X***, is assumed to belong to a predefined class as determined by another database attribute called the class label attribute.[1]

|  |  |  |  |
| --- | --- | --- | --- |
| Training Data | | | |
| **CITIZENSHIP** | **Gender** | **SCHOOL** | **SCHOOL\_CITY** |
| UAE-Um Alquwain | M | UAE University | Al Ain |
| UAE-Sharjah | F | Ajman Univ of Science | AJMAN |
| UAE-Sharjah | M | Police Sciences Academy | SHARJAH |
| UAE-Sharjah | M | Police Sciences Academy | SHARJAH |
| UAE-Ras Alkhaimah | N | Islamic & Arabic Studies Coll | DUBAI |
| UAE-Abu Dhabi | M | Islamic & Arabic Studies Coll | DUBAI |
| UAE-Sharjah | N | Police Sciences Academy | SHARJAH |

Classification Algorithm

Classification Rules

If school is UAE University then school city is Al Ain.

If gender is M (male) and citizenship is UAE-Sharjah then School is police sciences academy.

If citizenship is UAE-Abu Dhabi and gender is M(male) then school is Islamic & Arabic Studies Coll

Fig 2.1(a)

Classification Rules

New Data

Test Data

|  |  |  |  |
| --- | --- | --- | --- |
| **Citizenship** | **Gender** | **School** | **School City** |
| UAE-Abu Dhabi | M | Ajman Univ of Science | Ajman |
| UAE-Abu Dhabi | F | UAE University | AL Ain |

(Gender M, UAE-Sharjah,Sharjah??

Police Sciences Academy

Fig 2.1(b)

**Fig 2.1** The data classification process: (a) *Learning*: Training data are analyzed by a classification algorithm. Here, the class label attribute is *school*, and the learned model or classifier is represented in the form of classification rules. (b) *Classification*: Test data are used to estimate the accuracy of the classification rules. If the accuracy is considered acceptable, the rules can be applied to the classification of new data tuples.

Here the data analysis task is classification, where a model or classifier is constructed to predict *categorical labels* for student's school such as Police Sciences Academy, Islamic & Arabic Studies Coll, Ajman Univ of Science, Police College In Abu Dhabi, University of Sharjah and UAE University and Kuwait University.

Suppose data analyst wants to predict what would be student major with respect to their gender then here student major would be class label and we can visualize student behavior well via decision tree as its results are shown in section 3.

## Evaluation of Classification Methods

Classification and prediction methods can be compared and evaluated according to the following criteria:

**Accuracy:** The accuracy of a classifier refers to the ability of a given classifier to correctly predict the class label of new or previously unseen data (i.e., tuples without class label information). The *confusion matrix* is a useful tool for analyzing how well your classifier can recognize tuples of different classes, the confusion matrix for multiple classes show in section 3.

### Classification by Decision Tree Induction Algorithms

Decision tree induction is the learning of decision trees from class-labeled training tuples.

There are many algorithms:

Hunt's Algorithm

CART

1D3, C4.5

**ID3** uses **information gain** as its attribute selection measure.

**C4.5** uses **Gain Ratio**

**CART** uses **Gini Index**

### Why Decision Tree?

Conducting analysis of decision making under uncertainty using decision trees serves several purposes.

* First, a decision tree is a visual representation of a decision situation (and hence aids communication).
* Second, the branches of a tree explicitly show all those factors within the analysis that are considered relevant to the decision (and implicitly those that are not).
* Third, and more subtly, a decision tree generally captures the idea that if different decisions were to be taken then the structural nature of a situation (and hence of the model) may have changed dramatically.
* Fourth, and arguably the most powerful, a decision tree allows for forward and backward calculation paths to happen and hence the choice of the correct decision to take (optimality of decision making, or optimal exercise if embedded real options) is made automatically.

## Cluster Analysis

Clustering partitions the large data-sets into groups according to their *similarity values*. A cluster is a collection of data objects that are *similar* to one another within the same cluster and are *dissimilar* to the objects in other clusters.

Cluster analysis can be used as a stand-alone tool **to gain insight into the distribution of data, to observe the characteristics of each cluster**, and to focus on a particular set of clusters for further analysis

Additional advantages of such a clustering-based process are that it is adaptable to changes and helps single out useful features that distinguish different groups.

1. Performance evaluation of data using K-Mean with different values of K(# of clusters) and would report precision, recall, accuracy and classification error
2. Performance evaluation of data using K-Mediods with different values of K(# of clusters) and would report precision, recall, accuracy and classification error
3. ***Evaluation of clusterings using the validation operators ClusterCentroidEvaluator***
4. ***Evaluation of clusterings using the validation operators ClusterDensityEvaluator***

### Centroid-based clustering

### Density-based clustering

**CHAPTER 3**

**IMPLEMENTATION**

# Tools used:

Rapid Miner 4.6.

This software and the new version and all RapidMiner plugins are available at  
   
<http://rapid-i.com>

It can be downloaded from this link:

<http://sourceforge.net/projects/rapidminer/files/1.%20RapidMiner/4.6/>

## Why to do use any software for this data analysis?

I'm having 7,000 student records for which I've to extract some useful information for future. For example I need information according to different patterns. E.g.

How many bachelor students are under 30?

Which school-city is most popular in era 2010-2011?

What is Student Major for students having age >30?

What are mostly student major, gender and school for students having age 30-37?

I can't do it manually. It would take much time for analysis of such patterns from a large domain if I do it by counting data or records myself. There are some algorithms that give me such information within few minutes. For applying that algorithms, for knowledge extraction from big data there are many algorithms and tools available and Rapid Miner is ranked 1 for such analysis that’s why I'm taking help from this software for ease in doing correct and precise knowledge discovery.

## Why Rapid Miner?

Data mining is an essential process where intelligent methods are applied in order to extract data patterns. Rapid Miner provides much more analysis steps (operators) than any other data Mining tool e.g. as compared to Weka and much more possibilities to combine them. It provides an additional set of about 400 operators for many aspects of Data Mining not covered by Weka. E.g. Preprocessing methods, IO, Learner, validation, Clustering and visualization techniques algorithms which are not available within Weka.

### Data Mining and RapidMiner:

RapidMiner offers a huge number of learning schemes for: support vector machines (SVM),Statistical Analysis, decision tree rule learners, lazy learners, Bayesian learners, Logistic learners, association rule mining and clustering Meta learning schemes including Bayesian Boosting etc.

#### Decision Trees:

This operator learns decision trees from both nominal and numerical data. Decision trees are powerful classification methods which often can also easily be understood. This decision tree learner works similar to Quinlan‘s C4.5 or CART. The actual type of the tree is determined by the criterion, e.g. using gain ratio or Gini for CART / C4.5.

#### Efficient and Scalable Frequent Item set Mining Methods by Rapid Miner

Data can be analyze using frequent item-set mining by following algorithms of Rapid Miner

* 1. **Apriori**
  2. **FPGrowth**

**Apriori** is the basic algorithm for finding frequent item sets**.** Apriori generates candidate sets whereas FPGrowth uses specialized data structures (no candidate sets) so I would be using FPGrowth as it is more efficient and consume less memory than Apriori.

#### FPGrowth

It found frequent item Set without candidate generation. It constructs a highly compact data structure (an *FP-tree*) to compress the original transaction database. Rather than employing the generate and-test strategy of Apriori-like methods, it focuses on frequent pattern (fragment) growth, which avoids costly candidate generation, resulting in greater efficiency.

#### Generating Association Rules

Rule support and confidence are two measures of rule interestingness. Once the frequent item sets from transactions in a database *D* have been found, it is straightforward to generate strong association rules from them (where *strong* association rules satisfy both minimum support and minimum confidence). This can be done using Equation for confidence, which is here:

***Confidence (A🡪 B*) = *P* (*B* | *A*) = *support count* (*A* U *B*) /*support count* (*A*).[1]**

In given data set I generated thousands of association rules and in appendix only those rules are given that have 100% confidence of occurring.

#### Clustering- ClusterCentroid Evaluator

This is an evaluator for centroid based clustering methods. The average within cluster distance is calculated by averaging the distance between the centroid and all examples of a cluster.

#### Cluster Density Evaluator

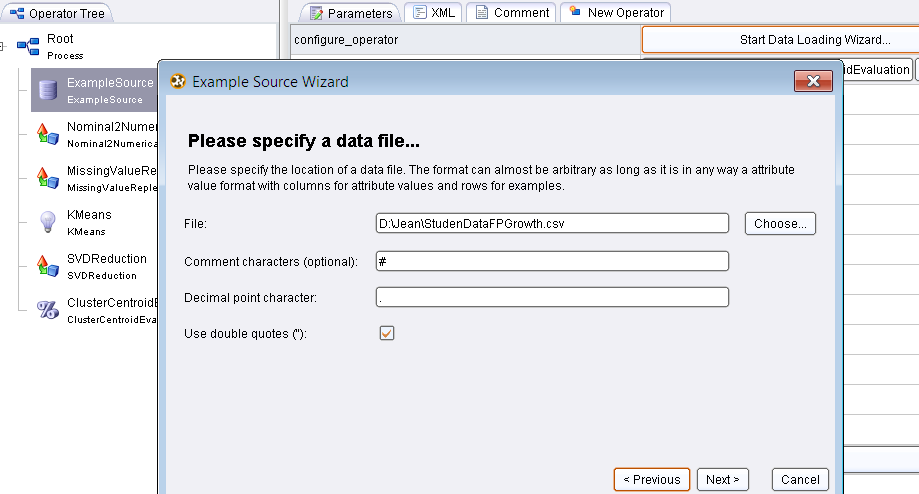
This operator is used to evaluate a non-hierarchical cluster model based on the average within cluster similarity/distance. It is computed by averaging all similarities / distances between each pair of examples of a cluster.

### Rapid Miner 4.6

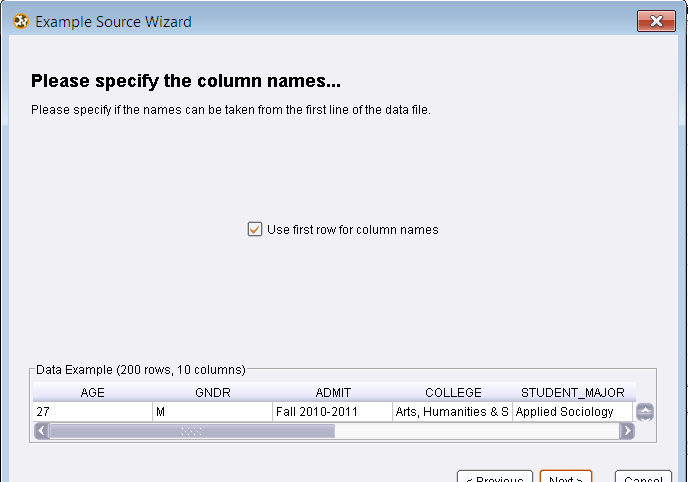


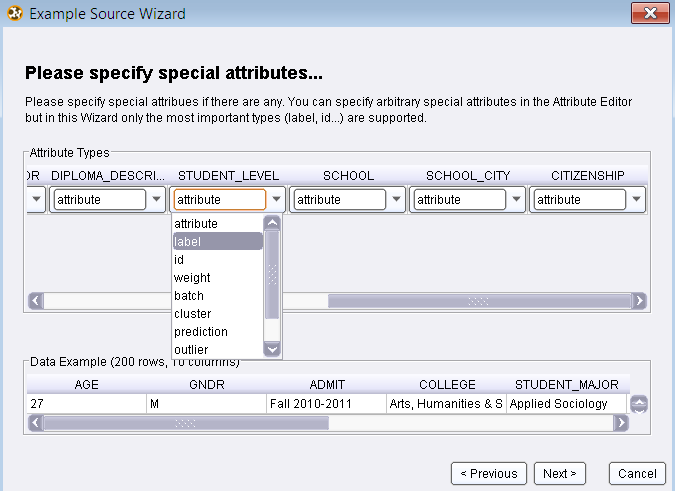
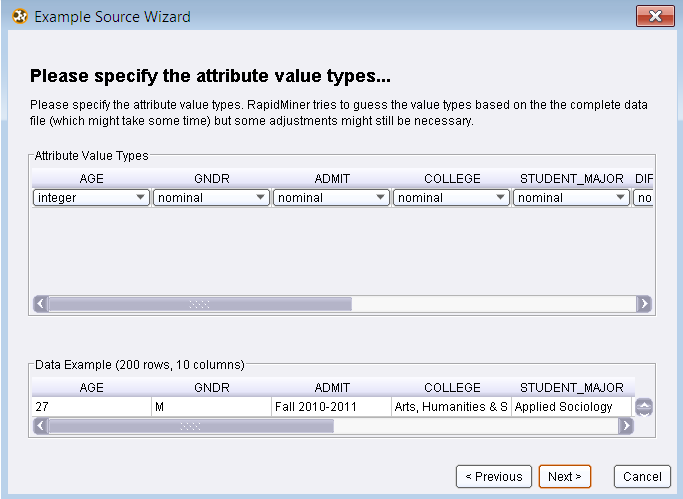
## Input Data File

At first example source is loaded by 'start data loading wizard'. Following are input screens for loading data-set file in required format.

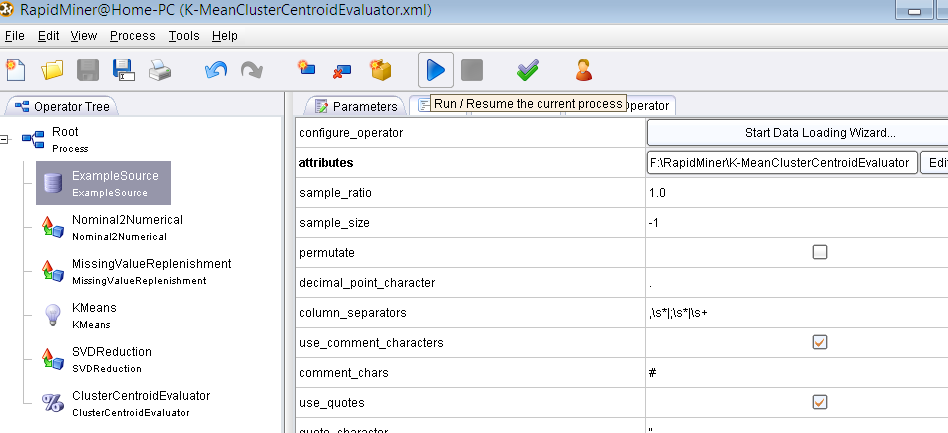


## 









After loading ExampleSource, file is run by pressing 'Run' button. Nearly for running all algorithms this file needs to be loaded. My input file is in CSV format.

## Statistical Analysis

I am having 7599 records of student data set. It is containing 15 nominal,1 integer and 1 real attribute. For determining valuable information and for all attribute behavior I need to find most well-liked and accepted value for each attribute for example:

Citizenship: Mostly students belong to which country during 2010-2011?

Student\_Major: Which was most studied course among students in 2010-2011?

Gender: which gender category mostly graduated in 2010-2011 either male or female?

College: Most graduated students belong to which college from 7000 records in 2010-2011?

Diploma Description: What was the diploma description for most of the students in 2010-2011?

School: Which school students mostly graduated in 2010-2011?

School city: Which school city students mostly did graduation in 2010-2011?

Such statistics can be found by applying DataStatistics operator on our input data set as shown in fig 3.1.

This operator calculates some very simple statistics about the given example set. These are the ranges of the attributes and the average or mode values for numerical or nominal attributes respectively. These informations are automatically calculated and displayed by the graphical user interface of RapidMiner.

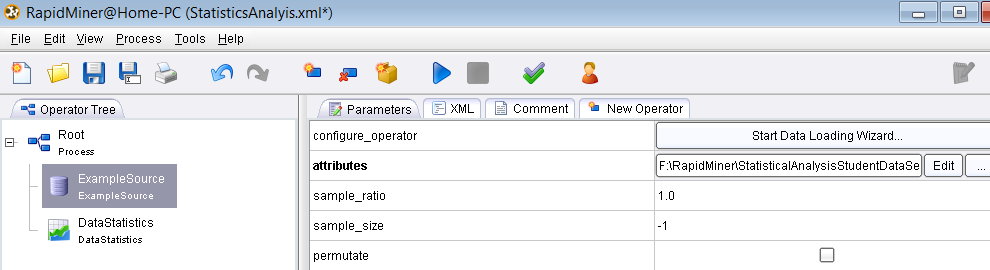
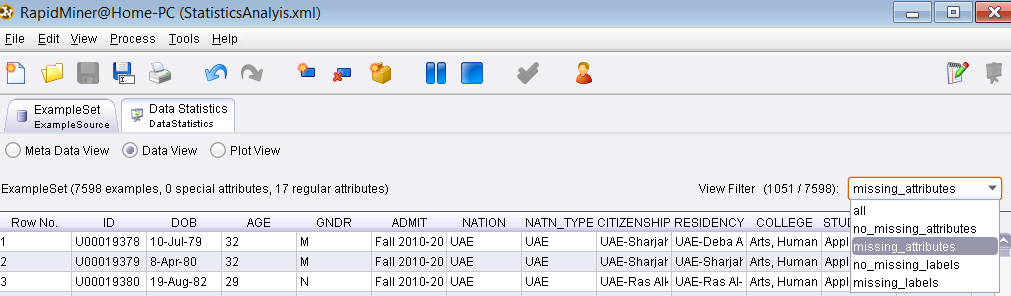


Fig 3.1

### Data Preprocessing

There are 1051 attributes having missing values as shown in fig 3.2. As data cleaning is part of data preprocessing. So I need to clean this dataset before mining any knowledge. Hence RapidMiner operator missingvaluereplishment replace all missing values by min, max or mode of attribute value.

