Using Data Mining Techniques for Investigating of Performance in ICT at G.C.E Advanced Level

P.N.W.A.L.K.Premarathne

149225D

Faculty of Information Technology

University of Moratuwa

.

May 2017

Using Data Mining Techniques for Investigating of Performance in ICT at G.C.E Advanced Level

P.N.W.A.L.K.Premarathne

149225D

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the requirements of the Degree of Master of Science in Information Technology.

May 2017

Declaration

I declare that this is my own work and has not been submitted any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student

P.N.W.A.L.K.Premarathne

Signature of Student

……

Date :

Supervised by

Name of supervisor

S.C Premarathne

Senior Lecturer

Faculty of Information Technology

University of Moratuwa

Signature of supervisor

……

Date :

Acknowledgement

A final thesis project like this is never the work of anyone alone. The contribution of many different people, in their different ways, have made this possible. I would like to extend my sincere thanks to all of them.

I would like to express my sincere gratitude to my supervisor Mr Saminda Premaratne the Senior Lecture of University of Moratuwa Sri Lanka, especially, Prof. Asoka Karunanada for the knowledge and assistance they provided throughout the research project. And I would like to thank the rest of my respective lecturers of the project committee, for their encouragement and insightful comments during project evaluation, which helped me to make necessary improvements.

Next I would like to express my special thanks to Mr Anura Pushpakumara the Commissioner General of Examination, Mrs Gayathri Abeygunasekara Commissioner of Examination Research and Development, Mr D.A Chandrasiri Commissioners of Examination Technology and other staff members of the Research and Development and the ICT branch for their support in providing data on exam results.

And I would like to thank all my friends who helped me to distribute my questionnaire among selected students as a sample.

I thank my colleagues who were behind me since last two years, sharing all good and bad moments in this hard time period.

Last but not least, I would like to thank my family for their unconditional support throughout my postgraduate degree. Your constant inspiration and guidance kept me focused and motivated. Thank you very much.

Abstract

The statistical analysis of the examination results plays a vital role in the field of education as the quality of the education is revealed by the evaluation process. The conclusions which follow the analysis, can be a feedback of the quality of examination papers or the supplementary of the students' feedback of teachers teaching process. Therefore it provides a summary of the prevailing standard of education which lead to educational research and reform. These reforms contribute to the positive changes applied to the examination system and to the learning and teaching process continuously.

In year 2011 Information and Communication Technology was introduced to the G.C.E Advanced Level as a main subject. The main objective of this research is to identify the problems related to the ICT subject of G.C.E Advanced Level Examination and provide suggestions for the teaching process of the education system. In this issue Association Rule technique of data mining is used to identify certain patterns in the examination results.

Table of Contents

Table of Figures

List of Tables

Chapter 1

1 Introduction

1.1 Prolegomena

The analysis of data and evaluation of examination results provide the theoretical basis for teaching quality and management and provide important work for the management of examination. Its conclusions are the theoretical basis for teaching evaluation, research and reform. By analyzing examination results, in one hand, the teachers can get to know how much knowledge students have obtained. For the other hand, it can be a feedback about the quality of examination papers, which is a benefit to modify the questions and make the test more standard. (Samaranayake & Caldera, 2012)Therefore statistical analysis of the examination results has been suggested for identifying the problems in the examination system as well as in the teaching process of the education system.

These results reveal the urgent need for a proper mechanism to select the most appropriate stream to follow at the G.C.E. Advanced level, mainly to reduce the failure rate. This has been the motivating factor behind this research.

Data mining is defined as a use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. It seems to be a suitable mechanism which can be used to resolve this issue. Different data mining techniques have been applied by researchers to solve different issues in the area of education. In this issue Association Rule technique of data mining is used to identify certain patterns in the examination results.

1.2 Background and Motivation

Today Data mining applications are everywhere but it started many years ago. 1763 Thomas Bayes' paper is published posthumously regarding a theorem for relating current probability to prior probability called the Bayes' theorem. It is fundamental to data mining and probability, since it allows understanding of complex realities based on estimated probabilities.

In 1805 Adrien-Marie Legendre and Carl Friedrich Gauss applied regression to determine the orbits of bodies about the Sun (comets and planets). The goal of regression analysis is to estimate the relationships among variables, and the specific method they used in this case is the method of least squares. Regression is one of the key tools in data mining.

In 1943 Warren McCulloch and Walter Pitts were the first to create a conceptual model of a neural network. In a paper entitled a logical calculus of the ideas immanent in nervous activity, they describe the idea of a neuron in a network.

1990s The term "data mining" appeared in the database community. Retail companies and the financial community are using data mining to analyze data and recognize trends to increase their customer base, predict fluctuations in interest rates, stock prices, and customer demand.

In 2003 Moneyball, by Michael Lewis, is published and changed the way many major league front offices do business. The Oakland Athletics used a statistical, data-driven approach to select for qualities in players that were undervalued and cheaper to obtain. In this manner, they successfully assembled a team that brought them to the 2002 and 2003 playoffs with 1/3 the payroll.

In February 2015, DJ Patil became the first Chief Data Scientist at the White House. Today, data mining is widespread in business, science, engineering and medicine just to name a few. Mining of credit card transactions, stock market movements, national security, genome sequencing and clinical trials are just the tip of the iceberg for data mining applications. Terms like Big Data are now commonplace with the collection of data becoming cheaper and the proliferation of devices capable of collecting data.

Present (2016) one of the most active techniques being explored is Deep Learning. Capable of capturing dependencies and complex patterns far beyond other techniques, it is reigniting some of the biggest challenges in the world of data mining, data science and artificial intelligence.

Data mining is widely used in diverse areas. There are a number of commercial data mining system available today. Financial Data Analysis, Retail Industry, Telecommunication Industry, Biological Data Analysis, Other Scientific Applications and Intrusion Detection are some of them.

The data mining technique that is applied in a number of fields like sales and marketing fields is helpful to find which product were bought together and in which sequence, in banking and finance sectors helps to retain the credit card of the customer, in transportation helps to analyze the loading patterns, in medical area it helps to identify which therapy is successful for which deceases.

When considering modern research conducted in developed countries, Data Mining is used also in education sector to predict the grades ,final results of the students, find the interest area of student ,find the performance of student in various fields , help managements in pre-planning of the business, manage the record of students in efficient way , classify the institutes.[5]

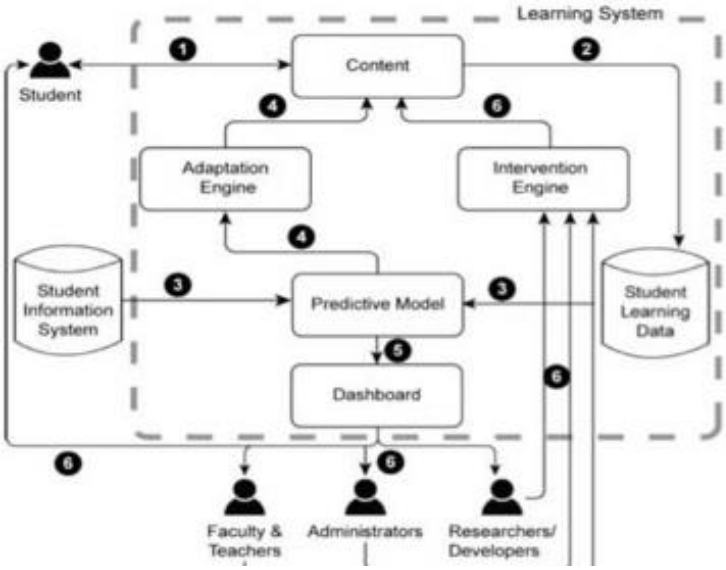


Figure 1.1 Modern Educational Data Mining

Developing countries also conducted researches relevant to Educational Data Mining. In India S. Lakshmi Prabha and Dr.A.R.Mohamed Shanavas (2014) have done research study on Educational Data Mining Applications. Johina and Vikas Kamra in India have done a study to review Data Mining Technique Used in Education Sector (2015). Manpreet Singh Bhullar, Member IAENG, Amritpal Kaur have also done a research on Use of Data Mining in Educational sector (2012). Pooja Thakar, Anil Mehta and Manisha conducted A Research Travelogue for Performance Analysis and Prediction in Educational Data Mining (2015).

Mohammed M. Abu Tair, Alaa M. El-Halees in Palestine also conducted a case study on Mining Educational Data to Improve Students' Performance (2012).

Considering about Sri Lankan context C.P. Samaranayake and H.A.Caldera have done a study on A Data Mining Solution on High Failure Rate in Physical Science Stream at the University Entrance Examination (2012).

Justification of the study

The secondary education system in Sri Lanka consists of two main examinations, the General Certificate of Education Ordinary Level (G.C.E. Ordinary level) and the General Certificate of Education Advanced Level (G.C.E. Advanced level), which are conducted by the Department of Examinations in Sri Lanka. After studying for 11 years in schools, pupils sit for the G.C.E. Ordinary level examination to get qualified for collegiate level (also called senior secondary level) education which lasts for another two years. Thereafter they sit for the competitive university entrance examination, the G.C.E. Advanced level Examination. Based on the results obtained at this examination, a limited number of students are selected to the state universities which are non-fee levying. This may be the cause for the highly competitive nature of this examination. The G.C.E. Advanced level examination is conducted under five main subject streams, namely Physical Science, Biological Science, Commerce, Arts and Technology [1]. Under each stream, students have to face three papers relevant to the stream. Apart from this, everybody has to sit for the General English paper and the Common General Knowledge paper. A certain level of competency has to be attained in these two papers to get qualified for university entrance.

However, selection to the university is based on the marks scored at the Advanced level examination for the three subjects relevant to the stream [1].

Information and communication technology (ICT) is an indispensable part of the contemporary world. In fact, culture and society have to be adjusted to meet the challenges of the knowledge age. The pervasiveness of ICT has brought about rapid technological, social, political, and economic transformation, which has eventuated in a network society organized around ICT [2] (Castells, 1996)

Despite coming of age with the Internet and other technology, many college students lack the information and communication technology (ICT) literacy skills locating, evaluating, and communicating information necessary to navigate and use the overabundance of information available today.

The discussions of Information Technology in Education typically emphasize the Technology rather than the Information. Widespread technology has meant that people encounter more information, in a greater variety of formats, than ever before. Technology is the portal through which we interact with information, but people's ability to handle information to solve problems and think critically about information tells us more about their future success than their knowledge of specific hardware or software. These skills known as Information and Communications Technology (ICT) Literacy comprise a 21st century form of literacy, in which researching and communicating information via digital environments are as important as reading and writing were in earlier centuries.[3]

A remarkable growth in the interest among the younger generation in Sri Lanka to embark on information technology related study programs is observed during the last decade or so. However, the introduction of Information technology as a subject in the secondary school curriculum is only a very recent development. According to that it has started to teach General Information Technology for Grade 12 students at year 2002. The Information & Communication Technology as a subject in the G.C.E. Ordinary Level Examination commenced from the year 2007. In year 2011 Information and Communication Technology introduced to the G.C.E Advanced Level as a main subject. A student of any stream can select ICT as a main subject. Therefore all the students of all five streams follow ICT as a subject to enter to the university. So Advanced level ICT is a more critical subject

compared to the other subjects. The main objective of this research is to identify problems related to the G.C.E Advanced Level Examination ICT subject and provide suggestions for the teaching process of the education system in a web based solution.[4]

1.3 Problem statement

In year 2011 Information and Communication Technology introduced to the G.C.E Advanced Level as a main subject. A student of any stream can select ICT as a main subject. Therefore students from all five streams follow ICT as a subject to enter to the university. So the Advanced level ICT is more critical subject than others. The main objective of this research is to identify problems related to the G.C.E Advanced Level Examination ICT subject and provide suggestions for the teaching process of the education system [4]

1.4 Hypothesis

Using data mining Techniques, performance in ICT at G.C.E Advanced Level could be investigated.

1.5 Objectives

The main objective of this research is to Identify problems related to the G.C.E Advanced Level Examination ICT subject and provide suggestions to enhance performance on ICT subject at G.C.E Advanced Level.

I. To define Ordinary Level exam qualifications for select ICT for Advanced Level.
II. To find, is there any effect for ICT subject results from the stream they sat for the Advanced Level examination.
III. To find, is there any effect from the Gender difference for ICT subject results

1.5.1 Sub Objectives

I. Provide theoretical basis for teaching quality and management
II. Get feedbacks of the quality of examination paper.
III. Modify the questions and tests up to standards.
IV. Identify problems in exam system & teaching process of education system.

V. To develop a model for investigating poor performance in G.C.E Advanced Level

VI. To evaluate the performance of the model

1.6 Data Mining Based Approach

The main objective of this research is to identify problems relating to the G.C.E Advanced Level Examination ICT subject and provide suggestions for the teaching process of the education system.

Therefore the study is based on analyzing mainly three criteria namely Analysis based on exam paper questions, Analysis based on the stream they sat for, Analysis based on the Ordinary Level Examination Results.

Analysis of examination paper considers the analysis of the most written questions and least mark scored questions. When considering the stream they sat for, it takes all the five steams which are Bio Science, Physical Science, Commerce, Arts and Technology. Grades obtained for Mathematics and ICT at Ordinary Level Examination are considered when analyzing the Ordinary Level Examination Results.

A questionnaire was used for the purpose of collecting above information. A questionnaire is a research instrument consisting of a series of questions and other prompts for the purpose of gathering information from respondents. Although they are often designed for statistical analysis of the responses.

The main focused audience of this questionnaires were the students who sat for the Advanced Level ICT subject in year 2016. Samples have been collected from island wide schools. After collecting data from above methods need to use suitable analysis method to get suitable results. Data mining system is used for these purpose.

In mining system the data preprocess is the phase where data cleaned from noise by overcoming the difficulties of recognizing its variables in order to be used as input to the next phase of pattern discovery. In the pattern mining phase various mining algorithms are incorporated into the system to mine different types of pattern. In the pattern analysis phase

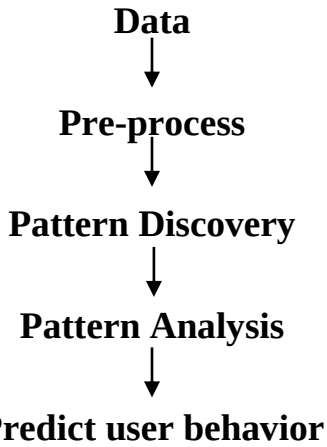
the mined patterns which are great number to be evaluated. Mining system is classified and explained. Commonly a mining system introduces three parts:

(i) Data Preprocessing

(ii) Pattern Discovery

(iii) Pattern Analysis

General Mechanism



1.7 Structure of the thesis

The rest of the thesis is organized as follows. Chapter 2 critically reviews the literature on Usage of Data Mining Techniques for investigating of poor performances in exams, for recognition and identify the research problems. Chapter 3 is about the Data Mining Techniques that can apply to solve above problem statement. Chapter 4 presents out new approach to data mining techniques. Chapter 5 and chapter 6 describe the design and implementation respectively.