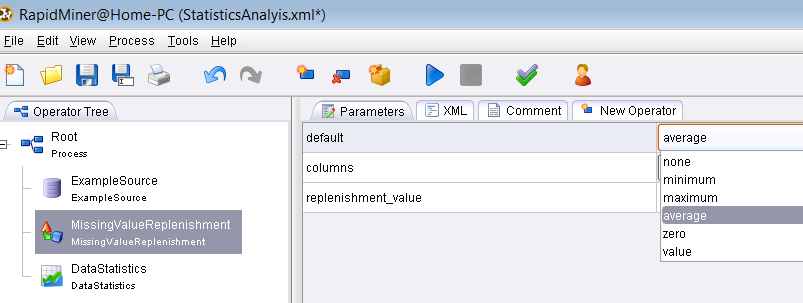
**Fig 3.2**

#### MissingValueReplenishment

It replaces missing values in examples. If a value is missing, it is replaced by one of the functions "minimum", "maximum", "average", and "none", which is applied to the non missing attribute values of the example set. "none" means, that the value is not replaced.

For nominal attributes the mode is used for the average, i.e. the nominal value which occurs most often in the data. For nominal attributes and replacement type zero the first nominal value defined for this attribute is used. The replenishment "value" indicates that the user defined parameter should be used for the replacement.

I would replace missing values by setting operator value as "average"; this is shown in fig 3.3 below.



**Fig 3.3**

After applying Data Statistics operator following information or knowledge about each attribute is mined. It is showing total number of distinct records for each value of corresponding attribute as there are 34 students having DOB 1st Jan 1993, 23 students have DOB 1st Jan 1994 etc.

### Date of Birth

1-Jan-93 (34), 1-Jan-94 (23), 28-Oct-92 (16), 19-Jun-93 (15), 5-Jul-93 (15), 1-Aug-93 (14), 18-May-93 (14), 19-Apr-92 (14), 1-Jul-92 (13), 16-Sep-93 (13), 17-Sep-93 (13), 20-May-93 (13), 21-Sep-92 (13), 22-Jul-92 (13), 1-Jan-92 (12), 14-Sep-93 (12), 17-Oct-92 (12), 20-Nov-92 (12), 27-Dec-92 (12), 29-Aug-93 (12), 3-Aug-93 (12), 3-Oct-92 (12), 31-Oct-93 (12), 8-Oct-92 (12), 1-Jan-90 (11), 1-Oct-92 (11), ... and 2624 more ... , 9-Jan-87 (1), 9-Jan-89 (1), 9-Jan-90 (1), 9-Jul-84 (1), 9-Jul-85 (1), 9-Jun-82 (1), 9-Jun-89 (1), 9-Jun-90 (1), 9-Mar-80 (1), 9-Mar-84 (1), 9-Mar-87 (1), 9-Mar-88 (1), 9-Mar-91 (1), 9-May-81 (1), 9-May-84 (1), 9-May-88 (1), 9-May-91 (1), 9-Nov-84 (1), 9-Nov-86 (1), 9-Oct-72 (1), 9-Oct-81 (1), 9-Oct-88 (1), 9-Oct-91 (1), 9-Sep-69 (1), 9-Sep-84 (1), 9-Sep-87 (1)

### Diploma Description

There are 3536 students having diploma description as 'Secondary school-scientific', 2377 students have secondary school-literature diploma, 782 students have High school-scientific diploma and so on.

Bachelor (291), Bachelor of Arts (26), Bachelor of Education (9), Bachelor of Science (19), Secondary School -Scientific (3536), Secondary School -Literature (2377), High School -Scientific (782), High School -Literature (134), IG -Literature (16), IG -Scientific (214), Bacalorate -Literature (11), High School -Commercial (21), Indian Board Cert. -Scientific (15), Secondary School -Commercial (10), Technical -Industrial (14), Indian Board Cert. -Literature (4), H.Sch. Pre-Uni. -Scientific (28), Secondary School -Industrial (23), IB -Scientific (8), Diploma in Nursing (6), Pak. Board Cert. -Scientific (9), Bacalorate -Scientific (7), H.Sch. Pre-Uni. -Literature (6), Pak. Board Cert. -Commercial (1), Indian Board Cert. -Commercial (4), Technical -Commercial (1), Secondary School -Shari'a (20), Pak. Board Cert. -Literature (3), Bachelor of Art (2), IB -Literature (1)

### CERT\_DATE

This is the date when student got his/her degree. Here 2237 students got their degree on 28th June 2011, 1858 students got their degree on 21st Jun 2010, 281 students got their degree on 1st July 2011, 243 students got their degree on 8th June 2009 and 213 students receive their degree on 1st July 2010 and so on…

28-Jun-11 (2237), 21-Jun-10 (1858), 1-Jul-11 (281), 8-Jun-09 (243), 1-Jul-10 (213), 3-Jun-08 (139), 7-Jul-10 (113), 13-Jul-11 (106), 27-Jun-11 (83), 26-Jun-11 (75), 14-Jun-07 (73), 1-Jul-09 (64), 21-Jun-11 (64), 23-Jun-10 (42), 2-Jul-09 (40), 15-Aug-11 (39), 20-Jun-11 (36), 1-Jul-08 (33), 21-Jul-10 (32), 6-Jun-10 (31), 26-Jun-08 (30), 20-Jun-10 (29), 6-Jun-11 (29), 6-Jun-09 (26), 1-Jun-10 (24), 15-Aug-10 (23), ... and 573 more ... , 7-Dec-10 (1), 7-Jan-01 (1), 7-Jan-02 (1), 7-Jun-04 (1), 8-Feb-05 (1), 8-Feb-92 (1), 8-Jan-09 (1), 8-Jul-01 (1), 8-Jul-79 (1), 8-Jun-08 (1), 8-Jun-44 (1), 8-Mar-10 (1), 8-Mar-11 (1), 8-Mar-98 (1), 8-May-11 (1), 8-Nov-98 (1), 8-Sep-11 (1), 9-Feb-03 (1), 9-Feb-10 (1), 9-Jan-99 (1), 9-Jul-05 (1), 9-Jul-09 (1), 9-Jul-11 (1), 9-Mar-10 (1), 9-May-11 (1), 9-Sep-04 (1)

### Gender

This student's dataset have 2854 male students, 4735 female students and 9 unknown.

M (2854), F (4735), N (9)

### Admission

There are 2788 students got admission in Fall 2010-2011, 585 students got admission in Spring 2010-2011, 23 students got admission in summer 2010-2011, 3443 students got admission in Fall 2011-2012, 758 students got admission in spring 2011-2012 and so on.

Fall 2010-2011 (2788), Fall 2011-2012 (3443), Spring 2010-2011 (585), Spring 2011-2012 (758), Summer 2010-2011 (23), Summer 2011-2012 (1)

### Student Major

852 student's major was Law, 729 student's major was common business program, 393 student's major was civil engineering, 380 student's major was Business administration, 362 student's major was medicine and surgery, 89 student's major was computer engineering and 337 student's major was pharmacy and so on.

Law (852), Common Business Program (729), Civil Engineering (393), Business Administration (380), Medicine & Surgery (362), Pharmacy (337), Public Relation (323), Dental Surgery (321), Electrical/Electronic Engr. (301), Mass Communication (297), Architectural Engineering (282), Information Technology (229), Sustainable/Renewable Enrg Eng (227), Biotechnology (217), Jurisprudence & its Foundation (185), Applied Sociology (175), Industrial Engineering & Mgt (172), Clinical Nutrition & Dietetics (135), Foundations of Religion (129), Sociology (122), English Language & Literature (108), Interior Architecture & Design (107), Computer Engineering (89), Medical Lab Technology (87), Graphics Design & Multimedia (85), Education (81), ... and 11 more ... , Computer Science (35), Non Degree Undergraduate (35), Health Services Administration (31), History & Islamic Civilization (31), Information Tech. - Multimedia (29), Secretariat & Office Mgt (29), Environmental Health (27), Chemistry (24), Sharia & Law (20), Fashion Design & Textile (19), Applied Physics (17), Library & Information Systems (17), Medical Diagnostic Imag - Bdg (15), Translation (14), Private Law (12), Dental Laboratory Technology (11), Mathematics (11), Environmental Health & Safety (9), Fine Arts (9), Communication (5), Dental Hygiene (4), Dental Assistant (1), Finance (1), Foundation (1), Non Degree Graduate (1), jewelry Design (1)

### College

827 student's college was engineering, 901 student's college was Law, 776 student's college was Community, 625 belonged to college 'Communication', 518 student's college was Arts, Humanities and social sci and in current data set there are 1855 students didn't designate any college.

Arts, Humanities & Social Sci. (518), Business Administration (348), Communication (625), Community (776), Dentistry (238), Engineering (827), Fine Arts & Design (148), Health Sciences (290), Law (901), Medicine (277), No College Designated (1855), Pharmacy (229), Sciences (183), Shari'a & Islamic Studies (383)

### Citizenship

There are 2574 student have citizenship UAE-Sharjah, 3073 students are non citizen, 849 student's citizenship GCC, 303 have citizenship UAE-Abu Dhabi and so on.

UAE-Um Alquwain (43), UAE-Sharjah (2574), UAE-Ras Alkhaimah (169), UAE-Abu Dhabi (214), Non Citizen (3073), UAE-Al Fujairah (111), GCC (849), UAE-Dubai (303), Nil (No Official Documents) (97), UAE-Ajman (68), UAE-Passport (96), Unknown Citizenship (1)

### School City

There are 3061 student's school city is Sharjah, 744 student's school city is Dubai, 101 were from Al Fujairah, 155 student's school city was KALBA and others are given:

SHARJAH (3061), Sharjah (851), DUBAI (744), ABU DHABI (555), Dubai (260), AJMAN (199), DABA AL HESSEN (156), KALBA (155), KUWAIT (126), Al Ain (123), Abu Dhabi (115), AL FUJAIRAH (101), RAS AL-KHAIMAH (99), SHARJAH - EMIRATE (92), KOR FAKAN (82), AL AIN (61), UMM AL QUWAIN (59), Ras Al-Khaima (50), Ajman (43), AL KHOBAR (40), Fujairah (39), RAS AL KHAIMAH - EMI (38), RIYADH (38), DOHA (34), Kuwait (31), Doha (29), ... and 113 more ... , Linhe (1), London (1), MADINAH (1), MOSANDAM (1), Madinah (1), Mississauga (1), Najran (1), Nigeria (1), OREGON (1), Ouargla (1), PAKISTAN (1), RABAT (1), RIFFA (1), SUR (1), Sabha (1), Saudi Arabian (1), Sterlitamak (1), TEHRAN (1), Tanger (1), Tanzania (1), The Eastern Region (1), Tirane (1), Tripoli (1), Yemen (1), abvja (1), canada (1)

### Nation-Type

3436 student's nation-type is UAE, 2640 student's nation-type is Other Arab, 856 are GCC, 580 are Non Arab and 86 are UAE-P.

UAE (3436), GCC (856), OTHER ARAB (2640), NON ARAB (580), UAE-P (86)

### Nation

There are 3436 student's nation is UAE, 509 nation was Jordan, 80 are without nationality, 56 student's nation is United States of America and others are given below.

UAE (3436), Jordan (509), Syria (496), Saudi Arabia (379), Palestine (354), Iraq (333), Egypt (285), Kuwait (213), Sudan (186), Iran (156), Yemen (149), Oman (133), UAE Passport (86), Lebanon (85), Qatar (82), Without Nationality (D.H.A.D) (80), Somalia (75), Pakistan (57), United States of America (56), Bahrain (49), Algeria (46), Canada (38), Comoros (33), India (32), Nigeria (21), Morocco (19), ... and 28 more ... , Burkina Faso (2), C?´te d'Ivoire (2), Eritrea (2), France (2), Mali (2), Philippines (2), Tanzania (2), Ukraine (2), Albania (1), Azerbaijan (1), Belarus (1), Belgium (1), Bulgaria (1), Cameroon (1), Denmark (1), Dominica (1), Finland (1), Indonesia (1), Mauritania (1), Netherlands (1), Norway (1), Romania (1), Seychelles (1), Singapore (1), Sweden (1), Venezuela (1)

### Residency

There are 2601 student's residency is UAE-Sharjah, 946 student's residency is UAE-Dubai, 652 student's residency is UAE-Abu Dhabi, 395 student's residency is UAE-Khor Fakkan and other student's residency is given:

UAE-Sharjah (2601), UAE-Dubai (946), UAE-Abu Dhabi (652), - (550), UAE-Kalba (448), UAE-Khor Fakkan (395), UAE-Ajman (246), UAE-Ras Al-Khaima (205), UAE-Deba Al-Hosn (189), UAE-Al Ain (166), UAE-Fujairah (114), Kuwait-Kuwait (106), UAE-Al-Dhaid (104), Qatar-Doha (85), UAE-Umm Al-Quwain (66), Saudi Arabia-Dammam (50), Saudi Arabia-Al Riyadh (45), Saudi Arabia-Al Khobar (38), Saudi Arabia-Dhahran (38), Saudi Arabia-jeddah (26), Oman-Muscat (25), UAE-Daba Al Hessen (19), Saudi Arabia-Al Quteef (18), UAE-Al Madam (16), UAE-Deba Al-Fujairah (16), UAE-Kor Fakan (14), ... and 221 more ... , Saudi Arabia-Saimahat (1), Saudi Arabia-Sharqiya (1), Saudi Arabia-khobar (1), Saudi Arabia-saihat (1), Senegal-Dakkar (1), Senegal-Kerevane (1), Sudan-Umm Durman (1), Sudan-om dorman (1), Sweden-Falkenberg (1), Switzerland-Gerere (1), Syria-Damascus (1), UAE-Abu Mousa (1), UAE-Al Fujairah (1), UAE-Al Himreya (1), UAE-Hatta (1), UAE-KHOR FAKKAN (1), UAE-Khor Fakan (1), UAE-Khur Kalba (1), UAE-Umm Al Quwain (1), UAE-Wadi Al-Hilo (1), United Kingdom-London (1), United States of America-Columbus (1), Yemen-Al Hodidah (1), Yemen-Hadida (1), Yemen-Sanaa (1), Yemen-TAIZ (1)

### SCHOOL

154 student's school is salma bint qais secondary G, 107 student's school is Al Ahlia Pvt secondary and others given below.

Undeclared (695), Al Sho'ala Private Sch. (214), Salma Bint Qais Secondary G (154), Jameela Bou Hurraid Sec G (148), Bahithat Al Badyeh Secondary G (135), Al Shifa Bint Abdulla Ctr (119), University of Sharjah (111), Fatima Al Zahra' Secondary G. (108), Al Ahlia Pvt Secondary (107), Al Reffa'a Secondary G. (107), Waset Secondary G (103), Al Gobaibah Secondary G. (93), Rouqaya Sec. Sch. G. (93), Al Nour Intrnational Pvt B (89), Al Ma'arifa Int. Private Sch. (88), Al Wahda Private School (85), Al Khalil Bin Ahmed Secondary (82), Sharjah American Intl School (77), Um Amarah Sec. Sch. Edu. G. (74), Al-Ibda'a Girls' Sec.Sch. (73), Umm Omarah Secondary G. (71), Dubai International Secondary (70), Dubai Modern Education School (66), Al Ma'refa Private School (65), Al Rashed Al Salleh Pvt Sec (65), Dubai National School. (65), ... and 796 more ... , The Tenth in Madinah (1), The Third Secondary School (1), Third Secondary School Dammam (1), Thirty Seventh Secondary Sch. (1), Uasin Gishu High School (1), Um Amer Alansarieh Secondary (1), Um-Kalthum Secondary (1), Umamah Bint Abi Al Aaas Sec. G (1), Umm Al Hammam Secondary (1), Umm Al Qurah Secondary School (1), Umm Habiba Secondary G (1), Umm Hakeem Secondary (1), Umm Salamah Secondary (1), Universal American School (1), University of Khartoum (1), University of South Carolina (1), University of Technology (1), Urwa Bin Mas'oud Secondary (1), Wadi Sfaini Boys Secondary (1), Walnut Ridge High School (1), Wasit Model Sch. for Sec. Edu. (1), Wdam Al Gafe Sec. Sch. (1), Yazeed Bin Al Muhalab Sec (1), Zaid The 1st Secondary (1), Zarka Private University (1), Zobaidah Secondary (1)

### Student Level

447 student's level is Graduate, 3811 student's level is undergraduate, 776 student's level is Diploma, 514 level is Foundation year, 1819 student's level is Intensive English, 81 higher diploma,148 level is Fine Art, 1 dentistry training student and 1 student level is Doctorate.