1.8 Summary

This chapter gave an overall picture of the ‘entire project presented in this thesis. As such we describe the background / motivation, problem definition, hypothesis objectives and a brief overview of the solution. Next present a critical review of literature on using Data Mining Techniques for Investigating Poor Performance in ICT at G.C.E Advanced Level.

Chapter 2

2 Literature Review - Development and Challenges in Data Mining

2.1 Introduction

Chapter one gave a comprehensive description of the overall project described in this thesis. This chapter provides a critical review of the literature in relation to developments and challenges in Data mining. For this purpose the review of the past researches have been presented under three major sections. Namely, early (Birth) developments, modern trends (achievements) and future challenges (unsolved problems). At the end, this chapter defines the research problem as the Poor performance in ICT at G.C.E Advanced Level and identifies the technology that can be used to addressed the problem and identify the technology to be used to address the problem.

2.2 Early developments

Framework 1

In the last 15 years educational data mining emerged as a new application area for data mining.(Romero & Ventura, 2007) [7] Provided a survey of educational data mining from 1995-2005 and extended their survey covering the latest development until 2009.

Framework 2

There is an increasing number of data mining applications in education, from enrollment management, graduation, academic performance, gifted education, web-based education, retention and other areas (Nandeshwar & Chaudhari, 2009)[8]. This section reviews only research where the main focus is on study outcome, successful or unsuccessful course completion.

Framework 3

(Kovacic, 2010) [9] found for courses in the mathematics and computing faculty at the Open University in UK, by using the binary logistic regression, that the most significant factors to whether students passed, failed or dropped out, were marks for the first

assignment, the number of maths courses passed in the previous two years, the course level, the points the course is worth and the occupation group of the student.

Framework 4

(Yuan, Deng, Zhu, & Li, 2012) [11] Used a data mining approach to differentiate the predictors of retention among freshmen enrolled at Arizona State University. Using the classification tree based on an entropy tree-splitting criterion they concluded that 'cumulated earned hours' was the most important factor contributing to retention. Gender and ethnic origin were not identified as significant.

Framework 5

(Sarker, Tiropanis, & Davis, 2013)[10] Predicted the secondary student grades of two core classes using past school grades, demographics, social and other school related data. The results were obtained using decision trees, random forests, neural networks and support vector machines. They achieved high level of predictive accuracy when the past grades were included. In some cases their models included also the school related features, demographics (student's age, parent's job and education) and social variables. Unfortunately most of their variables (e.g. student previous grades) were not available for the Open Polytechnic students.

Framework 6

El-Halees [12], gave a case study that used educational data mining to analyze students' learning behavior. The goal of his study is to show how useful data mining can be used in higher education to improve student' performance. He used students' data from database course and collected all available data including personal records and academic records of students, course records and data came from e-learning system. Then, he applied data mining techniques to discover many kinds of knowledge such as association rules and classification rules using decision tree. Also he clustered the student into groups using EMclustering, and detected all outliers in the data using outlier analysis. Finally, he presented how can we benefited from the discovered knowledge to improve the performance of student.

Framework 7

Baradwaj and Pal [13], applied the classification as data mining technique to evaluate student' performance, they used decision tree method for classification. The goal of their study is to extract knowledge that describes students' performance in end semester examination. They used students' data from the student' previous database including Attendance, Class test, Seminar and Assignment marks. This study helps earlier in identifying the dropouts and students who need special attention and allow the teacher to provide appropriate advising.

Framework 8

Chandra and Nandhini [14], applied the association rule mining analysis based on students' failed courses to identify students' failure patterns. The goal of their study is to identify hidden relationship between the failed courses and suggests relevant causes of the failure to improve the low capacity students' performances. The extracted association rules reveal some hidden patterns of students' failed courses which could serve as a foundation stone for academic planners in making academic decisions and an aid in the curriculum re-structuring and modification with a view to improving students' performance and reducing failure rate.

Framework 9

Latest work published in International Journal of Computer Science and Mobile Computing, 2014 describes the process of finding the set of weak students based on graduation and post graduation marks [15]. Another paper published in European Journal of Scientific Research in 2010 also analyzed students' learning behavior to predict weak students. [16]. P. Ramasubramanian, K. Iyakutti and P. Thangavelu, in year 2009 also predicted weak students using rough set theory. [17]

Framework 10

To assist in selecting students for enrollment in a particular course Surjeet Kumar Yadav and Saurabh Pal used Decision Trees technique of data mining in 2012 [18]. In another

paper In 2010 Zlatko J. Kovačić also investigated enrolment attributes to pre-identify success of students.[19]

Framework 11

A comprehensive evaluation method for undergraduates that can objectively distinguish the grades of students was developed by Xiewu, Huacheng Zhang, Huimin Zhang in year 2010 [20]. Another study by Dai Shangping, Zhang Ping, in year 2008 predicted final grades of students based on features extracted from log data in web-based system and published their work in IEEE [21]

Framework 12

Decision tree proved to be consistently 3-12% more accurate than the Bayesian network in predicting academic performance of undergraduate and postgraduate students in a paper titled "A Comparative Analysis of Techniques for Predicting Academic Performance" in 2007 [22], IEEE.

Framework 13

Al-Radaideh et al. [23], applied the data mining techniques, particularly classification to help in improving the quality of the higher educational system by evaluating student data to study the main attributes that may affect the student performance in courses. The extracted classification rules are based on the decision tree as a classification method, the extracted classification rules are studied and evaluated. It allows students to predict the final grade in a course under study.

Framework 14

Ayesha et al. [24], used k-means clustering algorithm as a data mining technique to predict students' learning activities in a students' database including class quizzes, mid and final exam and assignments. These correlated information will be conveyed to the class teacher before the conduction of final exam. This study helps the teachers to reduce the failing ratio by taking appropriate steps at right time and improve the performance of students.

2.3 Modern trends in data mining

There are two forms of data analysis that can be used for extracting models describing important classes or to predict future data trends. These two forms are as follows

- Classification
- Prediction

Classification models predict categorical class labels, and prediction models predict continuous valued functions. For example, we can build a classification model to categorize bank loan applications as either safe or risky, or a prediction model to predict the expenditures in dollars of potential customers on computer equipment given their income and occupation.

2.4 Major Issues and Future challenges of data mining

The major issue is preparing the data for Classification and Prediction. Preparing the data involves the following activities

- **Data Cleaning** − Data cleaning involves removing the noise and treatment of missing values. The noise is removed by applying smoothing techniques and the problem of missing values is solved by replacing a missing value with most commonly occurring value for that attribute.
- **Relevance Analysis** − Database may also have the irrelevant attributes. Correlation analysis is used to know whether any two given attributes are related.
- **Data Transformation and reduction** − The data can be transformed by any of the following methods.
- **Normalization** − The data is transformed using normalization. Normalization involves scaling all values for given attribute in order to make them fall within a small specified range. Normalization is used when in the learning step, the neural networks or the methods involving measurements are used.
- **Generalization** − The data can also be transformed by generalizing it to the higher concept. For this purpose we can use the concept hierarchies.

2.5 Problem Definition

The literature review has identified various unsolved problems including security, efficiency, and reliability of algorithms. Table 2-1 summaries the achievements and the limitations of the key research discussion in this chapter.

Table 2-1 achievements and the limitations of the key research discussion in this chapter.

Research	Technology Used	Key benefits	Limitations	Remarks
Johina Vikas Kamra	Classification, clustering and association	Classification based on learning and classification check the accuracy of classification rules finds the classes and assigns each object to a particular class.find interesting association and correlation between a large data set.	quality of clustering is also depends on its ability to discover the hidden patterns.	The data mining algorithms in education field used for the improvement in the performance of student and the institution
C.P. Samaranayake H.A.Caldera	Association Rule technique		lack of knowledge on other factors that influence	Other factors that may have an effect on the G.C.E.

			the performance at the G.C.E. Advanced level examination Comparing the results of students who sat the examination in different years is not practical	Advanced level results of students could be Considered
Mohammed M. Abu Tair, Alaa M. El-Halees	Classification, Association rules; Clustering, Outlier Detection.			

It is evident from the above summary that the research has been done to Identify problems relating to the G.C.E Advanced Level Examination ICT subject and provide suggestions for the teaching process of the education system.

2.6 Summary

This chapter presented a comprehensive literature review on the analysis of Educational Data mining problems and identifying the research problems relating to the G.C.E Advanced Level Examination ICT subject and provided suggestions for the teaching process of the education system. Also identified the Data Mining technology to address the above problem. Next chapter will discuss the technology to be used for the solution.

Chapter 3

3 Technology - Data Mining Tools and Techniques. Principles of opinion mining

3.1 Introduction

Data mining is the process of finding of hidden information from a huge amount of data. Data mining analyzing the data from different source and convert it into meaningful information [1]. Data mining is a new powerful technology that helps business to focus on important information like future trends, decision making, customer choice etc. A target dataset is prepared before applying the data mining algorithm. The common source of data is the data warehouse. Pre–processing is needed to analyze the data sets before applying the data mining. The data mining process is as follows shown in Figure 3.1

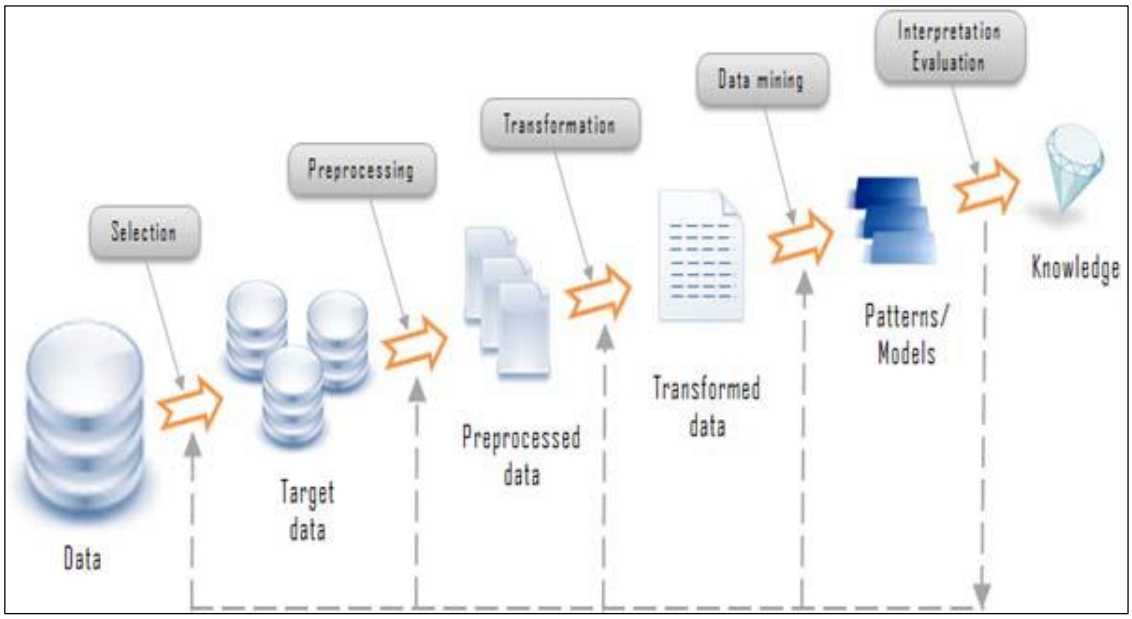


Figure 3.1 *The data mining process*

Data is collected from a different source and combined in one data store called the data set or target data. Data is then pre-processed and transformed into the required format. Then data mining technique is applied on the transformed data. The output is finally presented in forms of tables and graphs called the knowledge or information.

3.2 Data Mining In Education Sector

The data mining technique are applied in a number of fields likes in sales/marketing fields helps to find which product were bought together and in which sequence, in banking and finance sectors helps to retain the credit card of the customer, in transportation helps to analyze the loading patterns, in medical area it helps to identify which therapy is successful for which deceases but now a day's also used in education sectors. There are a lot of advantages of data mining technique in education sector. Some advantages of data mining in education sectors are as follows [3]:

- Data mining techniques helps to predict the grades, final results of the students.
- It helps to find the most interested area of student.
- It helps in to find the performance of student in various fields.
- It also helps managements to pre-plan of the business.
- Data mining is used in education sector to manage the record of students in efficient way.
- Data mining is used to classify the institutes.

There are various application of data mining in education sector. Data mining methods like prediction, clustering, classification, association rule, decision tree and relationship mining are hugely used in the field of marketing and finance. In the figure 2 shows how the data mining can be used in education sector. [25]

3.3 Data Mining Techniques

There are various data mining techniques that have been developing and used in data mining projects lately. These include classification, clustering, and association. Association is the selected data mining technique for investigate the performance of Advanced level ICT subject.

3.3.1 Association

Association rule mining technique is used to find the most frequent item in the large data set. The main goal is to find interesting association and correlation between a large data set. The association rule helps to make decisions such as customer shopping behavior, loss leader analysis and catalogue design. Association rule needs to generate rule with value less than one. Now a day's many companies are being interested in association rule to

increase the profits. The association rule is easy to use and helps to improve the profits of the companies.

Types of association rule

- Multi-level association rule
- Multi-dimensional association rule
- Quantitative association rule

3.3.1.1 Lift value

The lift value is a measure of importance of a rule. By using rule filters, you can define the desired lift range in the settings. The lift value of an association rule is the ratio of the confidence of the rule and the expected confidence of the rule. The expected confidence of a rule is defined as the product of the support values of the rule body and the rule head divided by the support of the rule body.

3.3.1.2 FP growth Algorithm

The FP-Growth Algorithm is an alternative way to find frequent item sets without using candidate generations, thus improving performance. For so much it uses a divide-and-conquer strategy. The core of this method is the usage of a special data structure named frequent-pattern tree (FP-tree), which retains the item set association information.