Graduate (447), Higher Diploma (81), Undergraduate (3811), Diploma (776), Foundation Year (514), Fine Art (148), Intensive English (1819), Dentistry Training (1), Doctorate (1)

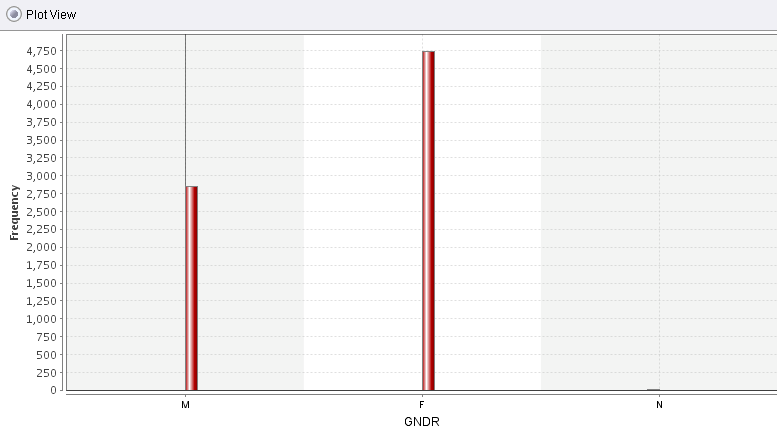
## Meta Data View/Statistics

After applying Data Statistics operator following information displayed.

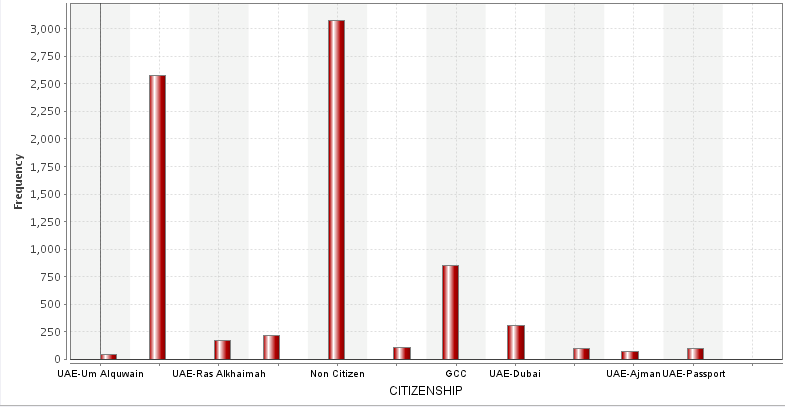
## 

|  |  |  |
| --- | --- | --- |
| **Attribute** | **Value Type** | **Mode/Max/Avg.** |
| AGE | Integer | avg = 20.624 +/- 4.468 |
| DOB[Date Of Birth] | Nominal | 1st Jan 1993 |
| Gender | Nominal | F(female) |
| Admission | Nominal | Fall 2011-2012 |
| Student\_Major | Nominal | Law |
| College | Nominal | No college designated |
| Citizenship | Nominal | Non Citizen |
| School City | Nominal | SHARJAH |
| Diploma-Description | Nominal | Secondary |
| Student Level | Nominal | Undergraduate |
| Residency | Nominal | UAE-Sharjah |
| CERT-Date | Nominal | 28th June 2011 |

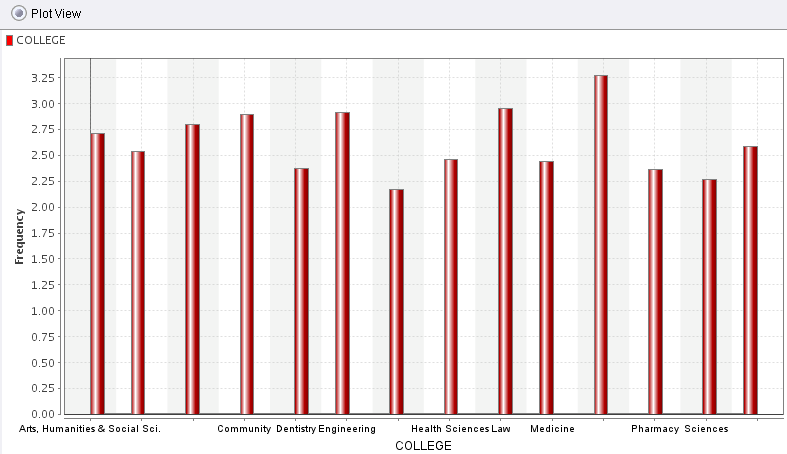
## Plot View

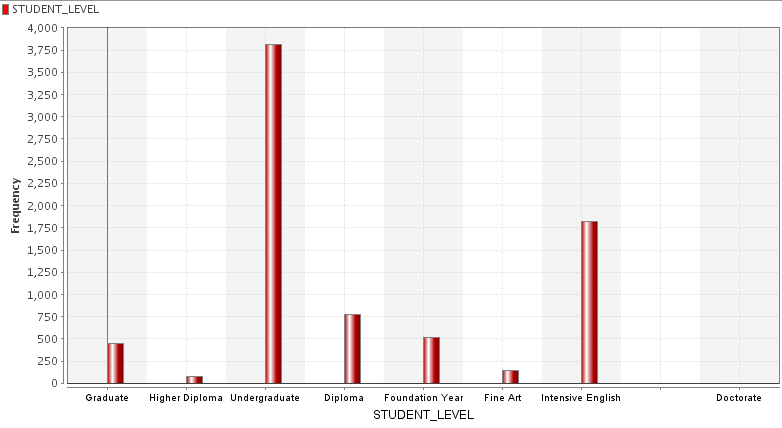


**Comment:** Above graph showing mostly students (nearly 4,750) in current dataset are females and then 2,750 are male students.



**Comment:** Most students are non citizen, nearly 850 have citizenship GCC and nearly 2600 students have citizenship UAE-Sharjah, and 300 students have citizenship UAE-Dubai and so on.





**Comment:**  3800 students are undergraduates, 1800 students have Intensive English level and nearly 490 students are graduates.

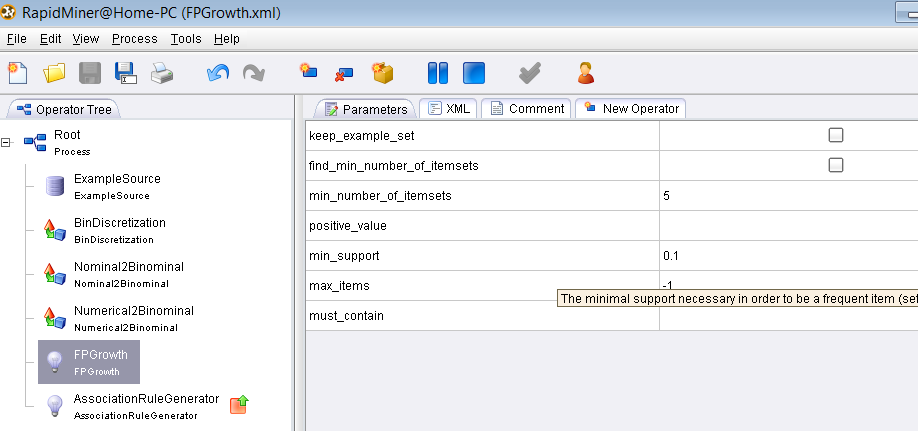
## Frequent Item set Mining using FPGrowth

Input operator's description in Appendix.

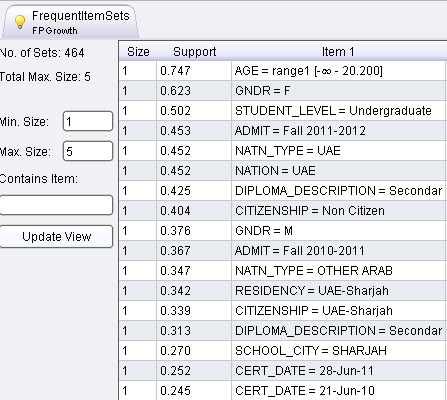
Output: No. of sets: 464

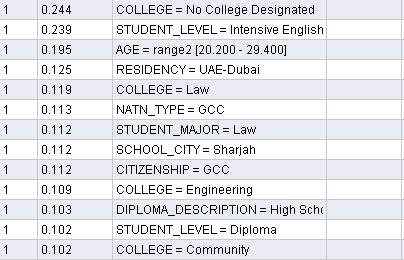
Total Max size: 5

Min Support: 0.1



After running this file I got 464 item sets, at first analysis of distinct item set is given. Now all item sets having minimum support 0.1 are given. Here 0.1 min support means 0.1\*7599= 759 such attribute's value having at least this occurring count (occurring values can be higher than min support) are displayed.

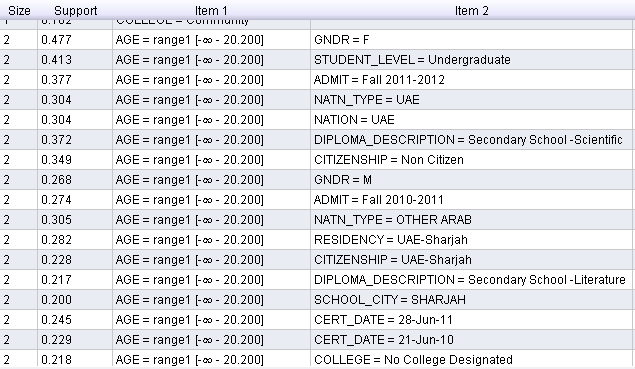




### One item set description

* First result showing that there is 5,676 student's age is under 20. [0.747\*7599=5676] or 74% student's age is under 20 and 19.5% student age range is 20-29.
* 4,734 students are female. [0.623\*7599=4734] or 62% students are female and 37.6% are male.
* 3,814 student's level is undergraduate.[0.502\*7599=3,814] or 50% students are undergraduate and 23.9% student level is Intensive English and 10.2% student's level is diploma.
* 45% student's got admission in Fall 2011-2012 and 36.7% got admission in Fall 2010-2011.
* 45% student's nation and nation type is UAE.
* 42% Student's diploma description is secondary school scientific and 31.3% student's diploma description is secondary school literature and 10.3% diploma description is High school scientific.
* 40% student's are non-citizens and 33.9% citizenship is UAE-Sharjah
* 34.7% student's nation type is other Arabs and 11.3% nation type is GCC.
* 34.2% student's residency is UAE-Sharjah and 12.5% student's residency is UAE-Dubai.
* 27% student's school city is Sharjah
* 25% student's Cert Date (date when obtained higher degree) is 28th June 2011 and 24.5% student's cert date is 21'st June 2010.
* 11.9% student's college is Law AND 10.9% student's college is engineering and 10.2% student's college is community.

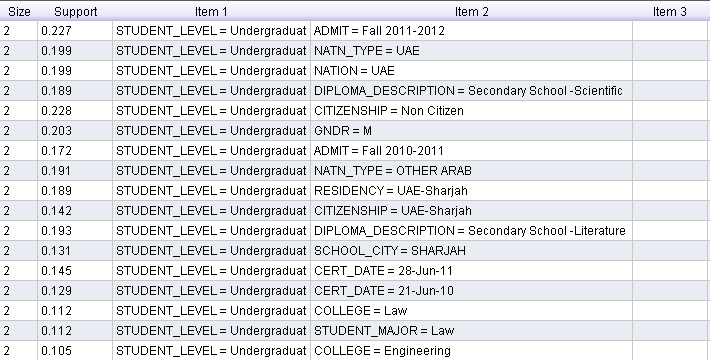
### Two item set description



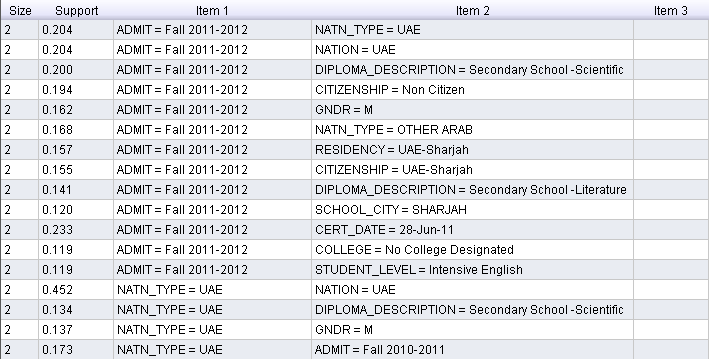
* There are 47.7% female student's age is under 20 and 26.8% male students are under 20.
* 41.3% undergraduate student's age is below 20.
* 37.7% students who got admission in Fall 2011-2012 having age under 20 and 27.4% got admission in Fall 2010-2011 are under 20.
* 30.4% students having age is under 20 and their nation and nation-type is UAE.
* 37.2% student's age is under 20 and DIPLOMA\_DESCRIPTION is Secondary School –Scientific.
* 34.9% students are noncitizen and having age under 20.
* 30.5% student's nation type is other Arabs and age is under 20.
* 28.2% having residency UAE-Sharjah are under 20.
* 22.8% have citizenship UAE-Sharjah are under 20.
* 22% student's age is under 20 and DIPLOMA\_DESCRIPTION is Secondary School –Literature.
* 20% having age below 20 and school city is Sharjah.
* 25% student's cert date[when obtained higher degree] is 28th june 2011 are uder age 20. And 23% student's cert date is 21'st June 2010 are under 20.
* 22% student's level is intensive English are under age 20.

## 

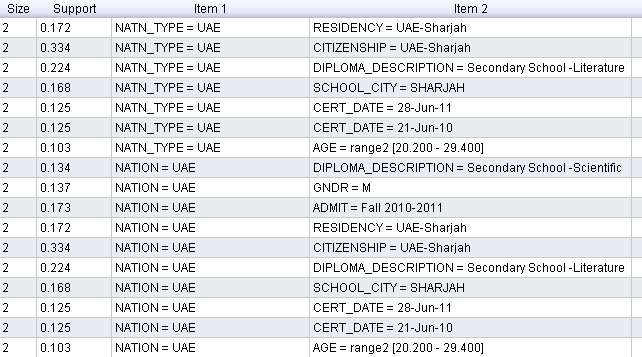
* Nearly 30% female student's level is undergraduate.
* 29% female student's got admission in Fall 2011-2012 and 24% female students got admission in Fall 2010-2011.
* 31.4% female student's nation type and nation is UAE.
* 26.3% female student's diploma description is secondary school scientific.
* 23% female students are non-citizen.
* 19% female student's nation type is other Arab.
* 23% female student's residency is UAE-Sharjah.
* 24% female student's citizenship is UAE-Sharjah.
* 22% female student's diploma description is secondary school literature.
* 15% female haven't designated any college.
* 15% female student's level is Intensive English.
* 11% female student's age range is 20-29.



* 22% undergraduates got admission in fall 2011-2012 and 17% undergraduates got admission in Fall 2010-2011.
* 20% undergraduate's nation type and nation is UAE and 19% undergraduate's nation type is other Arabs.
* 19% undergraduate's diploma description is secondary school scientific and 19% having diploma description 'Secondary school literature'.
* 23% undergraduates are non-citizen.
* 20% undergraduates are male.
* 19% undergraduate's residency is UAE-Sharjah and 14% undergraduate's citizenship is UAE-Sharjah
* 15% undergraduate got their degree on 28th June 2011 and 13% cert-date is 21st June 2010.
* 11% Undergraduates College and student major is 'LAW'.
* 10.5% undergraduate's college is 'Engineering'.