3.4 Summary

This chapter presented the technologies we have used in research. Data mining is decision making process used to find the useful pattern of information. Also data mining is of high importance in finding the patterns, forecasting, and discovery of knowledge in various business areas. Data mining techniques such as classification, clustering, association also help to find useful information in education system and managing information in effective way. The data mining algorithms in education field is used for the improvement in the performance of student and other related parties of education sector.

Chapter 4

4 Approach

4.1 Introduction

From this chapter discuss the approach of doing the research. This section illustrates how the above technologies are used for the research to address the problem defined. The approach is described under the headings of Hypothesis, Input to the system, Output of the system, process convert input to the output, users of the system and features of the system.

4.2 Hypothesis

Using data mining Techniques, performance in ICT at G.C.E Advanced Level could be investigated.

4.2.1 Input to the system

The system could be able to accept input data of Grade obtained for Ordinary Level Examinations ICT subject, Grade obtained for Ordinary Level Examinations Mathematics subject, Stream of the Advanced level and Grade obtained for Advanced Level ICT subject. Data should be preprocessed before entered. As well as selected features of exam questions (Mostly difficult questions) and the significant points of the syllabus (Most interesting chapters on the syllabus, Rate of the level of teaching, Need of the extra knowledge help) should be entered.

4.2.2 Output of the system

As the output of this process different data patterns, related to each input factor could be predicted according to the sub research questions. Best subject streams and most suitable subject combinations could be predicted. Could be able to identify most difficult areas of the current syllabus and provide suggestions for curriculum changes and enhancements match with the future opportunities.

4.3 Process

Data that have been collected from particular party are subjected for preprocessing and are loaded into the proposed tool using its interface, then data mining algorithm is applied. Following are the steps of data mining.

- Data Integration: First of all the data are collected and integrated from all the different sources.
- Data Selection: We may not use all the data we have collected in the first step. So in this step we select only those data which we think be useful for data mining.
- Data Cleaning: The data we have collected are not clean and may contain errors, missing values, noisy or inconsistent data. So we need to apply different techniques to get rid of such anomalies.
- Data Transformation: The data even after cleaning are not ready for mining as we need to transform them into forms appropriate for mining. The techniques used to accomplish this are smoothing, aggregation, normalization etc.
- Data Mining: Now we are ready to apply data mining techniques on the data to discover the interesting patterns. Techniques like clustering and association analysis are among the many different techniques used for data mining.
- Pattern Evaluation and Knowledge Presentation: This step involves visualization, transformation, removing redundant patterns etc from the patterns we generated.
- Decisions / Use of Discovered Knowledge: This step helps user to make use of the
- knowledge acquired to take better decisions.

The following Figure 4.2 shows the process of knowledge discovery

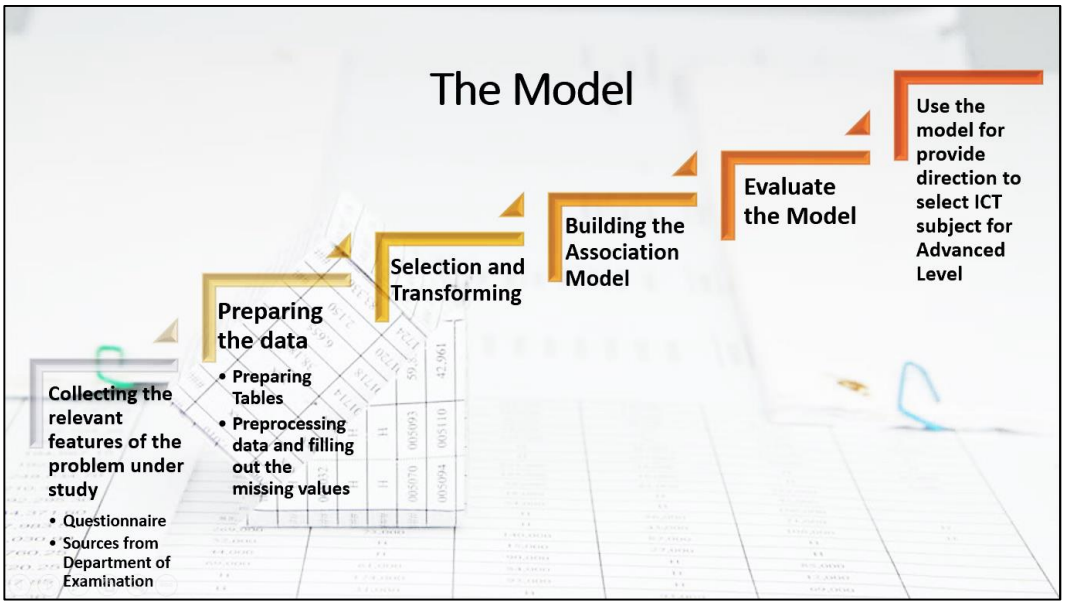


Figure 4.1 The Model

The main objective of this research is to identify problems relating to the G.C.E Advanced Level Examination ICT subject and provide suggestions for the teaching process of the education system in a web based solution.

Therefore the study is based on analyzing mainly three criteria's where are Analysis of exam paper questions, Analysis by the stream they sat for and Analysis by Ordinary Level Examination Results.

Analysis of examination paper consider the analysis of most written questions and least marks scored questions. When considering the stream they sat for, it takes all the five steams which are Bio Science, Physical Science, Commerce, Arts and Technology. Grades obtained for Mathematics and ICT for Ordinary Level Examination are consider when analysis the Ordinary Level Examination Results.

For the purpose of collecting above information questionnaires and Google forms were used. A questionnaire is a research instrument consisted of a series of questions and other prompts for the purpose of gathering information from respondents. Although they are often designed for statistical analysis of the responses, Google Forms is a tool that is part of Google Drive for creating surveys, tests, or web input forms. Google forms allow anyone to create an easy to use web form, tie it to a spreadsheet where you can track results and post it on the web without having to know programming.

The main focused audiences of this questionnaires are the students who sat for the Advanced Level ICT subject in year 2016. Samples have collected from schools Island wide. After collecting data from above methods a suitable analysis method is needed to get suitable results. Data mining systems were used for these purposes.

In mining system the data preprocess is the phase where data cleaned from noise by overcoming the difficulties of recognizing its variables in order to be used as input to the next phase of pattern discovery. In the pattern mining phase various mining algorithms are incorporated into the system to mine different types of pattern. In the pattern analysis phase the mined patterns which are in a great number are to be evaluated. Mining system is classified and explained. Commonly a mining system introduces three parts:

(i) Data Preprocessing

(ii) Pattern Discovery

(iii) Pattern Analysis

4.4 Users of system

Advanced Level Students who are eager to select the ICT as a subject, Teachers who teach ICT subject in schools, Educational Policy makers, Relevant officers of Ministry of Education and National Institute of Education are the users of the results of this analysis.

4.5 Features

- **Accuracy** − Accuracy of classifier refers to the ability of classifier. It predicts the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new datum.
- **Speed** − This refers to the computational cost in generating and using the classifier or predictor.
- **Robustness** − It refers to the ability of classifier or predictor to make correct predictions from given noisy data.
- **Scalability** − Scalability refers to the ability to construct the classifier or predictor efficiently; given large amount of data.
- **Interpretability** − It refers to what extent the classifier or predictor understands.

4.6 Summary

This chapter has described overall solution of the research. Have mentioned problem definition and assumption of the solution to that problem. Have described clearly the inputs to the system, outputs of the system and how to convert input to output, who have been benefited of this system, and incredible features of solution. Next chapter will describe in detail, extended design of process and what the system does.

Chapter 5

5 Design

5.1 Introduction

The approach to solve the problem defined in the chapter one was described in the previous chapter. Here from design chapter the process that stated in on approach is elaborated by using diagrams and detailed descriptions. The design section shows each module or component and the relationship among them.

5.2 System Design

According to the study it has identified major three modules in research which can be listed below.

1. Analysis of exam paper questions
    i. What are the mostly written questions
    ii. What are the least marks scored questions
2. Analysis by the stream they sat for
    i. Bio Science
    ii. Physical
    iii. Commerce
    iv. Arts
    v. Technology
3. Analysis by Ordinary Level Examination Results
    i. Mathematics
    ii. ICT

According to the above details an initial questionnaire is consist of following objectives, descriptions and problems was designed as shown in Table 5-1 *Details* an initial questionnaire. Initial Questionnaire is attached with appendices.

Table 5-1 Details an initial questionnaire

Base objectives	Description	Problems
1. **Questions on Exam paper**	• Exam paper is consists of two main parts. • Part I & Part II, • Part II is consists of two parts as part A and Part B • Full syllabus of A\L ICT is covered by the paper. • Main types of questions that could be asked are given. **Chapters of the syllabus** 1.Basic Concepts, Roll and History of ICT 2. Number systems & Logic Gates 3. Data Base Design & Development 4. Web designing 5. Operation systems & Memory Management 6. E- Commerce & Future Trends 7. Data Communication & Computer Networking. 8. Programming Languages (Python) 9. System Analysis and Design	1. Reasons for scoring low marks. 2. What are the most interesting chapters and why? 3. What are the most difficult chapters and why? 4. Percentage of coverage of the syllabus 5. Method of teaching, Good or Bad in each chapter 6. Contribution of the School & the Tution class

	In this area it was considered to find the mostly written questions and least mark scored questions.	
2. **Stream**	At the year 2009 Information and Communication Technology has been introduced to the Advanced Level Syllabus. Students sat for the exam initially in 2011. ICT subject can be combined with all the five streams equally. Student can select ICT subject with any two subjects in each stream and sit for the university entrance exam. There are five main streams currently in the SL school education system. 1. Bio Science 2. Physical Science 3. Commerce 4. Arts 5. Technology	1. What is the selected stream? 2. What are the other two subjects?

3. O\L results	• A data mining solution on high failure rate in Physical Science stream at the university entrance examination. • C.P. Samaranayake, H.A. Caldera has done the similar kind of research by analyzing O\L results. In here it is expected to consider, ✓ Whether the student followed ICT subject for Ordinary Level examination. ✓ Results obtained for Mathematics. ✓ whether there is an impact on Gender difference.	1. Followed ICT or Not for OL? 2. If "Yes" what is the grade obtained for ICT? 3. Grade obtained for Mathematics. 4. Impact on Gender

The initial questionnaire is attached with appendixes 1.

5.3 Top level Design

Top level design of a data mining system shown in following Figure 5.1

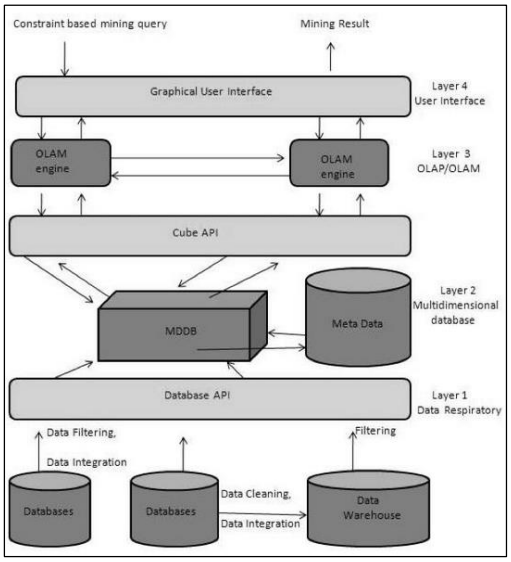


Figure 5.1 Top Level Design

5.4 User interface Design

User interface is the module of data mining system that helps the communication between users and the data mining system. User Interface allows the following functionalities

- Interact with the system by specifying a data mining query task.
- Providing information that help to focus on the search.
- Mining based on the intermediate data mining results.
- Browse database and data warehouse schemas or data structures.
- Evaluate mined patterns.
- Visualize the patterns in different forms.

5.5 Summary

This chapter provided details on research design and applicability of selected method for the research. Furthermore this chapter focuses on top level design for the research and how sub questions are structured with in the research. Subsequent section will be discussed about implementation details according to its design.

Chapter 6

6 Implementation

6.1 Introduction

This chapter provides implementation details of each of the module mentioned in previous chapter. Moreover, this presents computer hardware, equipment, software and algorithm, code segments used in each module with sample output.

6.2 Collection of Data

Data for analysis is collected from the Department of Examination and by distributing questionnaires among the for students who sat for Advanced Level Examination in Year 2016. The questionnaire is attached with appendixes.

6.3 RapidMiner

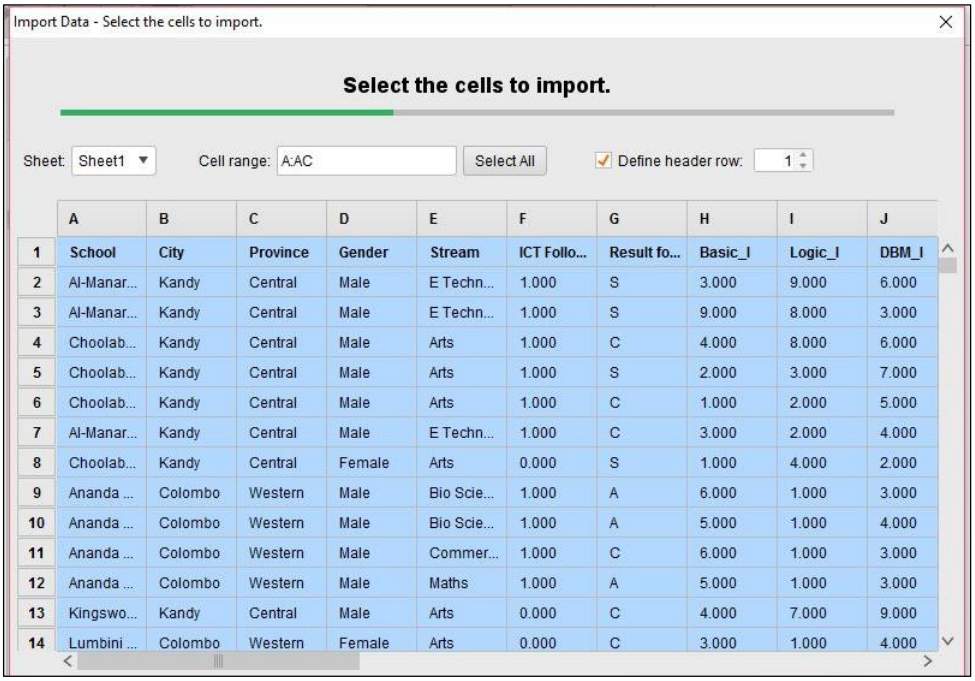
Software RapidMiner Studio combines technology and applicability to serve a user-friendly integration of the latest as well as established data mining techniques. Defining analysis processes with RapidMiner Studio is done by drag and drop of operators, setting parameters and combining operators. RapidMiner Studio is a code-free environment for designing advanced analytic processes with machine learning, data mining, text mining, predictive analytics and business analytics.