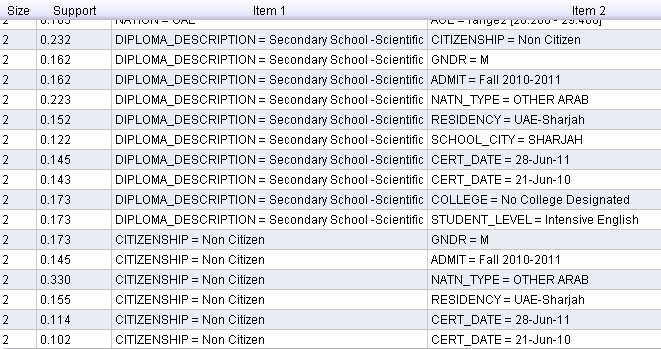


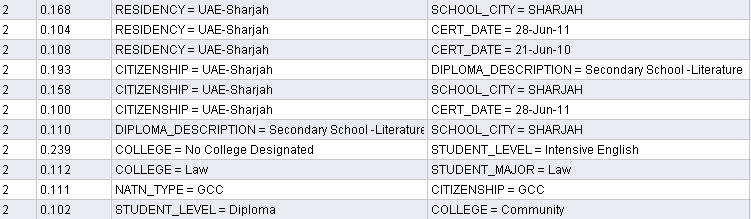
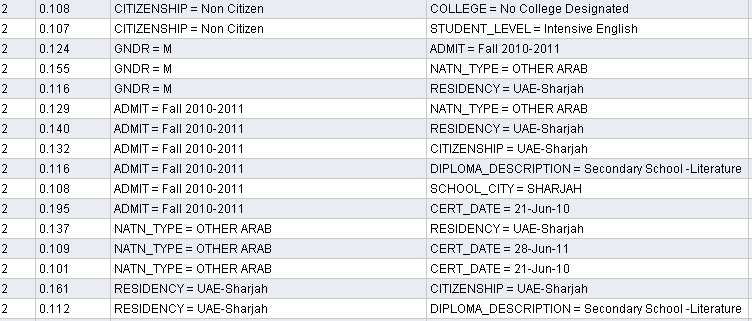
* 20% students who got admission in Fall2011-2012 have nation,nation type is UAE and 17% have nation-type 'Other Arabs'
* 45% student's nation and nation type is UAE.
* 14% male student's have nation type UAE.
* 20% students who got admission in Fall2011-2012 have diploma description as secondary school scientific and 14% have diploma description as secondary school literature.
* 19% students who got admission in Fall2011-2012 are non-citizen.
* 16% students who got admission in Fall2011-2012 are male.
* 16% students who got admission in Fall2011-2012 have residency and citizenship UAE-Sharjah
* 12% students who got admission in Fall2011-2012 have school city Sharjah.
* 23% students who got admission in Fall2011-2012 have cert-date 28th June 2011.
* 12% students who got admission in Fall2011-2012 haven't designated any college.
* 12% students who got admission in Fall2011-2012 have student level 'Intensive English'.
* 13% male student's have nation type UAE.
* 17% students who got admission in Fall2010-2011 have nation type UAE.



* 33% have nation,nation-type UAE and citizenship is UAE-Sharjah
* 22% have nation type UAE have diploma description as secondary school literature.
* 13% male student's belongs to Nation UAE.
* 17% students who got admission in fall 2010-2011 belong to nation UAE.
* 10% have age-range 20-29 belong to nation UAE.

Other remaining two-item set results is shown in figure below.



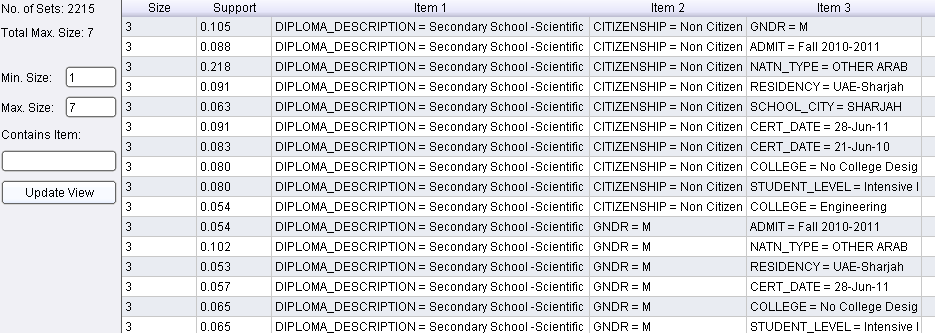


* 22% students have diploma description 'Secondary school scientific' and nation type 'other Arab'
* 33% have nation type 'Other Arab' are non-citizen.
* 12% male student's got admission in Fall 2010-2011.
* 20% students who got admission in Fall 2010-2011 they obtained their degree/certificate on 21st June 2010.
* 24% student's level is Intensive English didn't designate any college.
* 19% student's citizenship is UAE-Sharjah and their diploma description is secondary school literature.

### Three item-set description

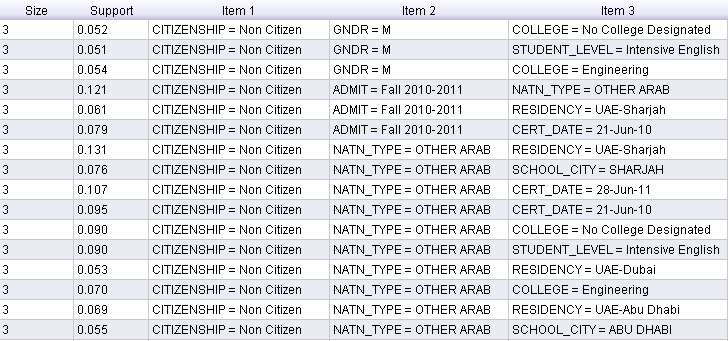
Min\_Support= 0.05

Now only items are included which have their occurring value >= 379 as now min-support 5%=0.05 means .05\*7599[total records]= 379.9 or 380

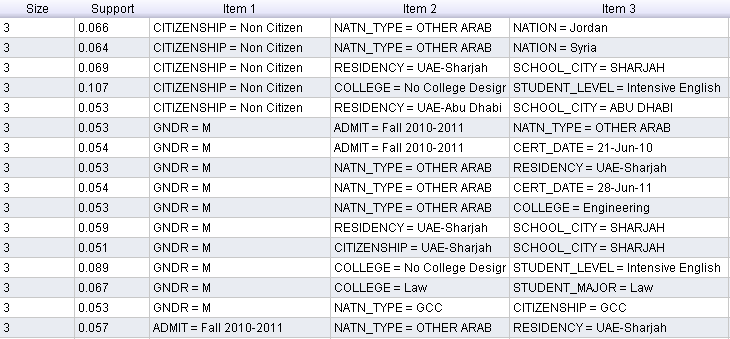


There are 410 non-citizen students having diploma secondary school scientific and college Engineering

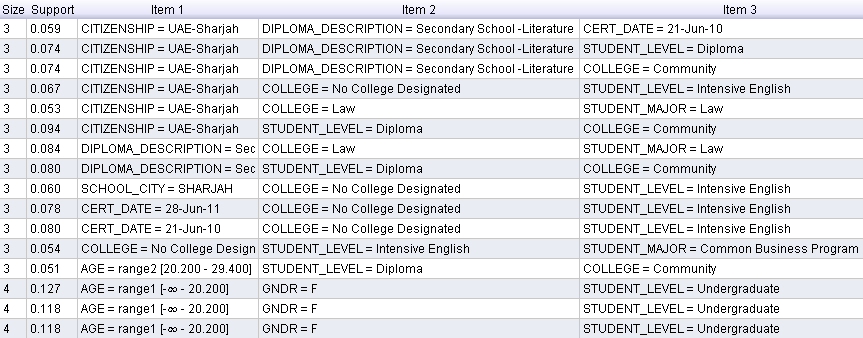
10% male students have nation type 'other arab' and diploma description secondary school scientific.



* 12% non-citizen students who got admission in fall 2010-2011 have nation type 'Other Arab'.
* 6% non-citizen students who got admission in fall 2010-2011 have residency 'UAE-Sharjah'.
* 5% non-citizen students have nation-type 'Other Arab' and residency 'UAE-Dubai'.
* 7% non-citizen students have nation-type 'Other Arab' and residency 'UAE-Abu Dhabi'

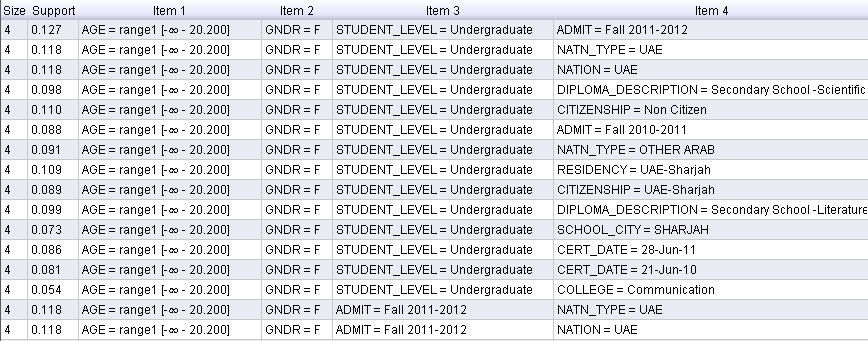


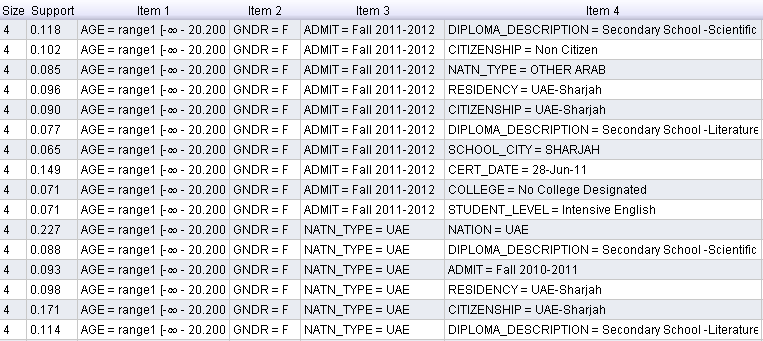
* 6.6% non-citizen students have nation-type 'Other Arab' and they belong to Nation 'Jordan'.
* 6% non-citizen students have nation-type 'Other Arab' and they belong to Nation 'Syria'.
* 9% male students don't designated any college have student level intensive English.
* 5% male student's college and student major is 'LAW'.
* 5% male student's nation type and Nation is 'GCC'.

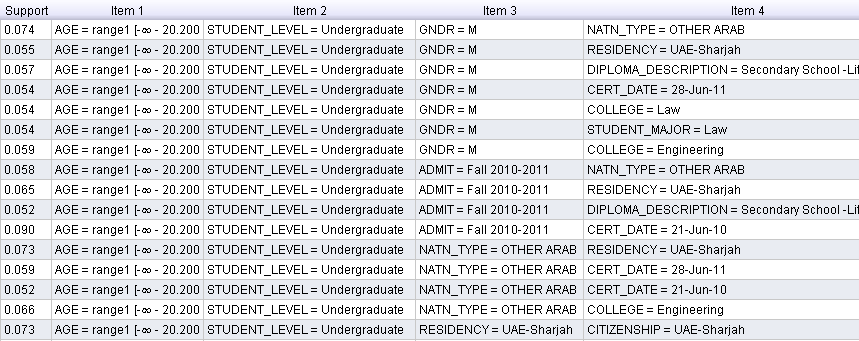


* 9% student's citizenship is 'UAE-Sharjah', student level is diploma and college is 'Community'.
* 7% student's citizenship is 'UAE-Sharjah', diploma 'secondary school literature' and college 'community'.
* 6% student's citizenship is 'UAE-Sharjah', diploma 'secondary school literature' and cert-date [when obtained higher degree] is 21st June 2010.
* 8% students have diploma in 'secondary school literature', college is law and student major is also Law.
* 5% students don't designate any college have student level 'Intensive English' and student major 'Common Business Program'.

### Four Itemset Descriptions



* 12% female undergraduate students have age under 20 got admissions in fall 2011-2012.
* 11% female undergraduate students have age under 20 have nation-type UAE.
* 9.8% female undergraduate students have age under 20 have DIPLOMA 'Secondary School scientific'.
* 9.9% female undergraduate students have age under 20 have DIPLOMA description 'Secondary School Literature'.
* 5.4% female undergraduate students have age under 20 studied in College 'Communication'.
* 15% female age less than 20 students who got admission in Fall 2011-2012 have diploma description secondary school scientific.
* 17% female age less than 20 students have nation-type UAE and citizenship 'UAE-Sharjah'.
* 23% female age less than 20 students have nation and nation type UAE.
* 7% female; age less than 20; students who got admission in Fall 2011-2012 have school-city 'Sharjah'.
* 11% female students age<20 have nation type UAE and diploma description 'Secondary School Literature
* 9% Female students age<20 have nation type UAE and diploma description 'Secondary School Scientific.



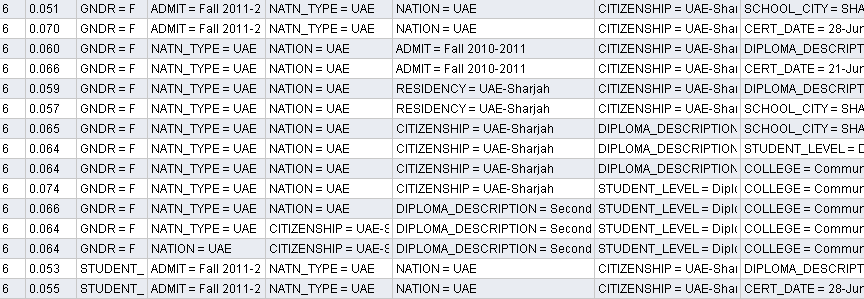
* 9% undergraduate age<20 students who got admission in Fall 2010-2011 have CERT\_DATE 21st June 2010.
* 6% undergraduate age<20 students have nation-type 'Other Arab' studied in college 'Engineering'.

Here only partial results are being display as complete results are in thousands so can't be displayed.

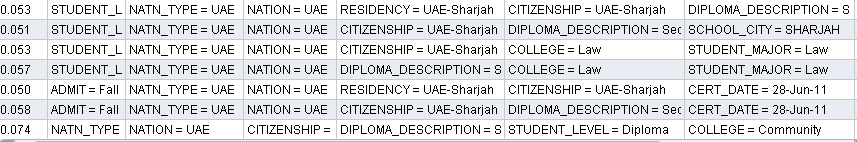
### Six item sets description

## 

* 9% undergraduate age<20 have nation, nation-type 'UAE', citizenship 'UAE Sharjah' and diploma description 'secondary school literature'.
* 5% undergraduate age<20 who got admission in Fall 2011-2012 are non-citizens have diploma description 'secondary school scientific' and nation-type 'Other Arab'.



* 7% female students who got admission in Fall 2011-2012 have citizenship 'UAE-Sharjah', nation and nation-type UAE got their higher degree in 28th June 2011.
* 7% female students, UAE-Sharjah citizens belong to nation,nation-type UAE; their student level is Diploma and college is community.
* 6.6% female students having nation,nation-type 'UAE'; their level is diploma and diploma-description is 'Secondary school literature' belongs to college 'Community'.
* 5% undergraduate students who got admission in Fall 2011-2012 having nation and nation-type UAE, citizenship is UAE-Sharjah have diploma description 'secondary school literature'.
* 5% undergraduate students who got admission in Fall 2011-2012 having nation and nation-type UAE, citizenship is UAE-Sharjah obtained their higher degree on 28th June 2011.



* 5.3% undergraduate students, UAE-Sharjah residents belong to Nation UAE have diploma description 'Secondary School Literature'.
* 5.1% undergraduate students, UAE-Sharjah residents belong to Nation UAE have diploma description 'Secondary School Literature' their school city is Sharjah.
* 6% students who got admission in Fall 2011-2012; UAE-Sharjah citizens belong to Nation and nation-type UAE have diploma description 'Secondary School Literature', have cert-date[when obtained higher certificate] is 28th June 2011.
* 7% UAE-Sharjah citizen, UAE nationality holder have diploma description 'Secondary School Literature', student-level is diploma belonged to college Community.
* 5.7% undergraduate students belong to Nation and nation-type UAE have diploma description 'Secondary School Literature', their major was 'LAW' and college also 'LAW'.

### Seven ItemSets Description

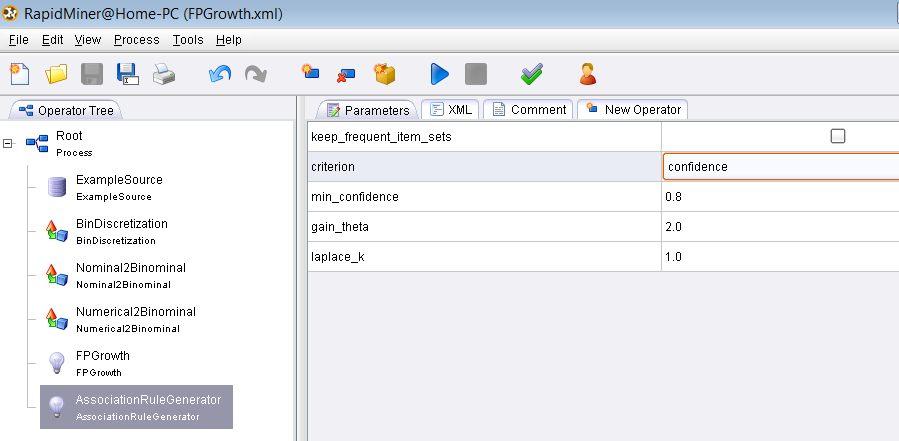


* 5% female undergraduates students <20 years age have nation UAE and UAE-Sharjah residents and citizens.
* 6% female undergraduates students <20 years age have nation UAE and UAE-Sharjah citizens have diploma description 'Secondary School Literature'.
* 5.3% female students <20 years age have nation and nation-type UAE who got admission in Fall 2011-2012 and UAE-Sharjah citizens have diploma description 'Secondary School Literature'.
* 6.8% female students <20 years age have nation and nation-type UAE who got admission in Fall 2011-2012 and UAE-Sharjah citizens obtained their high school certificate on 28th June 2011.
* 6.1% female students <20 years age have nation and nation-type UAE who got admission in Fall 2010-2011 and UAE-Sharjah citizens obtained their high school certificate on 21st June 2010.
* 5.3% undergraduates students <20 years age have nation and nation-type UAE and UAE-Sharjah citizens who got admission in Fall 2011-2012 have cert-date 28th June 2011.
* 5.4% students <20 years age have nation and nation-type UAE who got admission in Fall 2011-2012 and UAE-Sharjah citizens have diploma description 'Secondary School Literature' obtained their high-school/college degree on 28th June 2011.
* 6.4% female students have nation and nation-type UAE and UAE-Sharjah citizens have diploma description 'Secondary School Literature', Student level is diploma and college is 'Community'.

## Association Rule Mining

RapidMiner have one operator named **'Association Rule Generator'** for rule generations. This operator generates association rules from frequent item sets. In RapidMiner, the process of frequent item set mining is divided into two parts: first, the generation of frequent item sets and second, the generation of association rules from these sets.

For the generation of frequent item sets, I used for example the operator *FPGrowth*. Then result will be a set of frequent item sets which could be used as input for this operator.



### Minimum Confidence=0.8

As I set min\_confidence value 0.8, it means it would generate and display only those rules which are having confidence value>=80%. Rules generated are in thousands But here only describing those rules are being described which have 100% confidence that if part (premises) happened then conclusion (then part) must be happen.

1. [STUDENT\_MAJOR = Law] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
   There is 100% confidence if student major is 'LAW' then student-level is undergraduate.
2. [STUDENT\_LEVEL = Intensive English] --> [COLLEGE = No College Designated] (confidence: 1.000)

There is 100% confidence if student-level is Intensive English then student didn't designate any college.

1. [STUDENT\_MAJOR = Law] --> [COLLEGE = Law] (confidence: 1.000)  
   There is 100% confidence if student major is 'LAW' then student-College is Law.
2. [STUDENT\_LEVEL = Diploma] --> [COLLEGE = Community] (confidence: 1.000)
3. [COLLEGE = Community] --> [STUDENT\_LEVEL = Diploma] (confidence: 1.000)

There is 100% confidence if student-level is 'Diploma' then student-College is Community and vice versa.

1. [AGE = range1 [-∞ - 20.200], COLLEGE = Law] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
   There is 100% confidence if student age is<=20 and college is law then student-level would be 'undergraduate'.
2. [AGE = range1 [-∞ - 20.200], STUDENT\_MAJOR = Law] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)

There is 100% confidence if student age is<=20 and Student\_major is law then student-level would be 'undergraduate'.

[SCHOOL = Jameela Bou Hurraid Sec G] --> [GNDR = F] (confidence: 1.000)

[SCHOOL = Bahithat Al Badyeh Secondary G] --> [GNDR = F] (confidence: 1.000)

[STUDENT\_MAJOR = Sociology] --> [GNDR = F] (confidence: 1.000)

[SCHOOL = Al Shifa Bint Abdulla Ctr] --> [GNDR = F] (confidence: 1.000)

[STUDENT\_MAJOR = English Language & Literature] --> [GNDR = F] (confidence: 1.000)  
[SCHOOL = Fatima Al Zahra' Secondary G.] --> [GNDR = F] (confidence: 1.000)  
[SCHOOL = Al Reffa'a Secondary G.] --> [GNDR = F] (confidence: 1.000)  
[SCHOOL = Waset Secondary G] --> [GNDR = F] (confidence: 1.000)  
[SCHOOL = Rouqaya Sec. Sch. G.] --> [GNDR = F] (confidence: 1.000)  
[SCHOOL = Al Gobaibah Secondary G.] --> [GNDR = F] (confidence: 1.000)  
[STUDENT\_MAJOR = Education] --> [GNDR = F] (confidence: 1.000)  
[STUDENT\_LEVEL = Higher Diploma] --> [GNDR = F] (confidence: 1.000)

1. 100% confidence if school is 'Jameela Bou Hurraid Sec G' or school is 'Bahithat….' Or student major is 'Sociology' or school is 'Al shifa Bint Abdulla CTr' or student major is 'English lang and literature' or school is 'Fatima Al Zahra Secondry G' or school is 'Al Reffa'a Secondary G' or school is 'Waset Secondary' or school is 'Al Gobaibah Secondary G' or student major is 'Education' or student level is 'Higher Diploma' **then Gender is female.**

[STUDENT\_MAJOR = Law] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[STUDENT\_MAJOR = Public Relation] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[STUDENT\_MAJOR = Mass Communication] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[COLLEGE = Health Sciences] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[COLLEGE = Pharmacy] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[STUDENT\_MAJOR = Foundations of Religion] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[STUDENT\_MAJOR = Sociology] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)

1. 100% confidence if student\_major is 'Law'or 'Public Relation' or 'Mass Communication' or 'Foundations of Religion' or 'Sociology' or college is 'Health Sciences' or 'Pharmacy' then student-level is undergraduate.

[SCHOOL = Al Ahlia Pvt Secondary] --> [DIPLOMA\_DESCRIPTION = Secondary School -Scientific] (confidence: 1.000)

1. 100% confidence if school is 'Al Ahlia Pvt Secondary' then diploma description is 'secondary school scientific'.

[SCHOOL = Al Khalil Bin Ahmed Secondary] --> [GNDR = M] (confidence: 1.000)

1. 100% confidence if school is 'Al khalil Bin Ahmed..' then gender is male.

[SCHOOL = Al Sho'ala Private Sch.] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
[SCHOOL = Jameela Bou Hurraid Sec G] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
[SCHOOL = Bahithat Al Badyeh Secondary G] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
[SCHOOL = University of Sharjah] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
[SCHOOL = Fatima Al Zahra' Secondary G.] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
[SCHOOL = Al Reffa'a Secondary G.] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
[SCHOOL = Al Ahlia Pvt Secondary] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
[SCHOOL = Waset Secondary G] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
[SCHOOL = Al Gobaibah Secondary G.] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
[SCHOOL = Al Khalil Bin Ahmed Secondary] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)

[SCHOOL = Rouqaya Sec. Sch. G.] --> [SCHOOL\_CITY = Sharjah] (confidence: 1.000)

[SCHOOL = Al Ma'arifa Int. Private Sch.] --> [SCHOOL\_CITY = Sharjah] (confidence: 1.000)

1. There is 100% confidence if school is 'Al Sho'ala..' or 'Jameela bou hurraid ..' or 'Bahithat Al Badyeh..' or 'Uni of Sharjah' or 'Fatima Al zahra' or 'Al Reffa'a' or 'Al Ahlia pvt..' or 'wased secondary G.' or 'Al khalil…' or 'Rouqah sec. sch. G.' or 'Al Ma'arifa Int…' **then school city is SHARJAH.**

[STUDENT\_MAJOR = Information Technology] --> [STUDENT\_LEVEL = Diploma] (confidence: 1.000)  
[STUDENT\_MAJOR = Information Technology] --> [COLLEGE = Community] (confidence: 1.000)

1. 100% confidence **if student-major is 'Information Technology'** then college is 'community' or student\_level is 'Diploma'.

[STUDENT\_MAJOR = Public Relation] --> [COLLEGE = Communication] (confidence: 1.000)  
[STUDENT\_MAJOR = Mass Communication] --> [COLLEGE = Communication] (confidence: 1.000)

1. 100% confidence if student major is 'public relation' or 'mass communication' then college is 'communication'.

[STUDENT\_MAJOR = Applied Sociology] --> [COLLEGE = Arts, Humanities & Social Sci.] (confidence: 1.000)  
[STUDENT\_MAJOR = Sociology] --> [COLLEGE = Arts, Humanities & Social Sci.] (confidence: 1.000)  
[STUDENT\_MAJOR = Education] --> [COLLEGE = Arts, Humanities & Social Sci.] (confidence: 1.000)  
[STUDENT\_LEVEL = Higher Diploma] --> [COLLEGE = Arts, Humanities & Social Sci.] (confidence: 1.000)

1. 100% confidence if student major is 'Applied Sociology' or 'sociology' or 'Education' or 'Higher Diploma' then college is 'Arts, Humanities and Social sci'.

[COLLEGE = Medicine] --> [STUDENT\_LEVEL = Foundation Year] (confidence: 1.000)

1. If college is medicine then student-level is foundation year.

[DIPLOMA\_DESCRIPTION = Bachelor] --> [STUDENT\_LEVEL = Graduate] (confidence: 1.000)  
[STUDENT\_MAJOR = Applied Sociology] --> [STUDENT\_LEVEL = Graduate] (confidence: 1.000)  
[SCHOOL = University of Sharjah] --> [STUDENT\_LEVEL = Graduate] (confidence: 1.000)

1. If Diploma\_Description 'Bachelor' or 'Applied Sociology' or 'Uni of Sharjah' then student\_level is 'Graudate'

[STUDENT\_MAJOR = Jurisprudence & its Foundation] --> [COLLEGE = Shari'a & Islamic Studies] (confidence: 1.000)  
[STUDENT\_MAJOR = Foundations of Religion] --> [COLLEGE = Shari'a & Islamic Studies] (confidence: 1.000)

1. If student\_major is 'Jurisprudence…' or 'Foundation of religion' then college is 'Shari'a & Islamic Studies'.

[COLLEGE = Medicine] --> [STUDENT\_MAJOR = Medicine & Surgery] (confidence: 1.000)

1. If college is medicine then student major is 'Medicine & Surgery'.
2. [COLLEGE = Pharmacy] --> [STUDENT\_MAJOR = Pharmacy] (confidence: 1.000)
3. [COLLEGE = Dentistry] --> [STUDENT\_MAJOR = Dental Surgery] (confidence: 1.000)
4. [SCHOOL = Salma Bint Qais Secondary G] --> [SCHOOL\_CITY = DABA AL HESSEN] (confidence: 1.000)
5. [SCHOOL = Al Shifa Bint Abdulla Ctr] --> [SCHOOL\_CITY = KALBA] (confidence: 1.000)
6. [STUDENT\_LEVEL = Fine Art] --> [COLLEGE = Fine Arts & Design] (confidence: 1.000)
7. [COLLEGE = Fine Arts & Design] --> [STUDENT\_LEVEL = Fine Art] (confidence: 1.000)
8. [SCHOOL = Al Nour Intrnational Pvt B] --> [SCHOOL\_CITY = SHARJAH - EMIRATE] (confidence: 1.000)
9. [STUDENT\_MAJOR = Education] --> [STUDENT\_LEVEL = Higher Diploma] (confidence: 1.000)  
   [STUDENT\_LEVEL = Higher Diploma] --> [STUDENT\_MAJOR = Education] (confidence: 1.000)

[AGE = range1 [-∞ - 20.200], SCHOOL\_CITY = DABA AL HESSEN] --> [GNDR = F] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], SCHOOL = Salma Bint Qais Secondary G] --> [GNDR = F] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], SCHOOL = Jameela Bou Hurraid Sec G] --> [GNDR = F] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], SCHOOL = Bahithat Al Badyeh Secondary G]

[AGE = range1 [-∞ - 20.200], STUDENT\_MAJOR = Sociology] --> [GNDR = F] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], STUDENT\_MAJOR = English Language & Literature] --> [GNDR = F] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], SCHOOL = Fatima Al Zahra' Secondary G.] --> [GNDR = F] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], SCHOOL = Al Reffa'a Secondary G.] --> [GNDR = F] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], SCHOOL = Waset Secondary G] --> [GNDR = F] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], SCHOOL = Rouqaya Sec. Sch. G.] --> [GNDR = F] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], SCHOOL = Al Gobaibah Secondary G.] --> [GNDR = F] (confidence: 1.000)

1. If age <=20 and school is 'Daba al Hessen' or 'Salma Bint Qais…' or 'Jameela bou hur…' or 'Bahithat al…' or 'Al Reff..' or 'Waset sec' or 'Rouqaya..' or 'Al Gobaibah' and student-major is 'Sociology' or 'English language& literature' then gender is female.

[AGE = range1 [-∞ - 20.200], COLLEGE = Law] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], STUDENT\_MAJOR = Law] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], COLLEGE = Engineering] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], COLLEGE = Communication] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], COLLEGE = Arts, Humanities & Social Sci.] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], COLLEGE = Shari'a & Islamic Studies] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], COLLEGE = Business Administration] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], STUDENT\_MAJOR = Public Relation] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], STUDENT\_MAJOR = Mass Communication] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], COLLEGE = Health Sciences] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], COLLEGE = Pharmacy] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], STUDENT\_MAJOR = Jurisprudence & its Foundation] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], COLLEGE = Sciences] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], STUDENT\_MAJOR = Sociology] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)

1. If age <=20 and college is 'Law' or 'Engineering' or 'Communication' or 'Arts, Humanities & Social sci' or 'shariah & Islamic studies' or 'Business adminis..' or 'Health sciences' or 'pharmacy' or 'sciences' and student-major is 'law' or 'Public relations' or 'Mass communication' or 'Sociology' or ' Jurisprudence & its Foundation' then **student-level is 'Undergraduate'**.

[ADMIT = Fall 2011-2012, COLLEGE = Medicine] -->[AGE = range1 [-∞ - 20.200]] (confidence: 1.000)

1. If student got admission in Fall 2011-2012 and college is Medicine then they are under 20 age.
2. [AGE = range1 [-∞ - 20.200], SCHOOL = Al Ahlia Pvt Secondary] --> [DIPLOMA\_DESCRIPTION = Secondary School -Scientific] (confidence: 1.000)

[AGE = range1 [-∞ - 20.200], SCHOOL = Al Sho'ala Private Sch.] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], SCHOOL = Jameela Bou Hurraid Sec G] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], SCHOOL = Bahithat Al Badyeh Secondary G] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], SCHOOL = Fatima Al Zahra' Secondary G.] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], SCHOOL = Al Reffa'a Secondary G.] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], SCHOOL = Al Ahlia Pvt Secondary] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], SCHOOL = Waset Secondary G] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], SCHOOL = Al Gobaibah Secondary G.] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)

1. If student age is<=20 and school is 'Al Sho'ala Private' or 'Jameela bou' or 'Bhithat Al' or 'Fatima Alzahra' or 'Al Reffa'a secondary' or 'Al Ahlia pvt..' or 'Waset second..' or 'Al Gobaia..' **then school city is Sharjah**.
2. [AGE = range1 [-∞ - 20.200], STUDENT\_LEVEL = Intensive English] --> [COLLEGE = No College Designated] (confidence: 1.000)

[COLLEGE = No College Designated, STUDENT\_MAJOR = Sustainable/Renewable Enrg Eng] --> [AGE = range1 [-∞ - 20.200]] (confidence: 1.000)

1. If student don't deignate any college and student major is 'sustainable/renewable Enrg Eng' then student age is<=20
2. [STUDENT\_LEVEL = Intensive English, STUDENT\_MAJOR = Sustainable/Renewable Enrg Eng] --> [AGE = range1 [-∞ - 20.200]] (confidence: 1.000)  
     
   [AGE = range1 [-∞ - 20.200], SCHOOL = Rouqaya Sec. Sch. G.] --> [SCHOOL\_CITY = Sharjah] (confidence: 1.000)  
   [AGE = range1 [-∞ - 20.200], SCHOOL = Al Ma'arifa Int. Private Sch.] --> [SCHOOL\_CITY = Sharjah] (confidence: 1.000)
3. If student age is<=20 and school is 'Rouqaya sec….' or 'Al Ma'arifa…' then school-city is 'Sharjah'.