RapidMiner Studio combines technology and applicability to serve a user-friendly integration of the latest as well as established data mining techniques. Defining analysis processes with RapidMiner Studio is done by drag and drop of operators, setting parameters and combining operators.

6.4 Data Pre-Processing

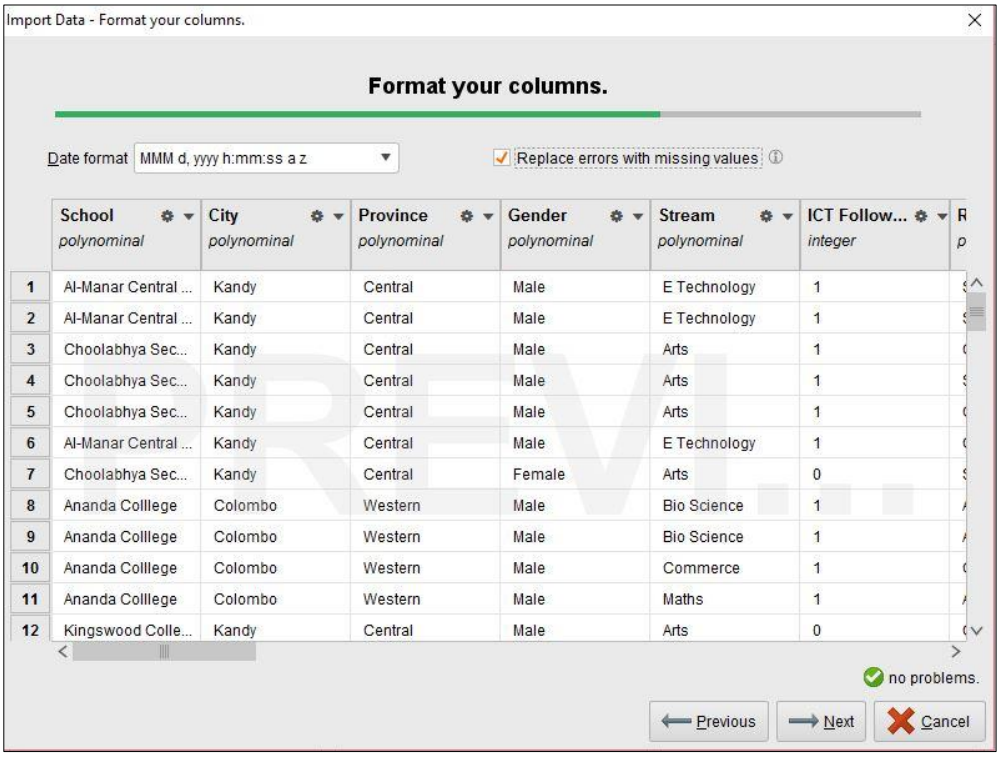
Collected data from questionnaire had lots of missing values and noisy values. Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. The data already collected from questionnaires and web based methods is often incomplete, inconsistent, and lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues.

In this analysis followed the process showed in Figure 6.1 to Figure 6.4 preprocess data replace missing values is done with Rapidminer using the Replace Missing Value operator. Which places all missing values for nominal and numeric attributes in a data set with the average value from the training data.



	A	B	C	D	E	F	G	H	I	J
1	School	City	Province	Gender	Stream	ICT Follo...	Result fo...	Basic_I	Logic_I	DBM_I
2	Al-Manar...	Kandy	Central	Male	E Techn...	1.000	S	3.000	9.000	6.000
3	Al-Manar...	Kandy	Central	Male	E Techn...	1.000	S	9.000	8.000	3.000
4	Choolab...	Kandy	Central	Male	Arts	1.000	C	4.000	8.000	6.000
5	Choolab...	Kandy	Central	Male	Arts	1.000	S	2.000	3.000	7.000
6	Choolab...	Kandy	Central	Male	Arts	1.000	C	1.000	2.000	5.000
7	Al-Manar...	Kandy	Central	Male	E Techn...	1.000	C	3.000	2.000	4.000
8	Choolab...	Kandy	Central	Female	Arts	0.000	S	1.000	4.000	2.000
9	Ananda ...	Colombo	Western	Male	Bio Scie...	1.000	A	6.000	1.000	3.000
10	Ananda ...	Colombo	Western	Male	Bio Scie...	1.000	A	5.000	1.000	4.000
11	Ananda ...	Colombo	Western	Male	Commer...	1.000	C	6.000	1.000	3.000
12	Ananda ...	Colombo	Western	Male	Maths	1.000	A	5.000	1.000	3.000
13	Kingswo...	Kandy	Central	Male	Arts	0.000	C	4.000	7.000	9.000
14	Lumbini ...	Colombo	Western	Female	Arts	0.000	C	3.000	1.000	4.000

Figure 6.2 Select the cells to import



	School *polynominal*	City *polynominal*	Province *polynominal*	Gender *polynominal*	Stream *polynominal*	ICT Follow... *integer*	R *p*
1	Al-Manar Central ...	Kandy	Central	Male	E Technology	1	
2	Al-Manar Central ...	Kandy	Central	Male	E Technology	1	
3	Choolabhya Sec...	Kandy	Central	Male	Arts	1	
4	Choolabhya Sec...	Kandy	Central	Male	Arts	1	
5	Choolabhya Sec...	Kandy	Central	Male	Arts	1	
6	Al-Manar Central ...	Kandy	Central	Male	E Technology	1	
7	Choolabhya Sec...	Kandy	Central	Female	Arts	0	
8	Ananda Colllege	Colombo	Western	Male	Bio Science	1	
9	Ananda Colllege	Colombo	Western	Male	Bio Science	1	
10	Ananda Colllege	Colombo	Western	Male	Commerce	1	
11	Ananda Colllege	Colombo	Western	Male	Maths	1	
12	Kingswood Colle...	Kandy	Central	Male	Arts	0	