[DIPLOMA\_DESCRIPTION = High School -Scientific, STUDENT\_MAJOR = Medicine & Surgery] --> [AGE = range1 [-∞ - 20.200]] (confidence: 1.000)  
[DIPLOMA\_DESCRIPTION = High School -Scientific, COLLEGE = Medicine] --> [AGE = range1 [-∞ - 20.200]] (confidence: 1.000)

1. If student diploma\_description is 'High school-scientific' and student\_major is 'Medicine & Surgery' or college is 'Medicine' then student age is<=20
2. [AGE = range1 [-∞ - 20.200], STUDENT\_LEVEL = Diploma] --> [COLLEGE = Community] (confidence: 1.000)  
   [AGE = range1 [-∞ - 20.200], COLLEGE = Community] --> [STUDENT\_LEVEL = Diploma] (confidence: 1.000)
3. [AGE = range1 [-∞ - 20.200], STUDENT\_MAJOR = Business Administration] --> [STUDENT\_LEVEL = Diploma] (confidence: 1.000)
4. [AGE = range1 [-∞ - 20.200], STUDENT\_MAJOR = Business Administration] --> [COLLEGE = Community] (confidence: 1.000)
5. [AGE = range1 [-∞ - 20.200], COLLEGE = Business Administration] --> [STUDENT\_MAJOR = Common Business Program] (confidence: 1.000)
6. [AGE = range1 [-∞ - 20.200], STUDENT\_MAJOR = Public Relation] --> [COLLEGE = Communication] (confidence: 1.000)
7. [AGE = range1 [-∞ - 20.200], STUDENT\_MAJOR = Mass Communication] --> [COLLEGE = Communication] (confidence: 1.000)
8. [AGE = range1 [-∞ - 20.200], STUDENT\_MAJOR = Sociology] --> [COLLEGE = Arts, Humanities & Social Sci.] (confidence: 1.000)

[AGE = range1 [-∞ - 20.200], COLLEGE = Medicine] --> [STUDENT\_LEVEL = Foundation Year] (confidence: 1.000)

[AGE = range1 [-∞ - 20.200], COLLEGE = Dentistry] --> [STUDENT\_LEVEL = Foundation Year] (confidence: 1.000)

1. If student age is<=20 and college is 'Medicine' or 'Dentistry' then student-level is foundation year.
2. [AGE = range1 [-∞ - 20.200], STUDENT\_MAJOR = Jurisprudence & its Foundation] --> [COLLEGE = Shari'a & Islamic Studies] (confidence: 1.000)
3. [AGE = range1 [-∞ - 20.200], COLLEGE = Medicine] --> [STUDENT\_MAJOR = Medicine & Surgery] (confidence: 1.000)
4. [AGE = range1 [-∞ - 20.200], COLLEGE = Pharmacy] --> [STUDENT\_MAJOR = Pharmacy] (confidence: 1.000)
5. [AGE = range1 [-∞ - 20.200], COLLEGE = Dentistry] --> [STUDENT\_MAJOR = Dental Surgery] (confidence: 1.000)
6. [AGE = range1 [-∞ - 20.200], SCHOOL\_CITY = DABA AL HESSEN] --> [SCHOOL = Salma Bint Qais Secondary G] (confidence: 1.000)  
   [AGE = range1 [-∞ - 20.200], SCHOOL = Salma Bint Qais Secondary G] --> [SCHOOL\_CITY = DABA AL HESSEN] (confidence: 1.000)

[AGE = range1 [-∞ - 20.200], STUDENT\_LEVEL = Fine Art] --> [COLLEGE = Fine Arts & Design] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], COLLEGE = Fine Arts & Design] --> [STUDENT\_LEVEL = Fine Art] (confidence: 1.000)

1. If student-level is 'Fine Art' and age <=20 then college is 'Fine Arts& Design' and vice versa.

[AGE = range1 [-∞ - 20.200], SCHOOL\_CITY = SHARJAH - EMIRATE] --> [SCHOOL = Al Nour International Pvt B] (confidence: 1.000)  
[AGE = range1 [-∞ - 20.200], SCHOOL = Al Nour Intrnational Pvt B] --> [SCHOOL\_CITY = SHARJAH - EMIRATE] (confidence: 1.000)

1. If student age<=20 and school city is 'Sharjah-Emirate' then school is 'Al Nour International Pvt B'.

[GNDR = F, STUDENT\_MAJOR = Law] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[GNDR = F, COLLEGE = Communication] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[GNDR = F, STUDENT\_MAJOR = Public Relation] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[GNDR = F, STUDENT\_MAJOR = Mass Communication] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[GNDR = F, COLLEGE = Health Sciences] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[GNDR = F, COLLEGE = Pharmacy] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[STUDENT\_MAJOR = Sociology] --> [GNDR = F, STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)  
[GNDR = F, STUDENT\_MAJOR = Sociology] --> [STUDENT\_LEVEL = Undergraduate] (confidence: 1.000)

1. There is 100% confidence if Student is female and college is 'Communication' or 'Health Sciences' or 'Pharmacy' OR student\_major is 'LAW' or 'Public relation' or 'Mass Communication' or 'Sociology'

**Then student-level is 'Undergraduate'.**

[STUDENT\_LEVEL = Undergraduate, STUDENT\_MAJOR = Sociology] --> [GNDR = F] (confidence: 1.000)

[DIPLOMA\_DESCRIPTION = Secondary School -Literature, SCHOOL = Jameela Bou Hurraid Sec G] --> [GNDR = F] (confidence: 1.000)  
[DIPLOMA\_DESCRIPTION = Secondary School -Literature, SCHOOL = Bahithat Al Badyeh Secondary G] --> [GNDR = F] (confidence: 1.000)  
[DIPLOMA\_DESCRIPTION = Secondary School -Literature, STUDENT\_MAJOR = Sociology] --> [GNDR = F] (confidence: 1.000)  
[DIPLOMA\_DESCRIPTION = Secondary School -Literature, SCHOOL = Al Shifa Bint Abdulla Ctr] --> [GNDR = F] (confidence: 1.000)

1. If student-level is 'Undergraduate' and student-major is 'Sociology' Or diploma\_descr is 'Secondary School Literature' and school is 'Jameela Bou Hurraid..' oR diploma\_descr is 'Secondary School Literature' and school is 'Bahithat Al…' OR student-major is sociology or school is Al Shifa Bint Ab.. **then student is Female.**
2. [GNDR = F, SCHOOL = Al Sho'ala Private Sch.] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)
3. [SCHOOL = Jameela Bou Hurraid Sec G] --> [GNDR = F, SCHOOL\_CITY = SHARJAH] (confidence: 1.000)
4. [GNDR = F, SCHOOL = Jameela Bou Hurraid Sec G] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)
5. [SCHOOL\_CITY = SHARJAH, SCHOOL = Jameela Bou Hurraid Sec G] --> [GNDR = F] (confidence: 1.000)
6. [SCHOOL = Bahithat Al Badyeh Secondary G] --> [GNDR = F, SCHOOL\_CITY = SHARJAH] (confidence: 1.000)
7. [GNDR = F, SCHOOL = Bahithat Al Badyeh Secondary G] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)
8. [SCHOOL\_CITY = SHARJAH, SCHOOL = Bahithat Al Badyeh Secondary G] --> [GNDR = F] (confidence: 1.000)
9. [SCHOOL = Fatima Al Zahra' Secondary G.] --> [GNDR = F, SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
   [GNDR = F, SCHOOL = Fatima Al Zahra' Secondary G.] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
   [SCHOOL\_CITY = SHARJAH, SCHOOL = Fatima Al Zahra' Secondary G.] --> [GNDR = F] (confidence: 1.000)
10. [SCHOOL = Al Reffa'a Secondary G.] --> [GNDR = F, SCHOOL\_CITY = SHARJAH] (confidence: 1.000)  
    [GNDR = F, SCHOOL = Al Reffa'a Secondary G.] --> [SCHOOL\_CITY = SHARJAH] (confidence: 1.000)

[STUDENT\_MAJOR = Education] --> [GNDR = F, COLLEGE = Arts, Humanities & Social Sci.] (Confidence: 1.000)  
[GNDR = F, STUDENT\_MAJOR = Education] --> [COLLEGE = Arts, Humanities & Social Sci.] (Confidence: 1.000)  
[COLLEGE = Arts, Humanities & Social Sci., STUDENT\_MAJOR = Education] --> [GNDR = F] (confidence: 1.000)

1. If student is female and student-major is 'Education' then student college is 'Arts, Humanities and Social Sci' and vice versa.
2. [GNDR = F, DIPLOMA\_DESCRIPTION = Bachelor] --> [STUDENT\_LEVEL = Graduate] (confidence: 1.000)
3. There is 83% confidence if COLLEGE = Arts, Humanities & Social Sci., AGE is 29-38 then student level would be Graduate.
4. There is 98% confidence if COLLEGE = Arts, Humanities & Social Sci., Gendre is female then student level would be Graduate.

Rule's premises and conclusion can easily be read by following way. Here then part (conclusion) is table heading.

|  |
| --- |
| **Student Level = Graduate IF** |
| STUDENT\_MAJOR = Applied Sociology Or |
| DIPLOMA\_DESCRIPTION = Bachelor Or |
| SCHOOL = University of Sharjah OR |
| GNDR = F, DIPLOMA\_DESCRIPTION = Bachelor OR  GNDR = M, DIPLOMA\_DESCRIPTION = Bachelor OR |
| GNDR = M, STUDENT\_MAJOR = Applied Sociology |
| ADMIT = Fall 2010-2011, DIPLOMA\_DESCRIPTION = Bachelor |
| SCHOOL\_CITY = SHARJAH, DIPLOMA\_DESCRIPTION = Bachelor |
| SCHOOL\_CITY = SHARJAH, SCHOOL = University of Sharjah |
| AGE = range [20- 29], DIPLOMA\_DESCRIPTION = Bachelor |
| AGE = range [20 - 29], STUDENT\_MAJOR = Applied Sociolog |
| AGE = range [20 - 29], SCHOOL = University of Sharjah |
| ADMIT = Spring 2011-2012, DIPLOMA\_DESCRIPTION = Bachelor |
| ADMIT = Spring 2011-2012, STUDENT\_MAJOR = Applied Sociology |
| COLLEGE = Arts, Humanities & Social Sci., DIPLOMA\_DESCRIPTION = Bachelor |
| AGE = range [29 - 38], DIPLOMA\_DESCRIPTION = Bachelor |
| GNDR = M, SCHOOL\_CITY = SHARJAH, DIPLOMA\_DESCRIPTION = Bachelor |
| GNDR = M, AGE = range [20 - 29], DIPLOMA\_DESCRIPTION = Bachelor |
| SCHOOL\_CITY = SHARJAH, AGE = range[20 - 29], DIPLOMA\_DESCRIPTION = Bachelor |
| SCHOOL\_CITY = SHARJAH, AGE = range [20 - 29], SCHOOL = University of Sharjah |
| AGE = range [20 - 29], COLLEGE = Arts, Humanities & Social Sci., STUDENT\_MAJOR = AppliedSociology |
| ADMIT = Spring 2011-2012, COLLEGE = Arts, Humanities & Social Sci., STUDENT\_MAJOR = Applied Sociology |
| GNDR = M, COLLEGE = Arts, Humanities & Social Sci., DIPLOMA\_DESCRIPTION = Bachelor, STUDENT\_MAJOR = Applied Sociology |

|  |
| --- |
| STUDENT\_MAJOR = Medicine & Surgery IF |
| COLLEGE = Medicine |
| ADMIT = Fall 2011-2012, COLLEGE = Medicine |
| DIPLOMA\_DESCRIPTION = Secondary School -Scientific, COLLEGE = Medicine |
| DIPLOMA\_DESCRIPTION = High School -Scientific, COLLEGE = Medicine |
| STUDENT\_LEVEL = Foundation Year, COLLEGE = Medicine |
| AGE = range [-∞ - 20], GNDR = F, COLLEGE = Medicine |
| AGE = range1 [-∞ - 20], ADMIT = Fall 2010-2011, COLLEGE = Medicine |
| AGE = range1 [-∞ - 20], GNDR = M, COLLEGE = Medicine |
| AGE = range1 [-∞ - 20.200], STUDENT\_LEVEL = Foundation Year, COLLEGE = Medicine |
| GNDR = M, STUDENT\_LEVEL = Foundation Year, COLLEGE = Medicine |
| DIPLOMA\_DESCRIPTION = High School -Scientific, STUDENT\_LEVEL = Foundation Year, COLLEGE = Medicine |
| AGE = range1 [-∞ - 20.200], ADMIT = Fall 2010-2011, STUDENT\_LEVEL = Foundation Year, COLLEGE = Medicine |
| GNDR = F, DIPLOMA\_DESCRIPTION = Secondary School -Scientific, STUDENT\_LEVEL = Foundation Year, COLLEGE = Medicine |
| AGE = range1 [-∞ - 20.200], DIPLOMA\_DESCRIPTION = Secondary School -Scientific, ADMIT = Fall 2010-2011, STUDENT\_LEVEL = Foundation Year, COLLEGE = Medicine |

|  |
| --- |
| GNDR = Male IF |
| COLLEGE = Engineering, STUDENT\_MAJOR = Civil Engineering |
| DIPLOMA\_DESCRIPTION = Secondary School -Scientific, ADMIT = Fall 2010-2011, STUDENT\_MAJOR = Civil Engineering |
| ADMIT = Fall 2010-2011, STUDENT\_MAJOR = Civil Engineering |
| STUDENT\_LEVEL = Undergraduate, STUDENT\_MAJOR = Civil Engineering |
| STUDENT\_LEVEL = Undergraduate, COLLEGE = Engineering, STUDENT\_MAJOR = Civil Engineering |
| STUDENT\_LEVEL = Undergraduate, DIPLOMA\_DESCRIPTION = Secondary School -Scientific, STUDENT\_MAJOR = Civil Engineering |
| DIPLOMA\_DESCRIPTION = Secondary School -Scientific, COLLEGE = Engineering, STUDENT\_MAJOR = Civil Engineering |
| STUDENT\_LEVEL = Undergraduate, DIPLOMA\_DESCRIPTION = Secondary School -Scientific, COLLEGE = Engineering, STUDENT\_MAJOR = Civil Engineering |
| AGE = range1 [-∞ - 20.200], STUDENT\_LEVEL = Undergraduate, COLLEGE = Engineering, STUDENT\_MAJOR = Civil Engineering |
| AGE = range1 [-∞ - 20.200], ADMIT = Fall 2011-2012, DIPLOMA\_DESCRIPTION = Secondary School -Scientific, STUDENT\_MAJOR = Civil Engineering |
| AGE = range1 [-∞ - 20.200], DIPLOMA\_DESCRIPTION = Secondary School -Scientific, STUDENT\_LEVEL = Intensive English, STUDENT\_MAJOR = Civil Engineering |
| AGE = range1 [-∞ - 20.200], DIPLOMA\_DESCRIPTION = Secondary School -Scientific, COLLEGE = No College Designated, STUDENT\_LEVEL = Intensive English, STUDENT\_MAJOR = Civil Engineering |
| SCHOOL\_CITY = SHARJAH, STUDENT\_MAJOR = Civil Engineering |
| STUDENT\_LEVEL = Undergraduate, SCHOOL\_CITY = KUWAIT |
| DIPLOMA\_DESCRIPTION = Secondary School -Scientific, SCHOOL\_CITY = SHARJAH, STUDENT\_MAJOR = Civil Engineering |
| SCHOOL\_CITY = SHARJAH, SCHOOL = Al Khalil Bin Ahmed Secondary |

|  |
| --- |
| COLLEGE = Communication IF |
| AGE = range [-∞ - 20], GNDR = F, STUDENT\_LEVEL = Undergraduate, DIPLOMA\_DESCRIPTION = Secondary School -Literature, SCHOOL\_CITY = SHARJAH, STUDENT\_MAJOR = Public Relation |
| AGE = range1 [-∞ - 20], STUDENT\_LEVEL = Undergraduate, ADMIT = Fall 2011-2012, STUDENT\_MAJOR = Mass Communication |
| AGE = range [-∞ - 20], ADMIT = Fall 2010-2011, STUDENT\_MAJOR = Mass Communication |
| AGE = range [-∞ - 20], STUDENT\_LEVEL = Undergraduate, ADMIT = Fall 2011-2012, STUDENT\_MAJOR = Public Relation |
| AGE = range [-∞ - 20], ADMIT = Fall 2010-2011, STUDENT\_MAJOR = Public Relation |
| AGE = range[-∞ - 20], DIPLOMA\_DESCRIPTION = Secondary School -Literature, STUDENT\_MAJOR = Mass Communication |
| AGE = range[-∞ - 20], DIPLOMA\_DESCRIPTION = Secondary School -Literature, STUDENT\_MAJOR = Public Relation |
| GNDR = F, SCHOOL\_CITY = Sharjah, STUDENT\_MAJOR = Public Relation |
| AGE = range1 [-∞ - 20.200], GNDR = F, STUDENT\_LEVEL = Undergraduate, ADMIT = Fall 2011-2012, DIPLOMA\_DESCRIPTION = Secondary School -Literature, STUDENT\_MAJOR = Public Relation |
| GNDR = F, STUDENT\_LEVEL = Undergraduate, ADMIT = Fall 2010-2011, STUDENT\_MAJOR = Mass Communication |
| GNDR = F, STUDENT\_LEVEL = Undergraduate, ADMIT = Fall 2011-2012, DIPLOMA\_DESCRIPTION = Secondary School -Literature, STUDENT\_MAJOR = Public Relation |
| GNDR = F, STUDENT\_LEVEL = Undergraduate, ADMIT = Fall 2011-2012, DIPLOMA\_DESCRIPTION = Secondary School -Literature, STUDENT\_MAJOR = Mass Communication |

|  |
| --- |
| STUDENT\_MAJOR = Common Business Program IF |
| GNDR = F, STUDENT\_LEVEL = Undergraduate, COLLEGE = Business Administration |
| AGE = range [-∞ - 20], GNDR = F, STUDENT\_LEVEL = Undergraduate, COLLEGE = Business Administration |
| AGE = range [-∞ - 20], STUDENT\_LEVEL = Undergraduate, ADMIT = Fall 2011-2012, COLLEGE = Business Administration |
| AGE = range [-∞ - 20.200], GNDR = M, COLLEGE = Business Administration |
| STUDENT\_LEVEL = Undergraduate, ADMIT = Fall 2011-2012, COLLEGE = Business Administration |
| AGE = range [-∞ - 20], STUDENT\_LEVEL = Undergraduate, GNDR = M, COLLEGE = Business Administration |
| GNDR = M, DIPLOMA\_DESCRIPTION = Secondary School -Literature, STUDENT\_LEVEL = Intensive English |
| AGE = range [-∞ - 20], STUDENT\_LEVEL = Undergraduate, ADMIT = Fall 2010-2011, COLLEGE = Business Administration |
| GNDR = M, DIPLOMA\_DESCRIPTION = Secondary School -Literature, COLLEGE = No College Designated, STUDENT\_LEVEL = Intensive English |

|  |
| --- |
| COLLEGE = Arts, Humanities & Social Sci. IF |
| DIPLOMA\_DESCRIPTION = Bachelor, STUDENT\_MAJOR = Applied Sociology |
| AGE = range1 [-∞ - 20.200], GNDR = F, STUDENT\_LEVEL = Undergraduate, DIPLOMA\_DESCRIPTION = Secondary School -Literature, STUDENT\_MAJOR = Sociology |
| GNDR = F, STUDENT\_MAJOR = Education, STUDENT\_LEVEL = Higher Diploma |
| GNDR = M, STUDENT\_LEVEL = Graduate, STUDENT\_MAJOR = Applied Sociology |
| GNDR = M, DIPLOMA\_DESCRIPTION = Bachelor, STUDENT\_MAJOR = Applied Sociology |
| AGE = range [20 - 29], STUDENT\_LEVEL = Graduate, STUDENT\_MAJOR = Applied Sociology |
| ADMIT = Spring 2011-2012, STUDENT\_LEVEL = Graduate, STUDENT\_MAJOR = Applied Sociology |
| GNDR = M, STUDENT\_LEVEL = Graduate, DIPLOMA\_DESCRIPTION = Bachelor, STUDENT\_MAJOR = Applied Sociology |
| AGE = range [-∞ - 20], GNDR = F, DIPLOMA\_DESCRIPTION = Secondary School -Literature, STUDENT\_MAJOR = Sociology |
| AGE = range [-∞ - 20], STUDENT\_LEVEL = Undergraduate, DIPLOMA\_DESCRIPTION = Secondary School -Literature, STUDENT\_MAJOR = Sociology |
| GNDR = M, STUDENT\_LEVEL = Graduate, DIPLOMA\_DESCRIPTION = Bachelor, STUDENT\_MAJOR = Applied Sociology |
| ADMIT = Spring 2011-2012, STUDENT\_LEVEL = Graduate, STUDENT\_MAJOR = Applied Sociology |
| AGE = range2 [20 - 29], STUDENT\_LEVEL = Graduate, STUDENT\_MAJOR = Applied Sociology |

|  |
| --- |
| **COLLEGE = Business Administration if** |
| AGE = range1 [-∞ - 20.200], STUDENT\_LEVEL = Undergraduate, STUDENT\_MAJOR = Common Business Program |
| AGE = range1 [-∞ - 20.200], GNDR = F, STUDENT\_LEVEL = Undergraduate, STUDENT\_MAJOR = Common Business Program |
| AGE = range1 [-∞ - 20.200], STUDENT\_LEVEL = Undergraduate, ADMIT = Fall 2011-2012, STUDENT\_MAJOR = Common Business Program |
| STUDENT\_LEVEL = Undergraduate, DIPLOMA\_DESCRIPTION = Secondary School -Scientific, STUDENT\_MAJOR = Common Business Program |
| AGE = range1 [-∞ - 20], STUDENT\_LEVEL = Undergraduate, GNDR = M, STUDENT\_MAJOR = Common Business Program |
| AGE = range1 [-∞ - 20], STUDENT\_LEVEL = Undergraduate, ADMIT = Fall 2010-2011, STUDENT\_MAJOR = Common Business Program |

|  |
| --- |
| COLLEGE = Community IF |
| GNDR = F, DIPLOMA\_DESCRIPTION = Secondary School -Literature, AGE = range [29 - 38] |
| GNDR = F, AGE = range [20 - 29], STUDENT\_MAJOR = Business Administration |
| GNDR = F, ADMIT = Spring 2011-2012, STUDENT\_MAJOR = Business Administration |
| ADMIT = Fall 2010-2011, AGE = range [20 - 29], STUDENT\_MAJOR = Business Administration |
| GNDR = F, ADMIT = Fall 2010-2011, STUDENT\_MAJOR = Business Administration |
| GNDR = F, STUDENT\_LEVEL = Diploma |
| GNDR = F, STUDENT\_MAJOR = Information Technology |
| ADMIT = Fall 2011-2012, STUDENT\_MAJOR = Information Technology |
| GNDR = F, DIPLOMA\_DESCRIPTION = Secondary School -Literature, STUDENT\_LEVEL = Diploma |
| AGE = range [-∞ - 20], GNDR = F, ADMIT = Fall 2010-2011, STUDENT\_LEVEL = Diploma |
| DIPLOMA\_DESCRIPTION = Secondary School -Literature, STUDENT\_MAJOR = Information Technology |
| SCHOOL\_CITY = SHARJAH, STUDENT\_LEVEL = Diploma |
| AGE = range [20 - 29], STUDENT\_MAJOR = Information Technology |
| GNDR = F, ADMIT = Fall 2011-2012, DIPLOMA\_DESCRIPTION = Secondary School -Literature, AGE = range2 [20 - 29], STUDENT\_LEVEL = Diploma |
| STUDENT\_LEVEL = Diploma, SCHOOL\_CITY = KALBA |
| SCHOOL\_CITY = SHARJAH, STUDENT\_LEVEL = Diploma, SCHOOL = Jameela Bou Hurraid Sec G |
| STUDENT\_LEVEL = Diploma, SCHOOL\_CITY = KALBA, SCHOOL = Al Shifa Bint Abdulla Ctr |
| GNDR = F, DIPLOMA\_DESCRIPTION = Secondary School -Scientific, STUDENT\_LEVEL = Diploma |
| GNDR = F, SCHOOL\_CITY = SHARJAH, STUDENT\_MAJOR = Information Technology |
| GNDR = F, STUDENT\_LEVEL = Diploma, ADMIT = Spring 2011-2012 |
| GNDR = F, DIPLOMA\_DESCRIPTION = Secondary School -Literature, SCHOOL\_CITY = SHARJAH, STUDENT\_MAJOR = Business Administration |
| DIPLOMA\_DESCRIPTION = Secondary School -Literature, SCHOOL\_CITY = SHARJAH, STUDENT\_LEVEL = Diploma, STUDENT\_MAJOR = Information Technology |
| DIPLOMA\_DESCRIPTION = Secondary School -Literature, AGE = range[20 - 29], STUDENT\_LEVEL = Diploma, STUDENT\_MAJOR = Business Administration |
| GNDR = F, ADMIT = Fall 2010-2011, DIPLOMA\_DESCRIPTION = Secondary School -Literature, STUDENT\_LEVEL = Diploma, STUDENT\_MAJOR = Business Administration |
| GNDR = F, ADMIT = Fall 2011-2012, DIPLOMA\_DESCRIPTION = Secondary School -Literature, AGE = range [20 - 29], STUDENT\_LEVEL = Diploma |
| GNDR = F, DIPLOMA\_DESCRIPTION = Secondary School -Literature, AGE = range[20 - 29], STUDENT\_LEVEL = Diploma, STUDENT\_MAJOR = Information Technology |
| ADMIT = Fall 2010-2011, DIPLOMA\_DESCRIPTION = Secondary School -Literature, AGE = range [20 - 29], STUDENT\_LEVEL = Diploma |

|  |
| --- |
| COLLEGE = Sciences IF |
| AGE = range [-∞ - 20], GNDR = F, STUDENT\_LEVEL = Undergraduate, STUDENT\_MAJOR = Biotechnology |

## Classification by Decision Tree Induction

### Tree Analysis

Conducting analysis of decision making under uncertainty using decision trees serves several purposes.

* First, a decision tree is a visual representation of a decision situation (and hence aids communication).
* Second, the branches of a tree explicitly show all those factors within the analysis that are considered relevant to the decision (and implicitly those that are not).
* Third, and more subtly, a decision tree generally captures the idea that if different decisions were to be taken then the structural nature of a situation (and hence of the model) may have changed dramatically.
* Fourth, and arguably the most powerful, a decision tree allows for forward and backward calculation paths to happen and hence the choice of the correct decision to take (optimality of decision making, or optimal exercise if embedded real options) is made automatically.

# 

### Information Gain

#### Class attribute: Student-Level

accuracy: 92.56%

classification\_error: 7.44%

Confustion Matrix same as above.

#### Class attribute: Citizenship

accuracy: 77.84%

classification\_error: 22.16%

weighted\_mean\_recall: 40.62%

weighted\_mean\_precision: 62.77%

#### Class attribute: Gender

Accuracy, classification error, recall and precision is same as in case of Gini Index and GainRatio for class attribute Gender.

### Gain Ratio

#### Class attribute: Student-level

There are 9 classes for this attribute. After applying all operators as shown in above fig. I run the system and following results displayed.

Accuracy: 92.56%

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Graduate | Higher Diploma | Under  graduate | Diploma | Foundation Year | Fine Art | Intensive English | Dentistry Training | Doctorate | Precision |
| pred. Graduate | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00% |
| pred. Higher Diploma | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00% |
| pred. Undergraduate | 446 | 81 | 3776 | 0 | 0 | 0 | 0 | 0 | 1 | 87.73% |
| pred. Diploma | 0 | 0 | 0 | 776 | 0 | 0 | 0 | 0 | 0 | 100.00% |
| pred. Foundation Year | 0 | 0 | 0 | 0 | 514 | 0 | 0 | 1 | 0 | 99.81% |
| pred. Fine Art | 0 | 0 | 0 | 0 | 0 | 148 | 0 | 0 | 0 | 100.00% |
| pred. Intensive English | 1 | 0 | 35 | 0 | 0 | 0 | 1819 | 0 | 0 | 98.06% |
| pred. Dentistry Training | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00% |
| pred. Doctorate | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00% |
| Recall | 0.00% | 0.00% | 99.08% | 100.00% | 100.00% | 100.00% | 100.00% | 0.00% | 0.00% |  |

classification\_error: 7.44%

#### Class attribute: Citizenship

**There are 12 different values for this attribute so there are 12 classes. Performance measures are given below.**

accuracy: 59.66%

classification\_error: 40.34%

weighted\_mean\_precision: 51.82%

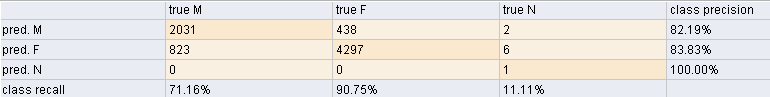
weighted\_mean\_recall: 27.89%

#### Class attribute: Gender

There are three classes as this attribute having three distince values. Male, Female, Not known/Not available. Performance measures are given below:

accuracy: 83.30%

Confusion matrix is:



classification\_error: 16.70%

weighted\_mean\_recall: 57.67%

weighted\_mean\_precision: 88.67%

### Gini Index

#### Class attribute: Student-Level

accuracy: 92.56%

classification\_error: 7.44%

Confustion Matrix same as above.

#### Class attribute: Citizenship

accuracy: 77.84%

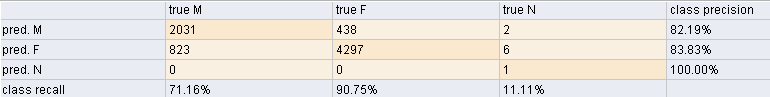
classification\_error: 22.16%

weighted\_mean\_recall: 40.62%

weighted\_mean\_precision: 62.77%

#### Class attribute: Gender

accuracy: 83.30%



classification\_error: 16.70%

### Accuracy

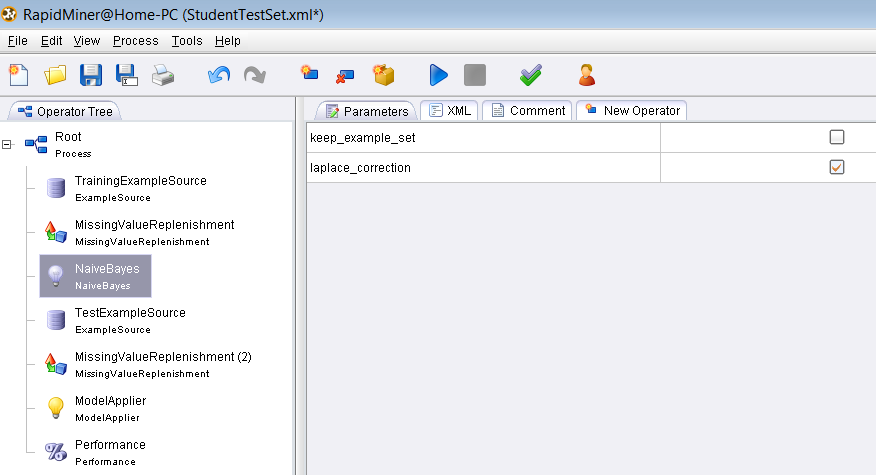
#### Class attribute: Student Level

accuracy: 62.31% , classification\_error: 37.69%

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | true Graduate | true Higher Diploma | true Undergraduate | true Diploma | true Foundation Year | true Fine Art | true Intensive English | true Dentistry Training | true Doctorate | class precision |
| pred. Graduate | 352 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100.00% |
| pred. Higher Diploma | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00% |
| pred. Undergraduate | 95 | 81 | 3455 | 288 | 427 | 119 | 1315 | 1 | 1 | 59.75% |
| pred. Diploma | 0 | 0 | 165 | 463 | 10 | 3 | 88 | 0 | 0 | 63.51% |
| pred. Foundation Year | 0 | 0 | 26 | 1 | 51 | 3 | 14 | 0 | 0 | 53.68% |
| pred. Fine Art | 0 | 0 | 2 | 0 | 1 | 14 | 3 | 0 | 0 | 70.00% |
| pred. Intensive English | 0 | 0 | 163 | 24 | 25 | 9 | 399 | 0 | 0 | 64.35% |
| pred. Dentistry Training | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00% |
| pred. Doctorate | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00% |
| class recall | 78.75% | 0.00% | 90.66% | 59.66% | 9.92% | 9.46% | 21.94% | 0.00% | 0.00% |  |

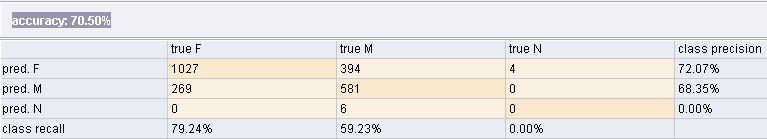
## NaiveBayes Classifier

Here 70% data is taken as training set and then NaiveBayes applied on train set and on test data remaining 30%data from student dataset is taken and performance is measured as follow.



### Class attribute: Gender

Accuracy: 70.50%



Weighted Mean Recall: 69.23%

Weighted Mean Precision: 70.21%

### Class attribute: Citizenship

Accuracy: 99.47%

Here total attribute values of class variable are 12 so confusion matrix can't be displayed but Recall and precision values are given:

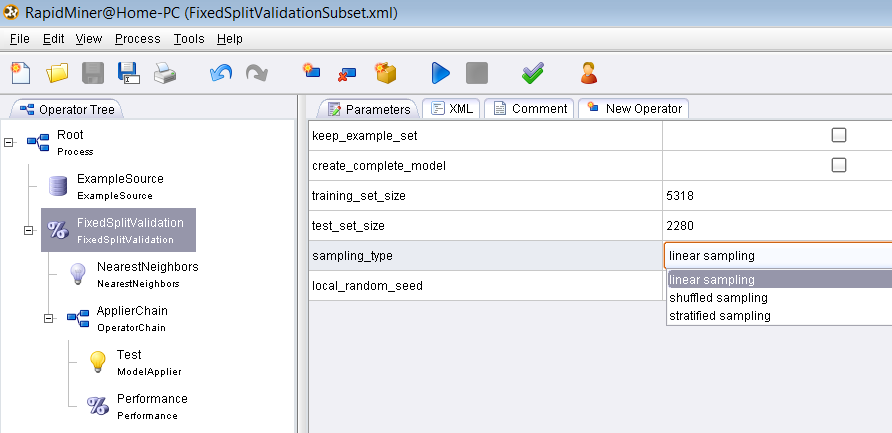
Weighted Mean Recall: 99.89%

Weighted Mean Precision: 98.31%

## *k*-Nearest-Neighbor Classifiers

**Fixed Split validation**

The FixedSplitValidation uses a fixed number of examples for either the training or the test set. This might be useful if both the training and the test data are given in the same data file and the split point is known beforehand.