Figure 6.1 Format your columns

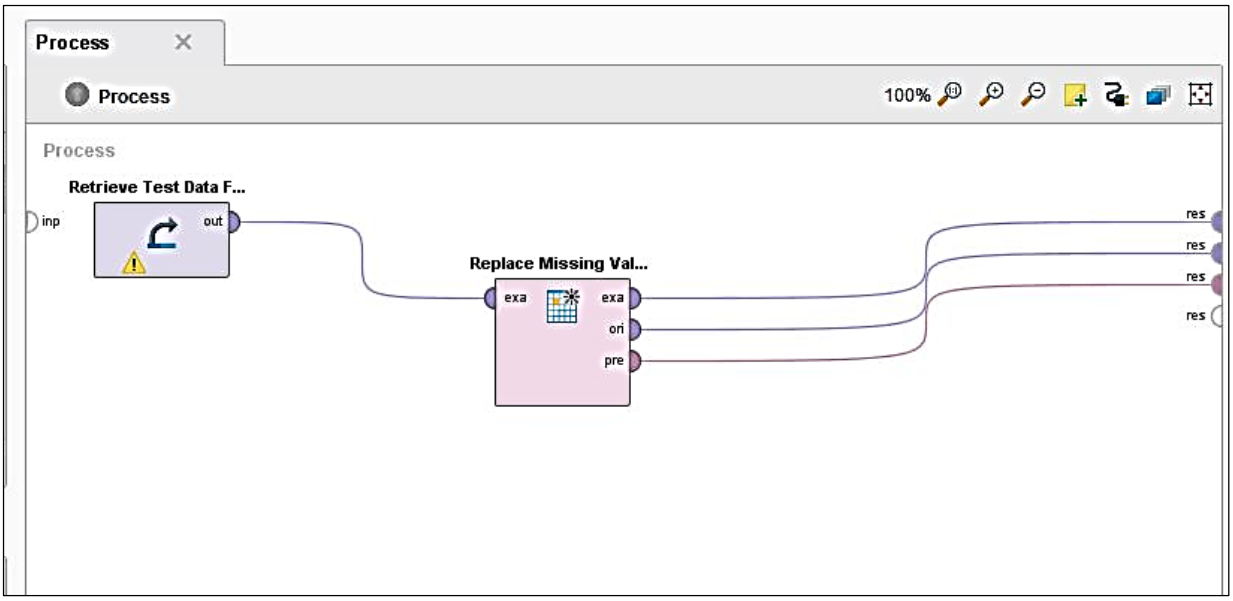


Figure 6.3 Replace missing value operator

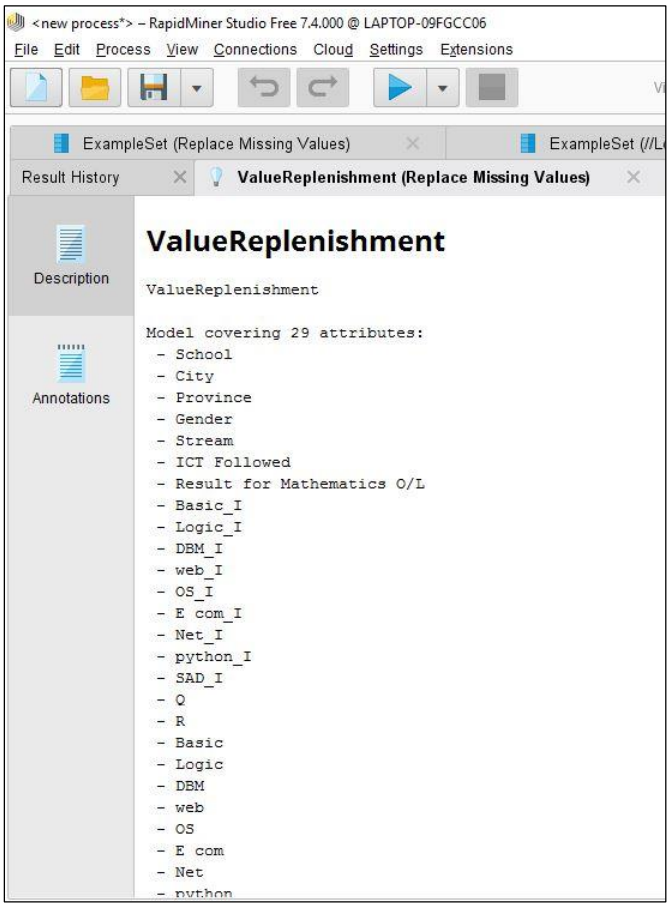


Figure 6.4 Values Replenishment

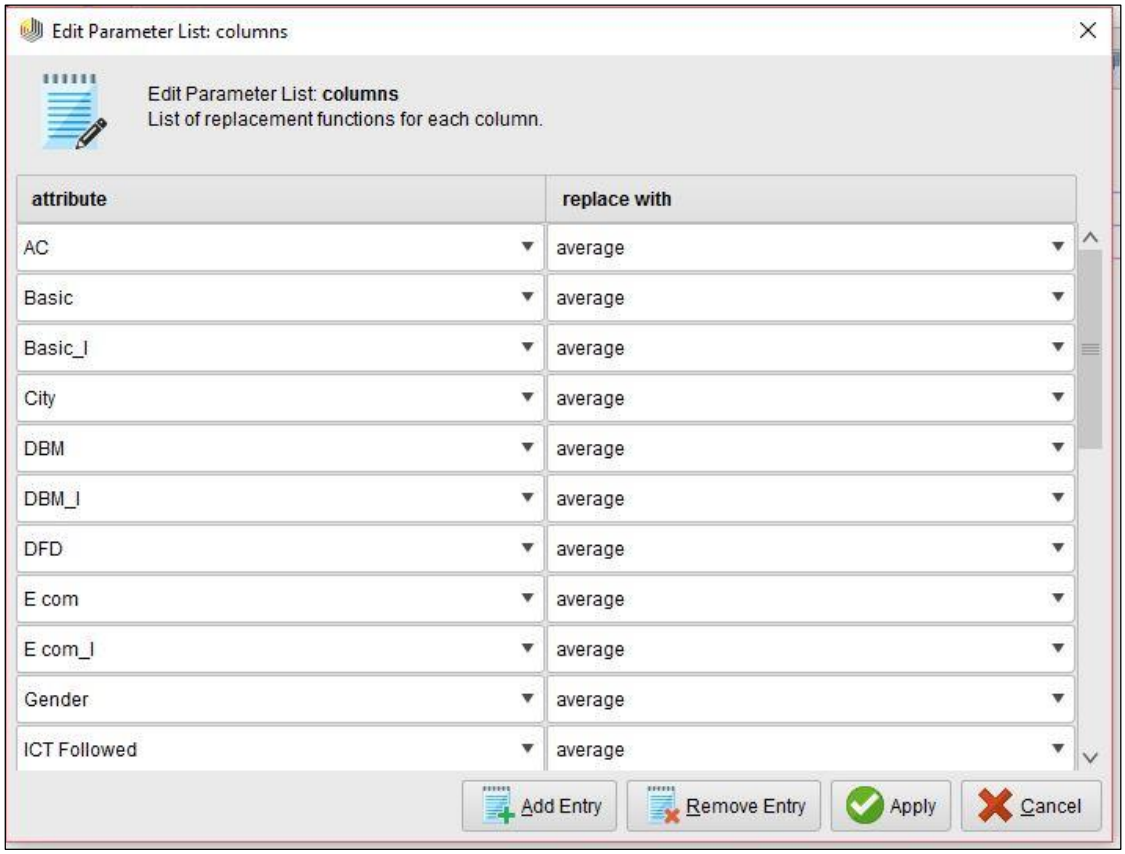


Figure 6.5 Edit Parameter list: columns

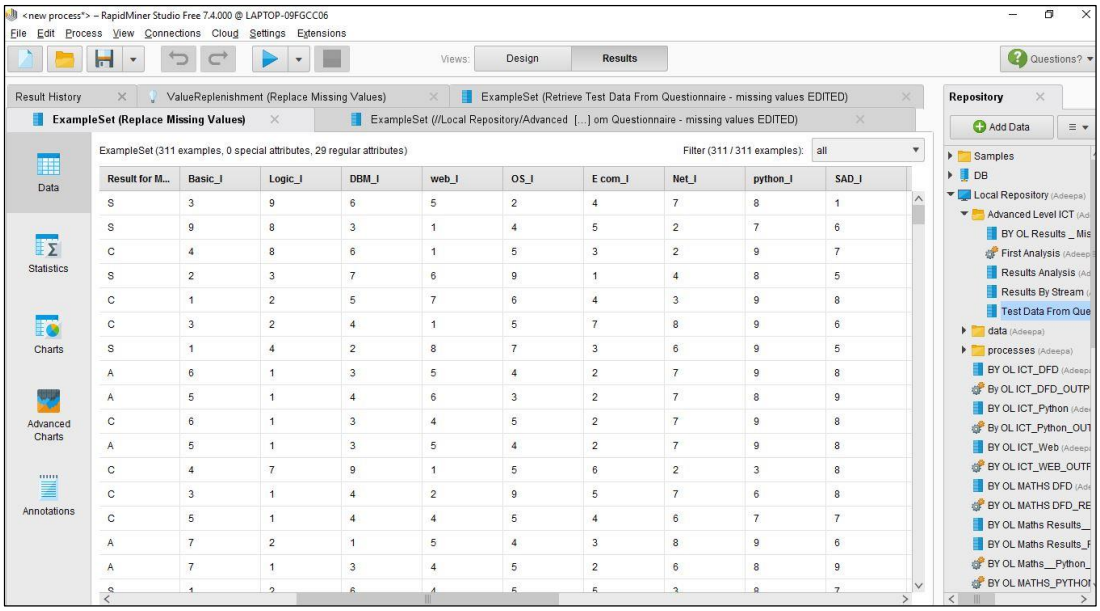


Figure 6.6 Missing values replaced data set

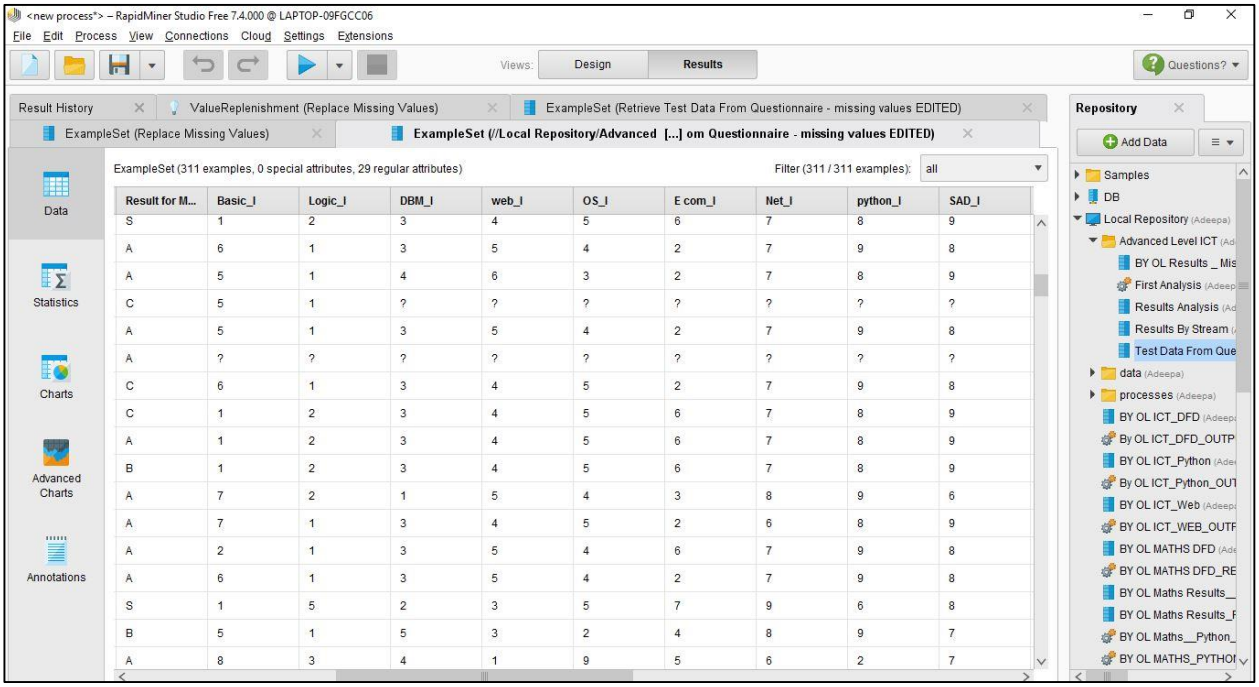


Figure 6.7 Missing values replaced data set

6.5 Analysis by Questions on Exam paper

By analyzing the questionnaire that have provided for students who sat for the Advanced Level examination, it has ranked the difficulty level of each question. According to that it has been identified three chapters which were considered by students by way of it difficulty.

1. Python
2. Web designing
3. DFD

Likewise these difficulty level was compared with their ordinary level examination results. These analysis can be divided in to two main parts.

1. Grade obtained for Mathematics
2. Whether followed ICT or Not followed
3. Gender Difference

Similarly each chapter separately was analyzed with the above two criteria. Grade obtained for mathematics subject at O\L examination. From the present criteria it was considered whether there was an impact on the level of difficulty from the grade obtained for the O\L mathematic subject, whether there was an impact following or not following ICT for O/L and the Gender Difference.

Association Rule

Association rules are created by analyzing data under frequent if/then patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in the database. Confidence indicates the number of times the if/then statements have been found to be true. The frequent if/then patterns are mined using the operators like the FP-Growth operator. The Create Association Rules operator takes these frequent item sets and generates association rules.

'By O\L results Maths' data set is loaded using the Retrieve operator. Thus the FP-Growth operator cannot be applied on it directly because the FP-Growth operator requires all attributes to be binominal. We have to do some preprocessing to mold the ExampleSet into desired form. The Nominal to Binominal operator is applied to change these nominal attributes to binominal attributes. Finally, the FP-Growth operator is applied to generate frequent itemsets. The frequent itemsets generated from the FP-Growth operator are provided to the Create Association Rules operator. The resultant association rules can be viewed in the Results Workspace.

6.5.1 Grade obtained for Mathematics

6.5.1.1 Python

There are 500 students who have been considered as the sample and 443 students had selected the python chapter as difficult. Look over the association between the O\L mathematics result and the level of difficulty of python chapter and its effect on the A\L ICT subject result, was done.

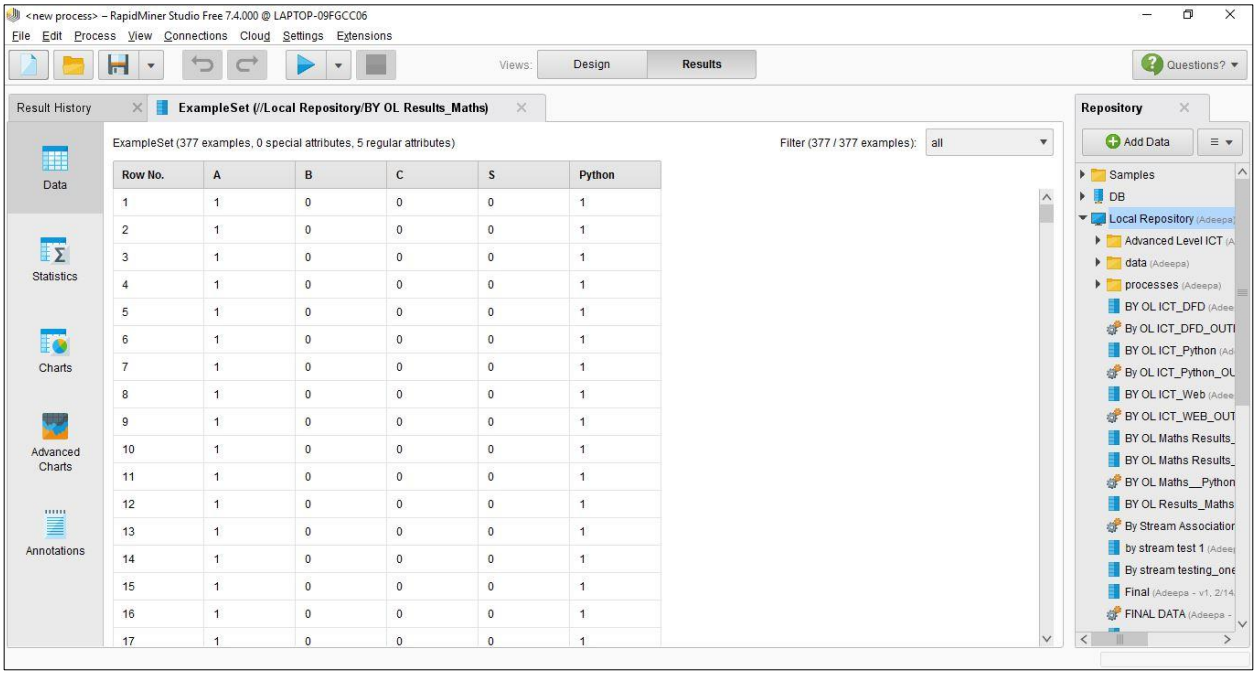


Figure 6.8 Level of difficulty in python chapter

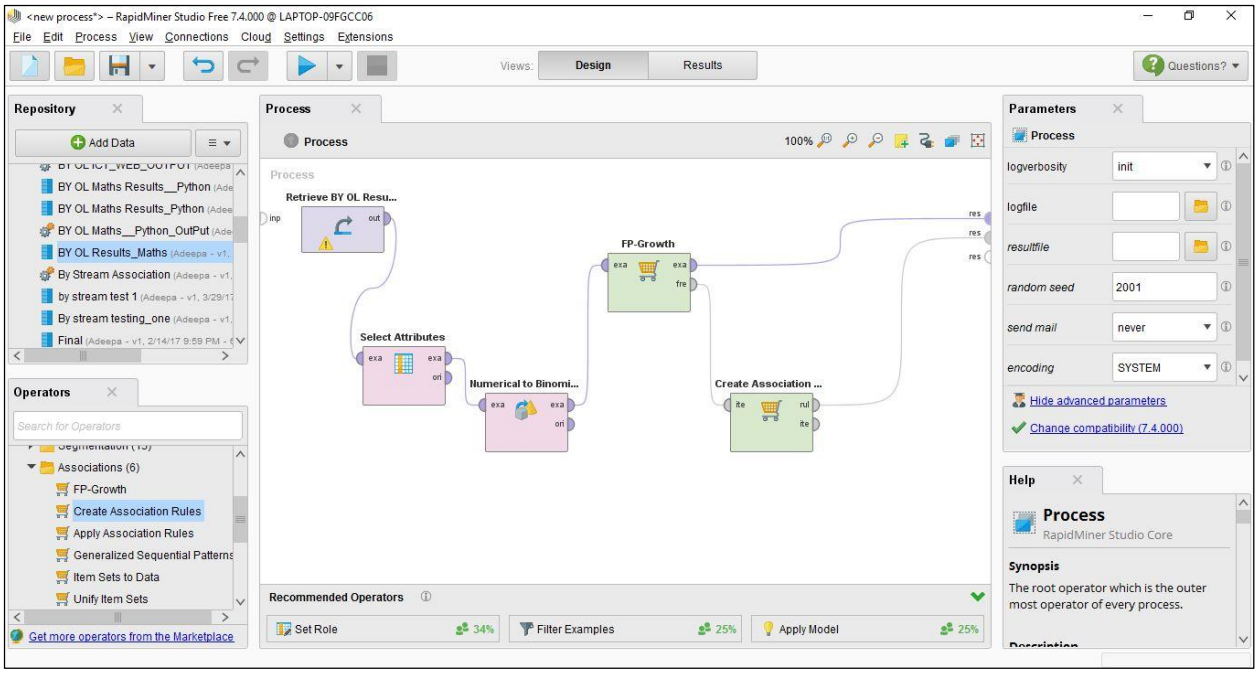
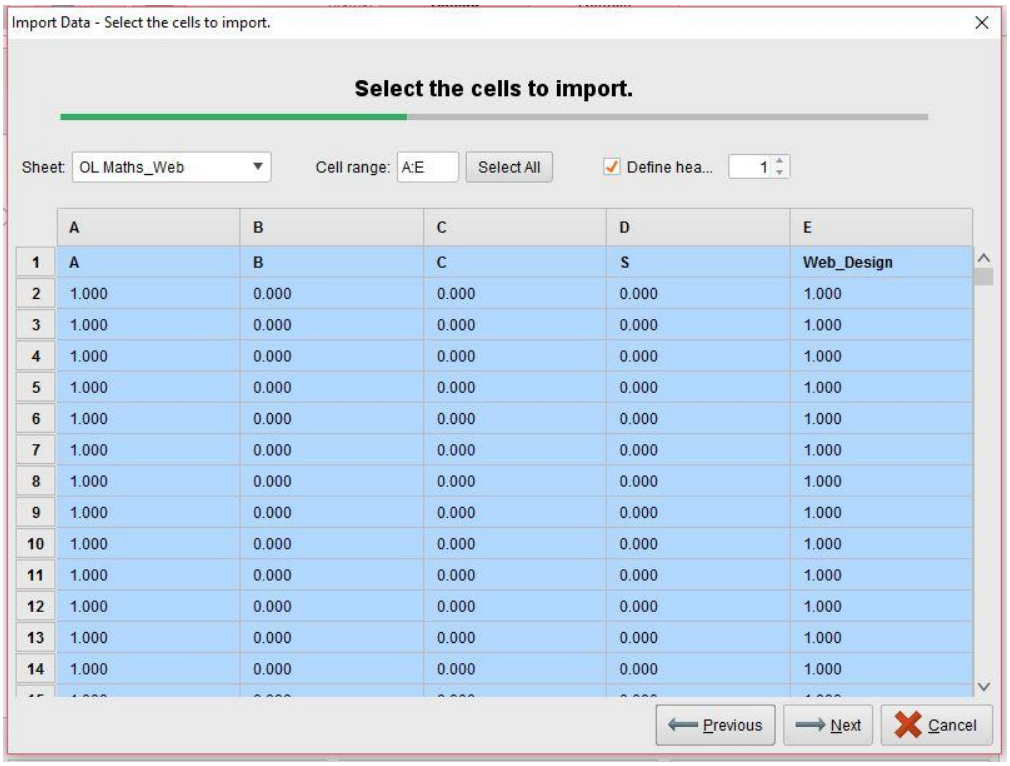