### Linear Sampling

# Accuracy: 75.26%

#### Class attribute: Student Level

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | true Graduate | true Higher Diploma | true Undergraduate | true Diploma | true Foundation Year | true Fine Art | true Intensive English | true Dentistry Training | true Doctorate | class precision |
| pred. Graduate | 177 | 0 | 62 | 8 | 0 | 0 | 11 | 0 | 0 | 68.60% |
| pred. Higher Diploma | 23 | 0 | 33 | 31 | 0 | 0 | 7 | 0 | 0 | 0.00% |
| pred. Undergraduate | 17 | 0 | 896 | 2 | 0 | 1 | 152 | 0 | 0 | 83.90% |
| pred. Diploma | 4 | 0 | 5 | 134 | 0 | 0 | 2 | 0 | 0 | 92.41% |
| pred. Foundation Year | 1 | 0 | 31 | 1 | 0 | 0 | 65 | 0 | 0 | 0.00% |
| pred. Fine Art | 0 | 0 | 5 | 0 | 0 | 6 | 12 | 0 | 0 | 26.09% |
| pred. Intensive English | 5 | 0 | 83 | 0 | 0 | 0 | 503 | 0 | 0 | 85.11% |
| pred. Dentistry Training | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00% |
| pred. Doctorate | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00% |
| class recall | 77.29% | 0.00% | 80.29% | 76.14% | 0.00% | 85.71% | 66.89% | 0.00% | 0.00% |  |

#### Class attribute: Citizenship

**Accuracy: 55.70%**

Confusion matrix is having 12 classes so can't be displayed here but Precision and recall is given:

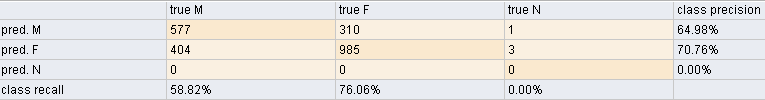
**Weighted Mean Recall: 77.26%**

**Weighted Mean Precision: 71.22%**

#### Class attribute: Gender

Accuracy: 68.51%

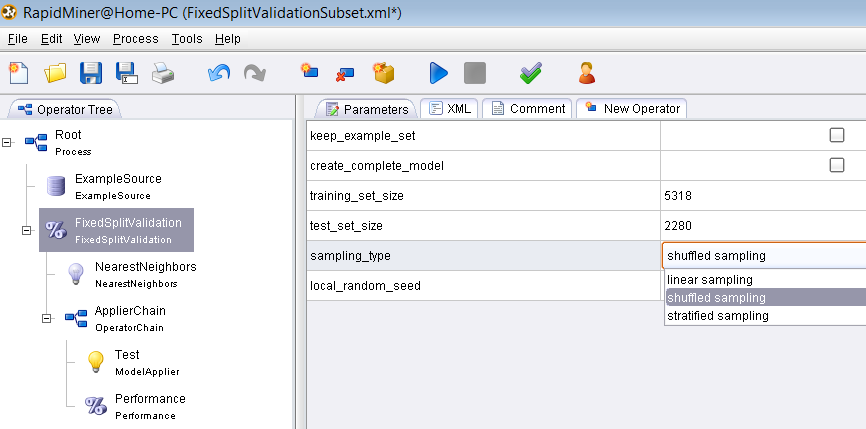
Confusion Matrix having Class recall and precision is as follow:



Weighted Mean Recall: 67.44%

Weighted Mean Precision:67.87%

### Shuffled Sampling



#### Class attribute: Student Level

# Accuracy: 88.46%

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | true Graduate | true Higher Diploma | true Undergraduate | true Diploma | true Foundation Year | true Fine Art | true Intensive English | true Dentistry Training | true Doctorate | class precision |
| pred. Graduate | 118 | 5 | 30 | 6 | 0 | 1 | 3 | 0 | 0 | 72.39% |
| pred. Higher Diploma | 9 | 25 | 15 | 6 | 0 | 0 | 1 | 0 | 0 | 44.64% |
| pred. Undergraduate | 5 | 0 | 1049 | 8 | 2 | 10 | 74 | 1 | 1 | 91.22% |
| pred. Diploma | 3 | 0 | 6 | 205 | 0 | 1 | 2 | 0 | 0 | 94.47% |
| pred. Foundation Year | 0 | 0 | 19 | 0 | 159 | 2 | 10 | 0 | 0 | 83.68% |
| pred. Fine Art | 0 | 0 | 1 | 0 | 0 | 25 | 2 | 0 | 0 | 89.29% |
| pred. Intensive English | 1 | 0 | 38 | 1 | 0 | 0 | 436 | 0 | 0 | 91.60% |
| pred. Dentistry Training | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00% |
| pred. Doctorate | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00% |
| class recall | 86.76% | 83.33% | 90.59% | 90.71% | 98.76% | 64.10% | 82.58% | 0.00% | 0.00% |  |

#### Class attribute: Citizenship

### Accuracy: 59.78%

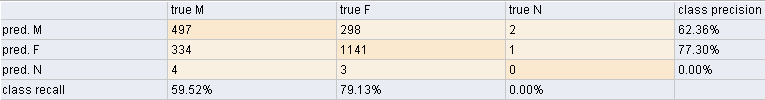
Confusion matrix is having 12 classes so can't be displayed here but Precision and recall is given:

Weighted Mean Recall: 27.79%

Weighted Mean Precision: 30.2%

#### Class attribute: Gender

Accuracy: 71.84%



Weighted Mean Recall: 69.32%

Weighted Mean Precision: 69.83%

### Stratified Sampling

#### Class attribute: Student Level

**Accuracy: 88.11%**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | true Graduate | true Higher Diploma | true Undergraduate | true Diploma | true Foundation Year | true Fine Art | true Intensive English | true Dentistry Training | true Doctorate | class precision |
| pred. Graduate | 108 | 0 | 48 | 5 | 0 | 0 | 6 | 0 | 0 | 64.67% |
| pred. Higher Diploma | 13 | 24 | 19 | 10 | 0 | 1 | 4 | 0 | 0 | 33.80% |
| pred. Undergraduate | 10 | 0 | 1028 | 3 | 3 | 8 | 71 | 0 | 0 | 91.54% |
| pred. Diploma | 2 | 0 | 4 | 215 | 0 | 0 | 2 | 0 | 0 | 96.41% |
| pred. Foundation Year | 0 | 0 | 10 | 0 | 151 | 1 | 11 | 0 | 0 | 87.28% |
| pred. Fine Art | 0 | 0 | 1 | 0 | 0 | 33 | 3 | 0 | 0 | 89.19% |
| pred. Intensive English | 1 | 0 | 33 | 0 | 0 | 1 | 449 | 0 | 0 | 92.77% |
| pred. Dentistry Training | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00% |
| pred. Doctorate | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00% |
| class recall | 80.60% | 100.00% | 89.86% | 92.27% | 98.05% | 75.00% | 82.23% | 0.00% | 0.00% |  |

#### Class attribute: Citizenship

**Accuracy: 60.20%**

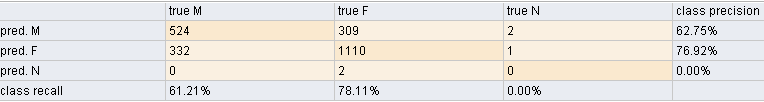
Confusion matrix is having 12 classes so can't be displayed here but Precision and recall is given:

**Weighted Mean Recall: 68.66%**

**Weighted Mean Precision: 61.74%**

#### Class attribute: Gender

**Accuracy:: 71.67%**

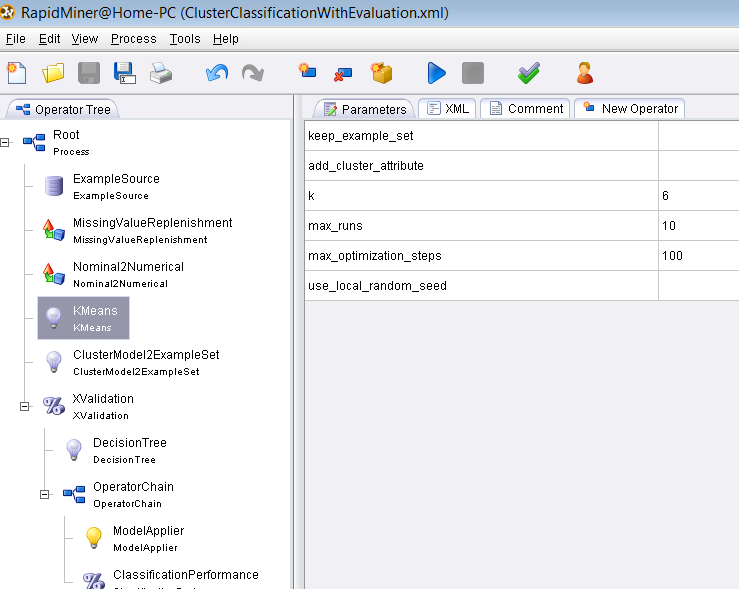
****

Weighted Mean Recall: 46.44%

Weighted Mean Precision: 46.55%

## Cluster Classification

### Cluster classification with Evaluation using K-Mean



Where K= no. of clusters so here performance measure is given when k=6 and classifying attribute is gender.

#### When k=6 AND class attribute: Gender

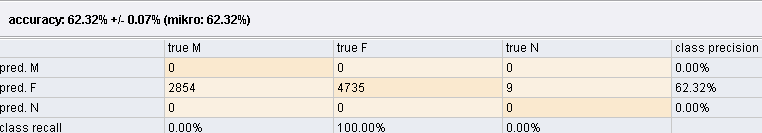
Accuracy: 62.32%

classification\_error: 37.68%

weighted Mean Recall: 33.33%

Weighted Mean Precision: 20.7%

**Confusion Matrix of three class classifier:**



#### When k=50 AND class attribute: Student-Level

**Accuracy: 99.45%**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | true Graduate | true Higher Diploma | true Undergraduate | true Diploma | true Foundation Year | true Fine Art | true Intensive English | true Dentistry Training | true Doctorate | class precision |
| pred. Graduate | 428 | 0 | 18 | 0 | 0 | 0 | 0 | 0 | 1 | 95.75% |
| pred. Higher Diploma | 0 | 81 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100.00% |
| pred. Undergraduate | 19 | 0 | 3793 | 0 | 0 | 1 | 0 | 0 | 0 | 99.48% |
| pred. Diploma | 0 | 0 | 0 | 776 | 0 | 0 | 0 | 0 | 0 | 100.00% |
| pred. Foundation Year | 0 | 0 | 0 | 0 | 512 | 0 | 0 | 1 | 0 | 99.81% |
| pred. Fine Art | 0 | 0 | 0 | 0 | 0 | 147 | 0 | 0 | 0 | 100.00% |
| pred. Intensive English | 0 | 0 | 0 | 0 | 0 | 0 | 1819 | 0 | 0 | 100.00% |
| pred. Dentistry Training | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0.00% |
| pred. Doctorate | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00% |
| class recall | 95.75% | 100.00% | 99.53% | 100.00% | 99.61% | 99.32% | 100.00% | 0.00% | 0.00% |  |

Classification Error: 0.55%

Weighted Mean Recall: 77.13%

Weighted Mean Precision: 77.22%

#### When k=20 AND class attribute: Citizenship

Accuracy: 89.63%

classification\_error: 10.37%

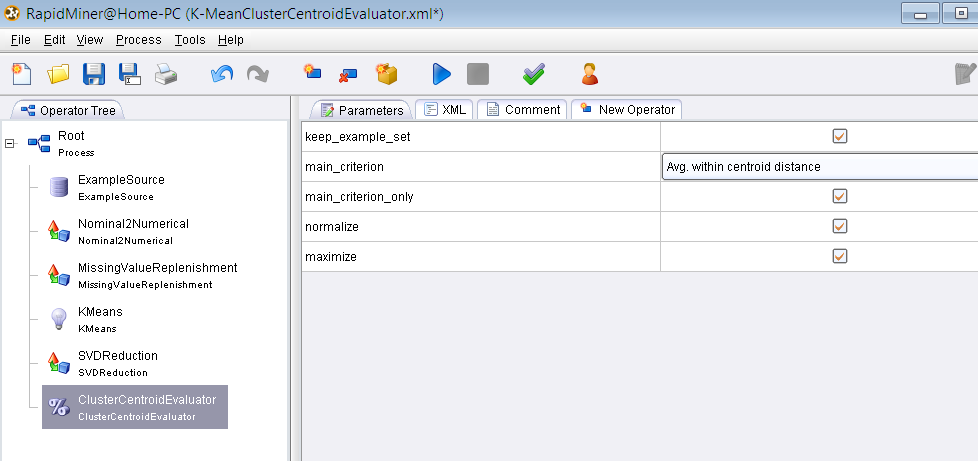
weighted\_mean\_recall: 48.56%

weighted\_mean\_precision: 50.28%

Class Attribute: **Gender**

### K-Mean Cluster Centroid Evaluator

#### Input



#### Output

**PerformanceVector:** class attribute-Student Level  
Avg. within centroid distance: 248.571  
Avg. within centroid distance\_cluster\_0: 248.571

**PerformanceVector:** **class attribute-Age**  
Avg. within centroid distance: 0.497

**PerformanceVector:** **class attribute-Gender**

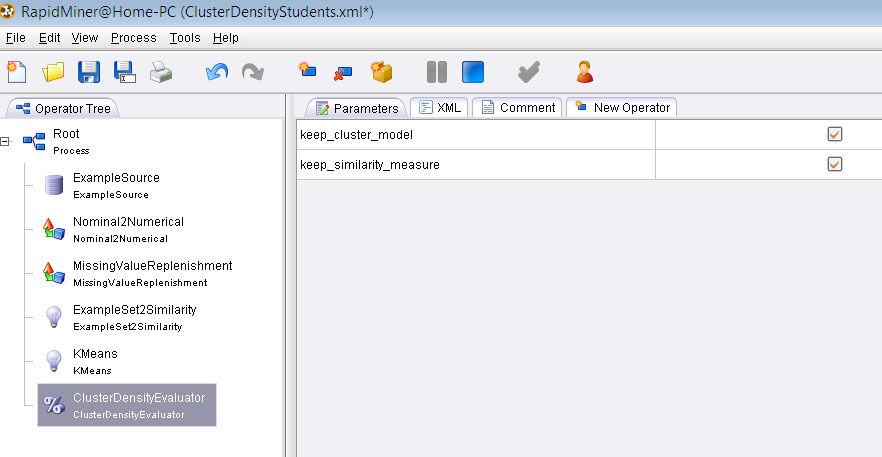
Avg. within centroid distance: 199.284

**PerformanceVector:** **class attribute-Citizenship**

Avg. within centroid distance: 198.994

### K-Mean Cluster Density Evaluator

#### Input



#### Output

**PerformanceVector:**   
Avg. within centroid distance: 230.598  
Avg. within centroid distance\_cluster\_0: 230.598

# 

# Chapter 4

**RESULTS**

# Glossary

## BinDiscretization:

This operator discretizes all numeric attributes in the dataset into nominal attributes. This discretization is performed by simple binning, i.e. the specified number of equally sized bins is created and the numerical values are simply sorted into those bins e.g. in current data set it is applied for age (a numeric value attribute)

## Nominal2Binominal:

This operator maps the values of all nominal values to binary attributes. For example, if a nominal attribute with name "costs" and possible nominal values "low", "moderate", and "high" is transformed, the result is a set of three binominal attributes "costs = low", "costs = moderate", and "costs = high". Only one of the values of each attribute is true for a specific example, the other values are false.

## Numerical2Binominal:

It converts all numerical attributes to binary ones. If the value of an attribute is between the specified minimal and maximal value, it becomes false, otherwise true. If the value is missing, the new value will be missing. The default boundaries are both set to 0, thus only 0.0 is mapped to false and all other values are mapped to true.

## FPGrowth:

This operator calculates all frequent items sets from a data set by building a FPTree data structure on the transaction data base. From this FPTree all frequent item set are derived. A major advantage of FPGrowth compared to Apriori is that it uses only 2 data scans and is therefore often applicable even on large data sets.

Given data set is only allowed to contain binominal attributes, i.e. nominal attributes with only two different values. That’s why I used the preprocessing operators in order to transform this given dataset. The necessary operators are the discretization operators for changing the value types of numerical attributes to nominal and the operator Nominal2Binominal for transforming nominal attributes into binominal / binary ones.

# Appendix

There are thousands of rules generated due to association mining. Only rules have 100% confidence are displayed here.