Figure 6.9 Process

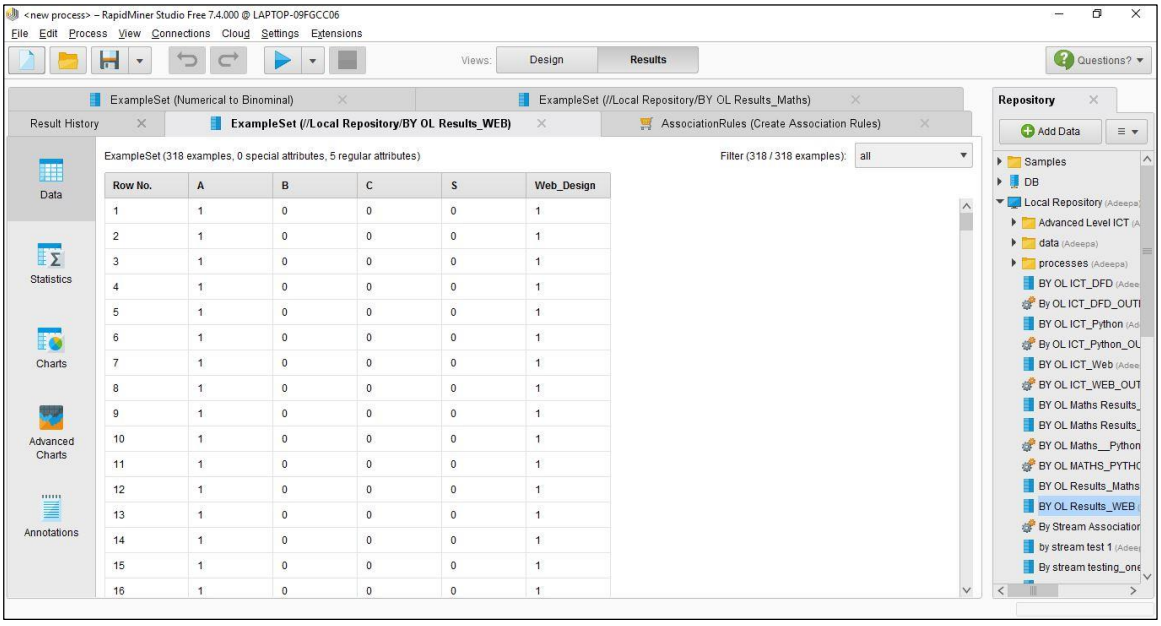
6.5.1.2 Web designing

Out of the 500 students who have been considered as the sample, 386 students had selected the web designing chapter as difficult. Look-over the association between the O\L mathematics result and the level of difficulty of web designing chapter and the effect on the A\L ICT subject result was done.



	A	B	C	D	E
1	A	B	C	S	Web_Design
2	1.000	0.000	0.000	0.000	1.000
3	1.000	0.000	0.000	0.000	1.000
4	1.000	0.000	0.000	0.000	1.000
5	1.000	0.000	0.000	0.000	1.000
6	1.000	0.000	0.000	0.000	1.000
7	1.000	0.000	0.000	0.000	1.000
8	1.000	0.000	0.000	0.000	1.000
9	1.000	0.000	0.000	0.000	1.000
10	1.000	0.000	0.000	0.000	1.000
11	1.000	0.000	0.000	0.000	1.000
12	1.000	0.000	0.000	0.000	1.000
13	1.000	0.000	0.000	0.000	1.000
14	1.000	0.000	0.000	0.000	1.000

Figure 6.10 Insert columns



Row No.	A	B	C	S	Web_Design
1	1	0	0	0	1
2	1	0	0	0	1
3	1	0	0	0	1
4	1	0	0	0	1
5	1	0	0	0	1
6	1	0	0	0	1
7	1	0	0	0	1
8	1	0	0	0	1
9	1	0	0	0	1
10	1	0	0	0	1
11	1	0	0	0	1
12	1	0	0	0	1
13	1	0	0	0	1
14	1	0	0	0	1
15	1	0	0	0	1
16	1	0	0	0	1

Figure 6.11 Selected data set

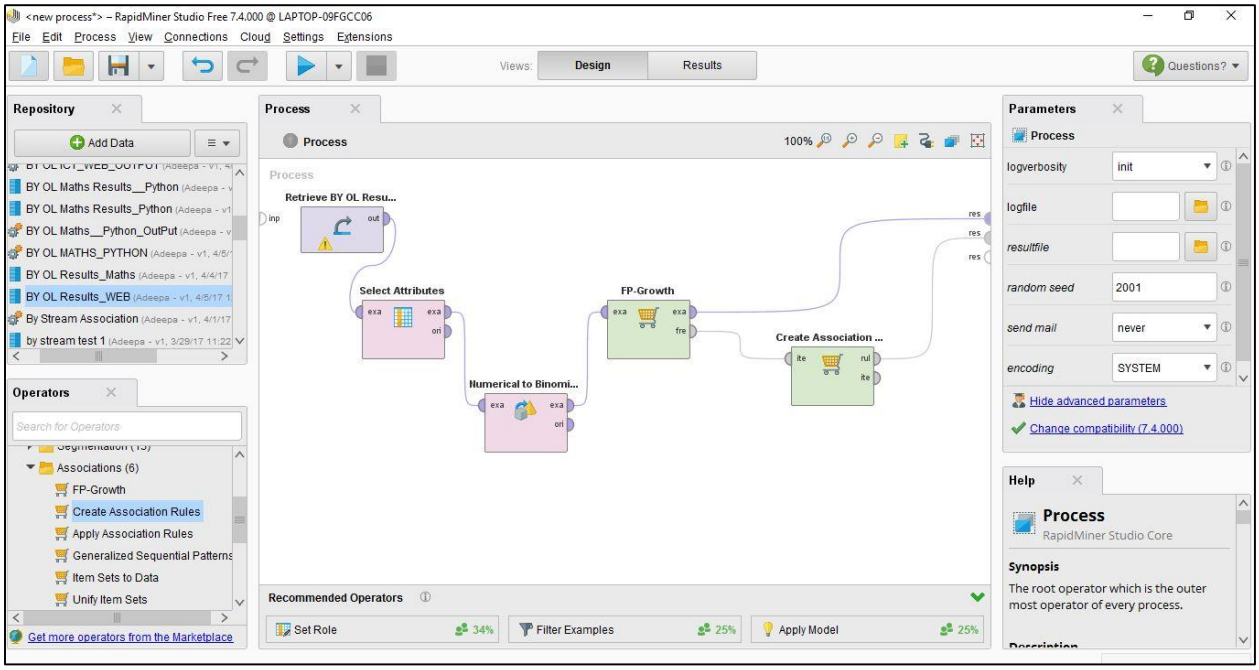
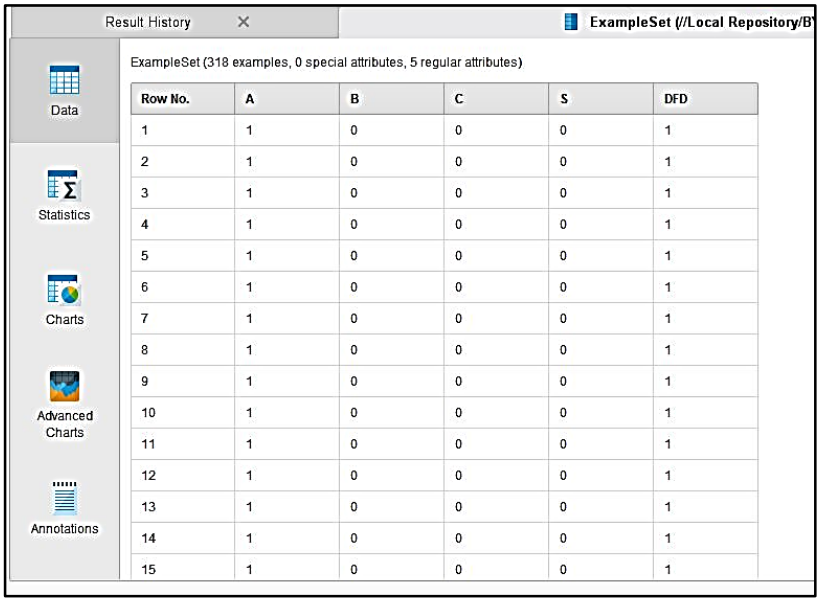


Figure 6.12 Process

6.5.1.3 DFD

Out of the 500 students who have been considered as the sample and 389 students had selected the DFD chapter as difficult. Look over the association between the O\L mathematics result and the level of difficulty of DFD chapter and its effect on the A\L ICT subject result was done.



Row No.	A	B	C	S	DFD
1	1	0	0	0	1
2	1	0	0	0	1
3	1	0	0	0	1
4	1	0	0	0	1
5	1	0	0	0	1
6	1	0	0	0	1
7	1	0	0	0	1
8	1	0	0	0	1
9	1	0	0	0	1
10	1	0	0	0	1
11	1	0	0	0	1
12	1	0	0	0	1
13	1	0	0	0	1
14	1	0	0	0	1
15	1	0	0	0	1

Figure 6.13 Selected Data set

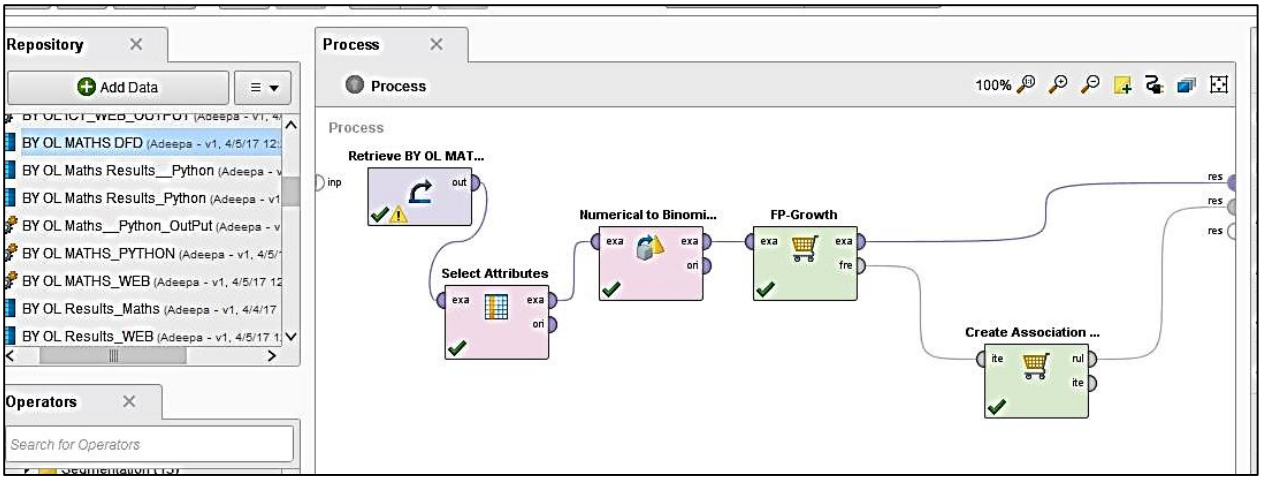
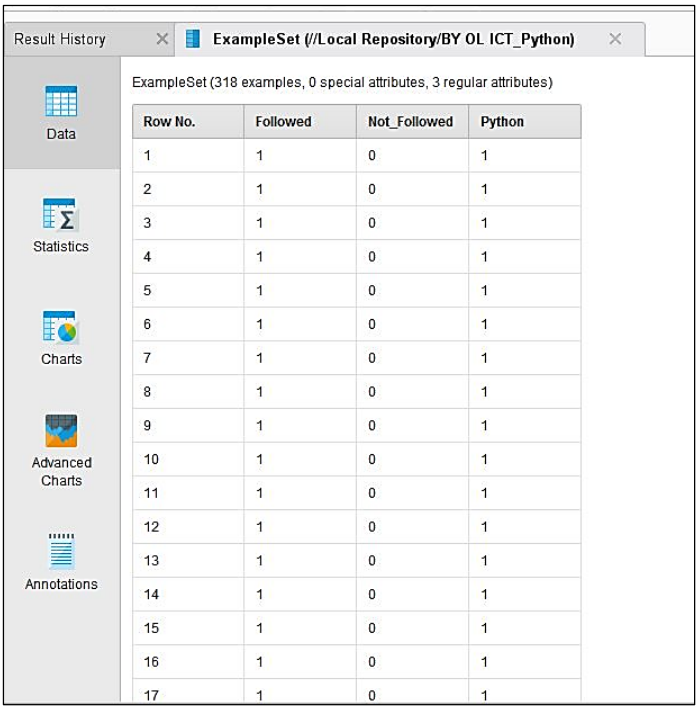


Figure 6.14 Process

6.5.2 Whether followed ICT or Not followed

6.5.2.1 Python

From the presence of selected sample, 443 students had selected the python chapter as difficult. From them, 173 students had followed ICT for O\L exam and 270 students had not followed ICT subject for O\L exam. Look- over the association between the level of difficulty of python chapter and effect on O\L ICT subject was done.



Row No.	Followed	Not_Followed	Python
1	1	0	1
2	1	0	1
3	1	0	1
4	1	0	1
5	1	0	1
6	1	0	1
7	1	0	1
8	1	0	1
9	1	0	1
10	1	0	1
11	1	0	1
12	1	0	1
13	1	0	1
14	1	0	1
15	1	0	1
16	1	0	1
17	1	0	1

Figure 6.15 Select columns

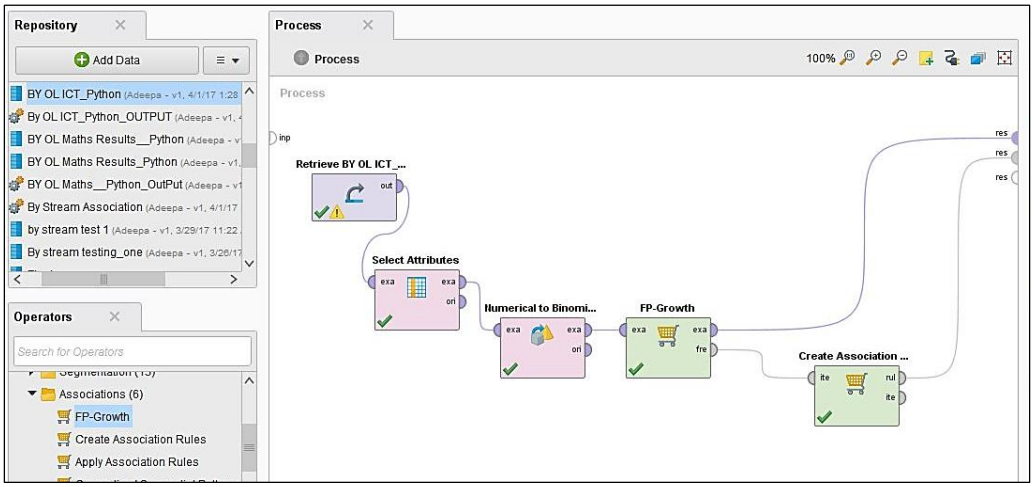
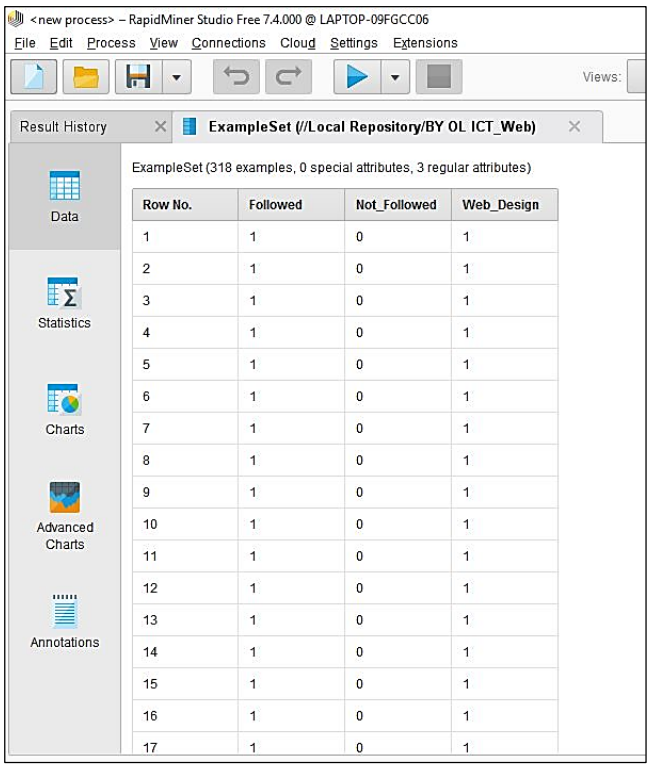


Figure 6.16 Process

6.5.2.2 Web Designing

From the presence of selected sample 386 students had selected the Web Design chapter as difficult from them, 87 students had followed ICT for O\L exam and 298 students had not followed ICT subject for O\L exam.Look over the association between the level of difficulty of Web Design chapter and effect on O\L ICT subject was done.



Row No.	Followed	Not_Followed	Web_Design
1	1	0	1
2	1	0	1
3	1	0	1
4	1	0	1
5	1	0	1
6	1	0	1
7	1	0	1
8	1	0	1
9	1	0	1
10	1	0	1
11	1	0	1
12	1	0	1
13	1	0	1
14	1	0	1
15	1	0	1
16	1	0	1
17	1	0	1

Figure 6.18 Select columns Web Design

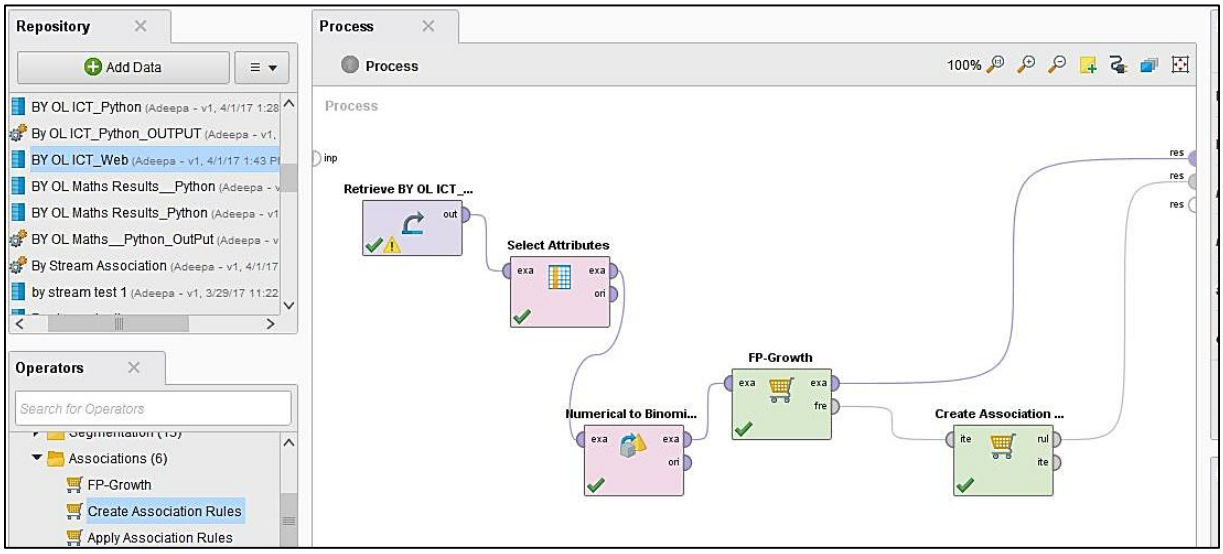
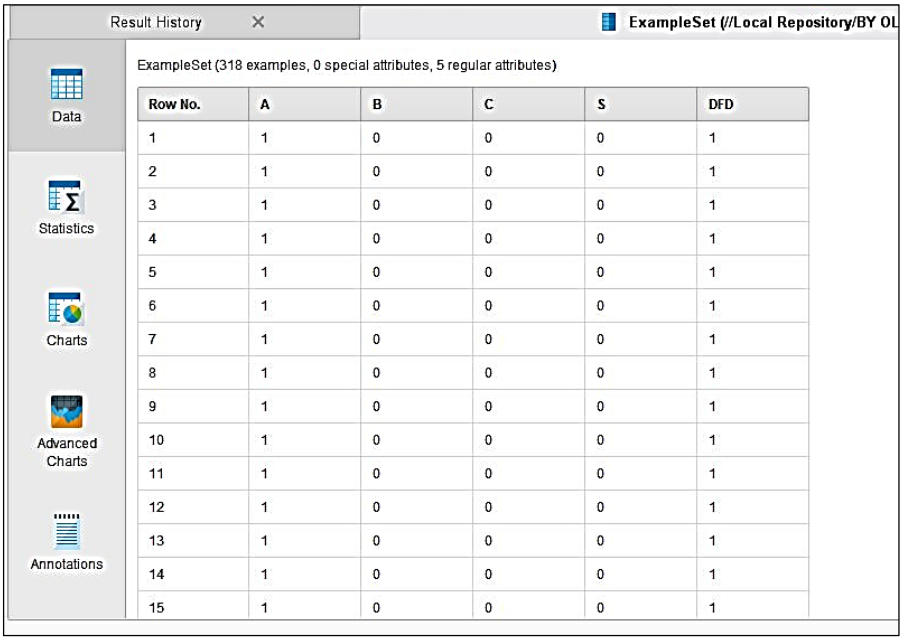


Figure 6.19 Process Web Design

6.5.2.3 DFD

From the presence of selected sample 389 students had selected the DFD chapter as difficult. From them, 187 students had followed ICT for O\L exam and 200 students had not followed ICT subject for O\L exam. Look-over the association between the level of difficulty of DFD chapter and effect on O\L ICT subject was done.



Row No.	A	B	C	S	DFD
1	1	0	0	0	1
2	1	0	0	0	1
3	1	0	0	0	1
4	1	0	0	0	1
5	1	0	0	0	1
6	1	0	0	0	1
7	1	0	0	0	1
8	1	0	0	0	1
9	1	0	0	0	1
10	1	0	0	0	1
11	1	0	0	0	1
12	1	0	0	0	1
13	1	0	0	0	1
14	1	0	0	0	1
15	1	0	0	0	1

Figure 6.20 select columns

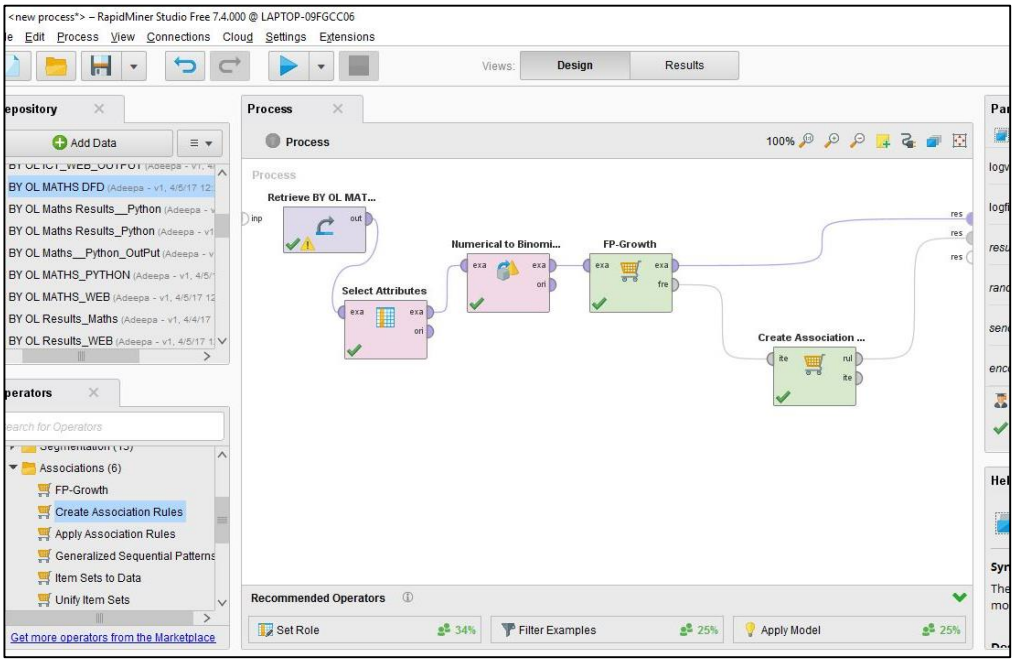
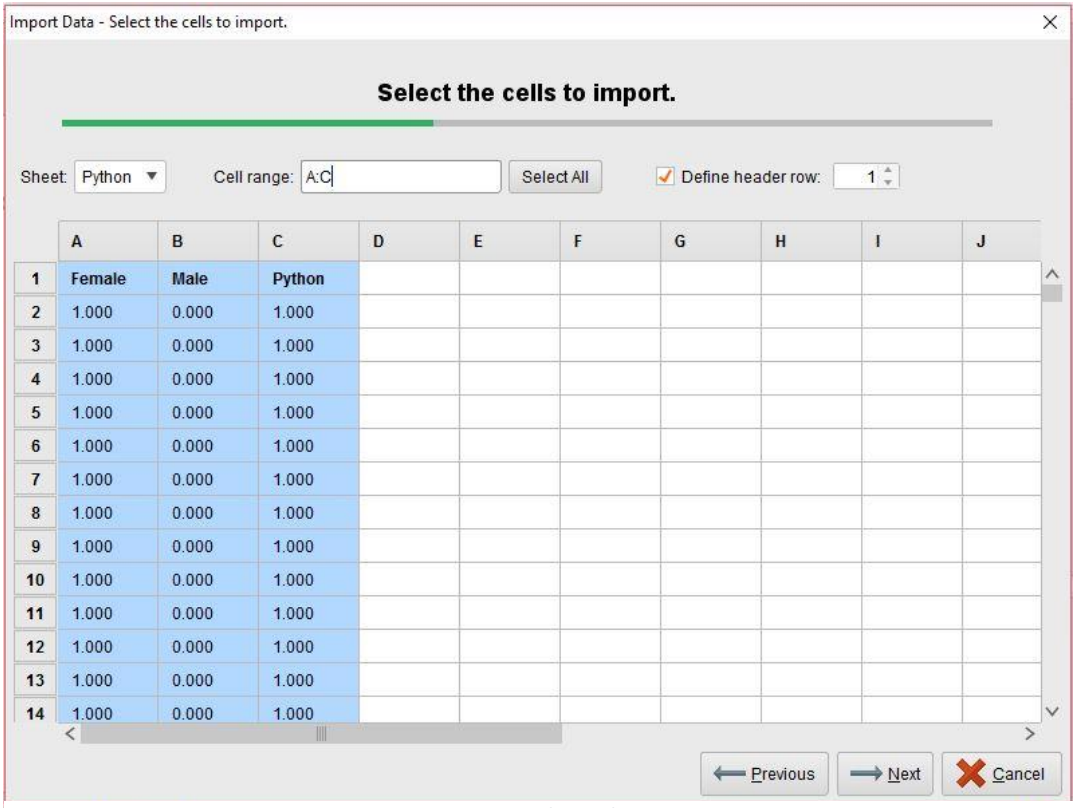


Figure 6.21 Process DFD

6.5.3 Analysis by Gender

6.5.3.1 By Python



	A	B	C	D	E	F	G	H	I	J
1	Female	Male	Python							
2	1.000	0.000	1.000							
3	1.000	0.000	1.000							
4	1.000	0.000	1.000							
5	1.000	0.000	1.000							
6	1.000	0.000	1.000							
7	1.000	0.000	1.000							
8	1.000	0.000	1.000							
9	1.000	0.000	1.000							
10	1.000	0.000	1.000							
11	1.000	0.000	1.000							
12	1.000	0.000	1.000							
13	1.000	0.000	1.000							
14	1.000	0.000	1.000							

Figure 6.22 Select columns

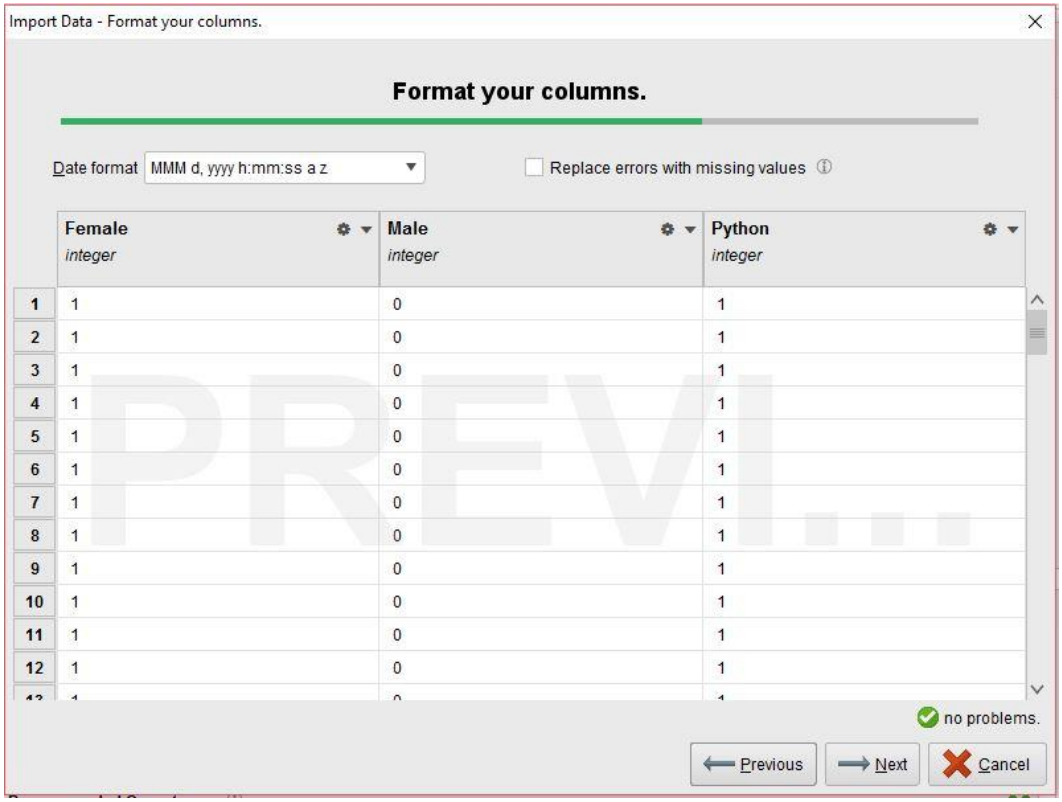


Figure 6.24 Format columns

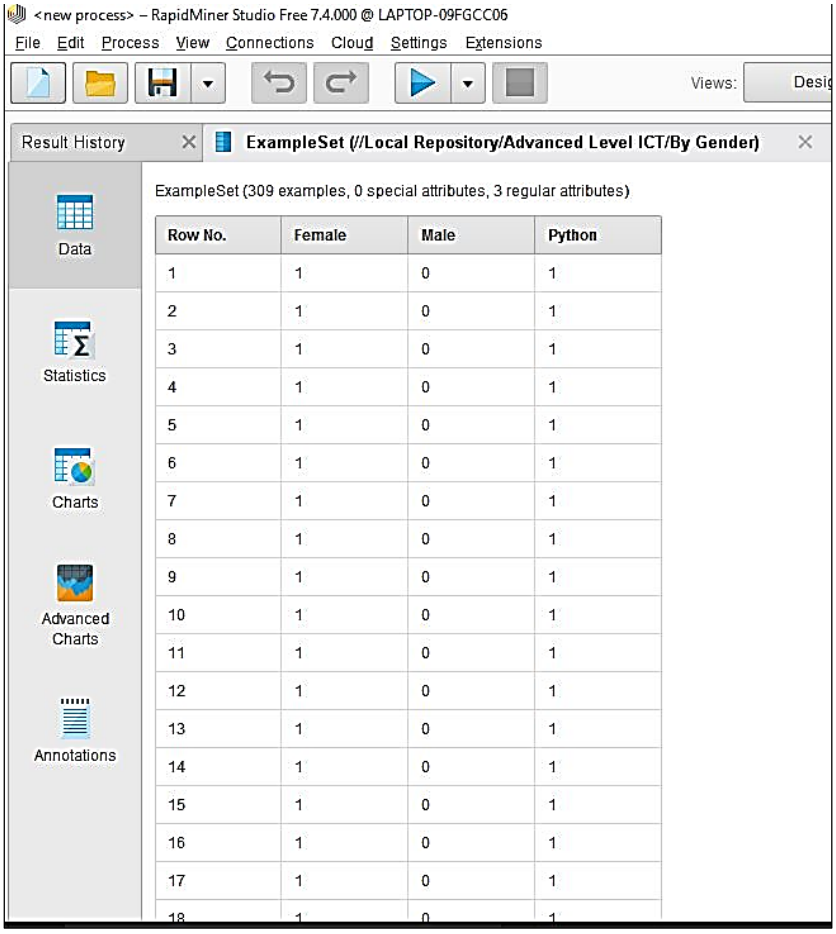


Figure 6.25 Selected data set

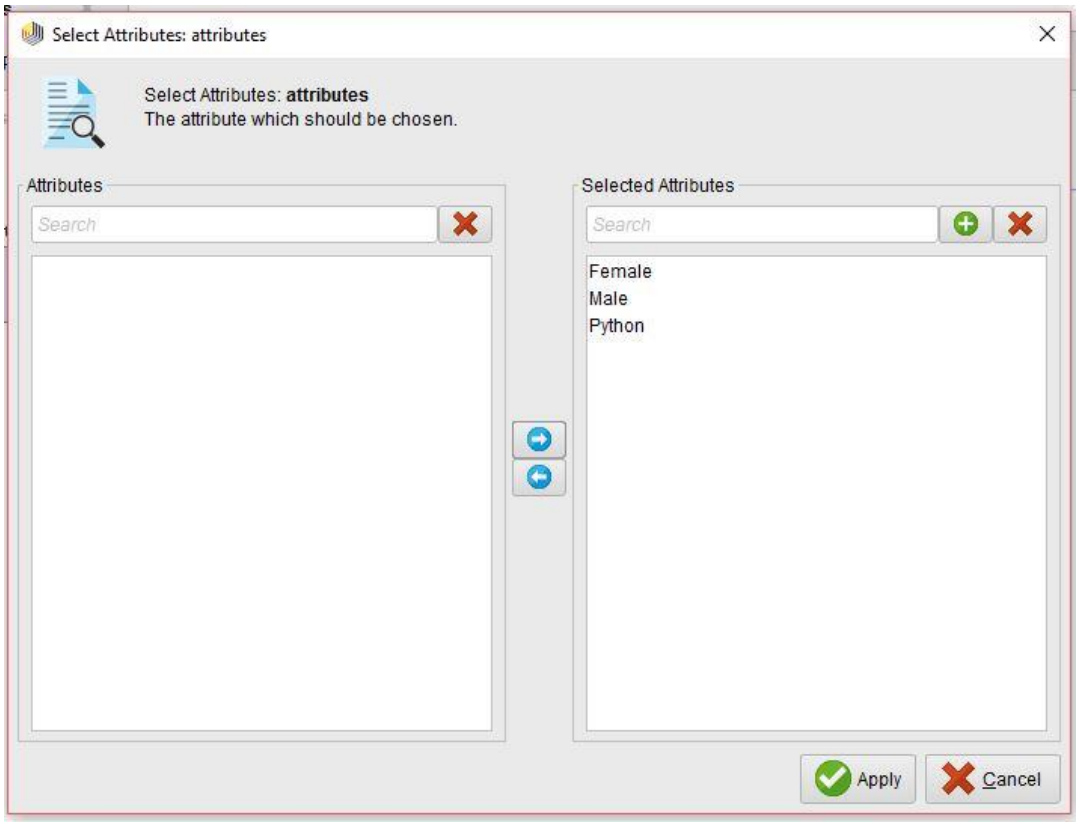


Figure 6.26 Select attribute

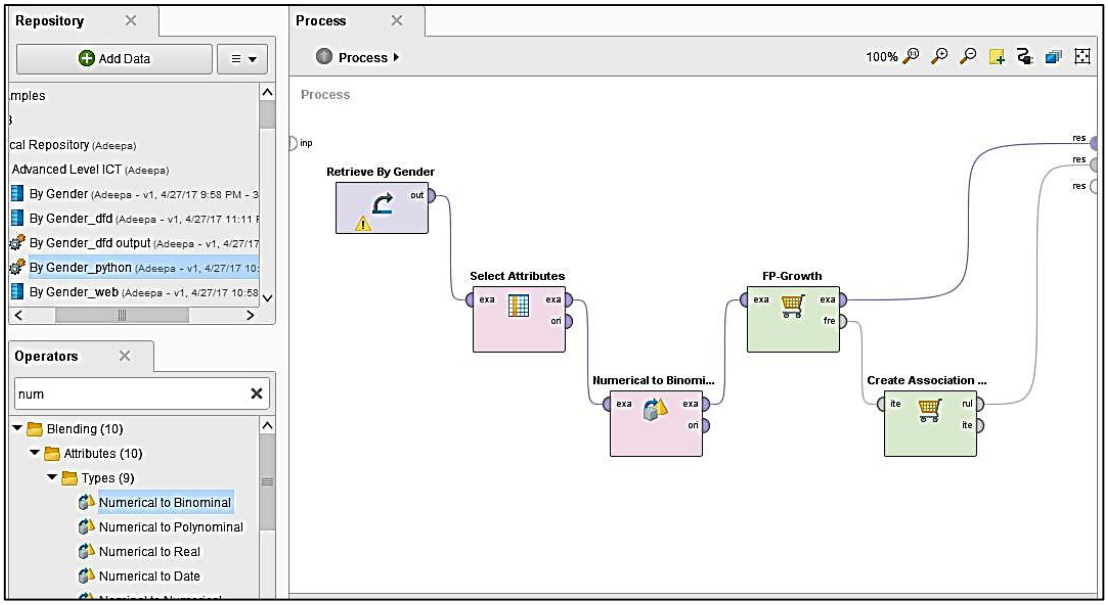


Figure 6.27 Process with operators

6.5.3.2 By web

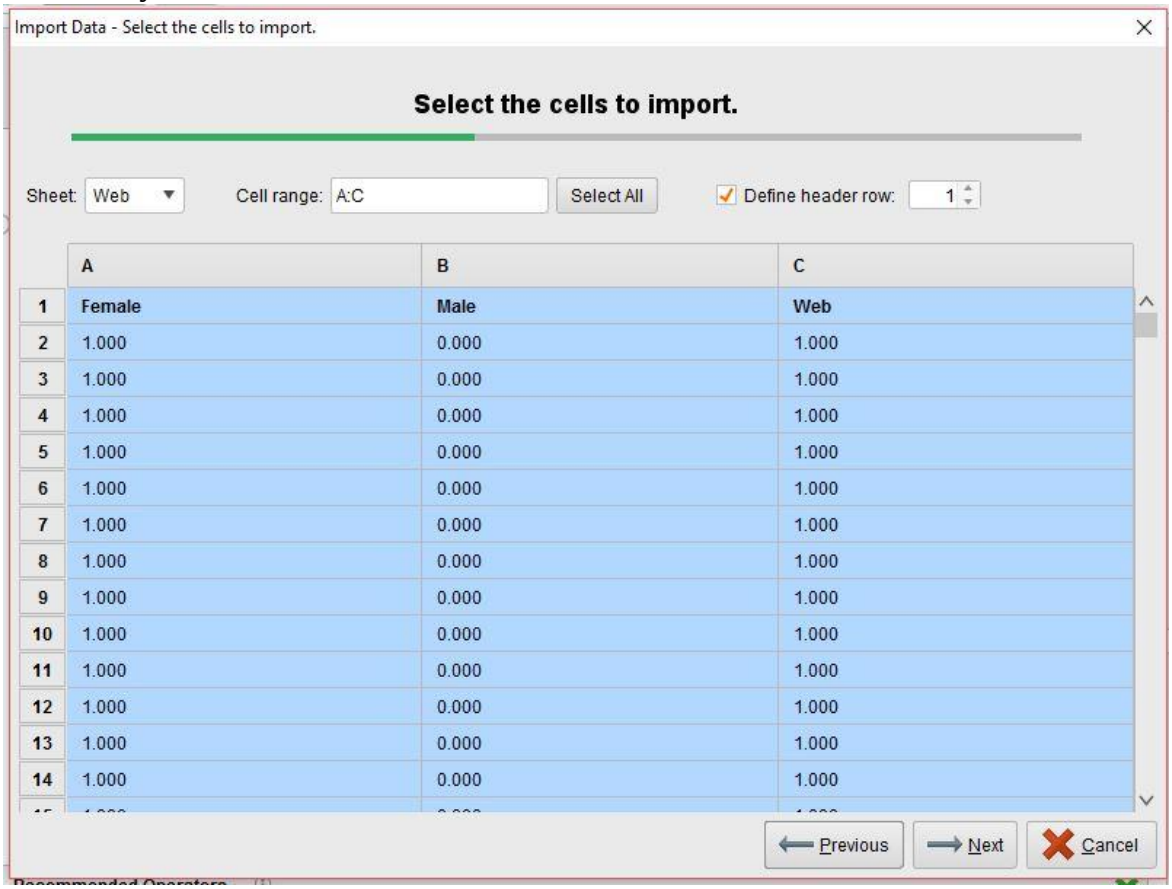


Figure 6.28 Select Cells to import

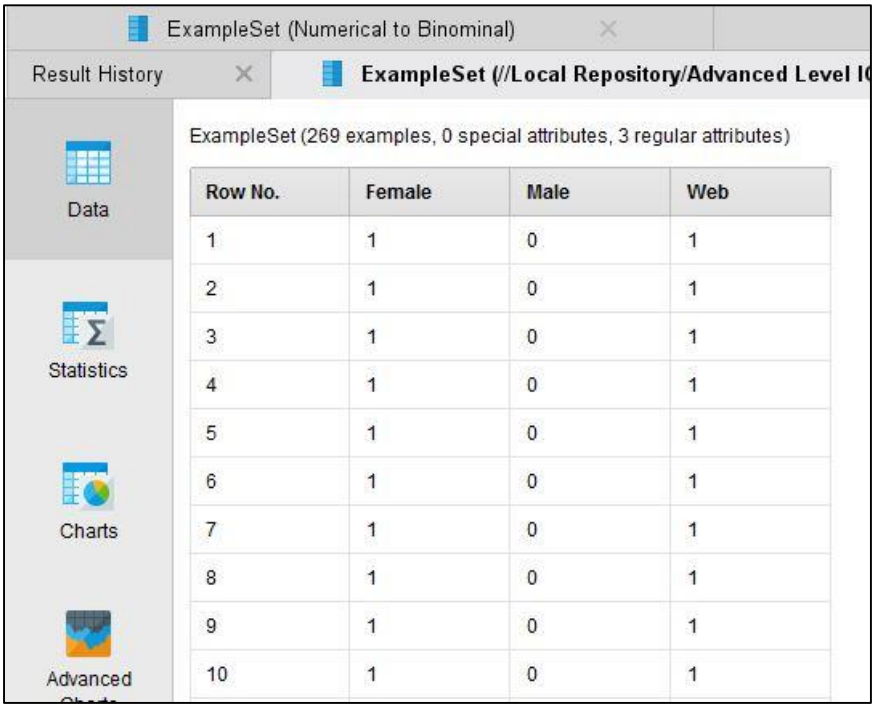


Figure 6.29 Converted Numerical to Binominal

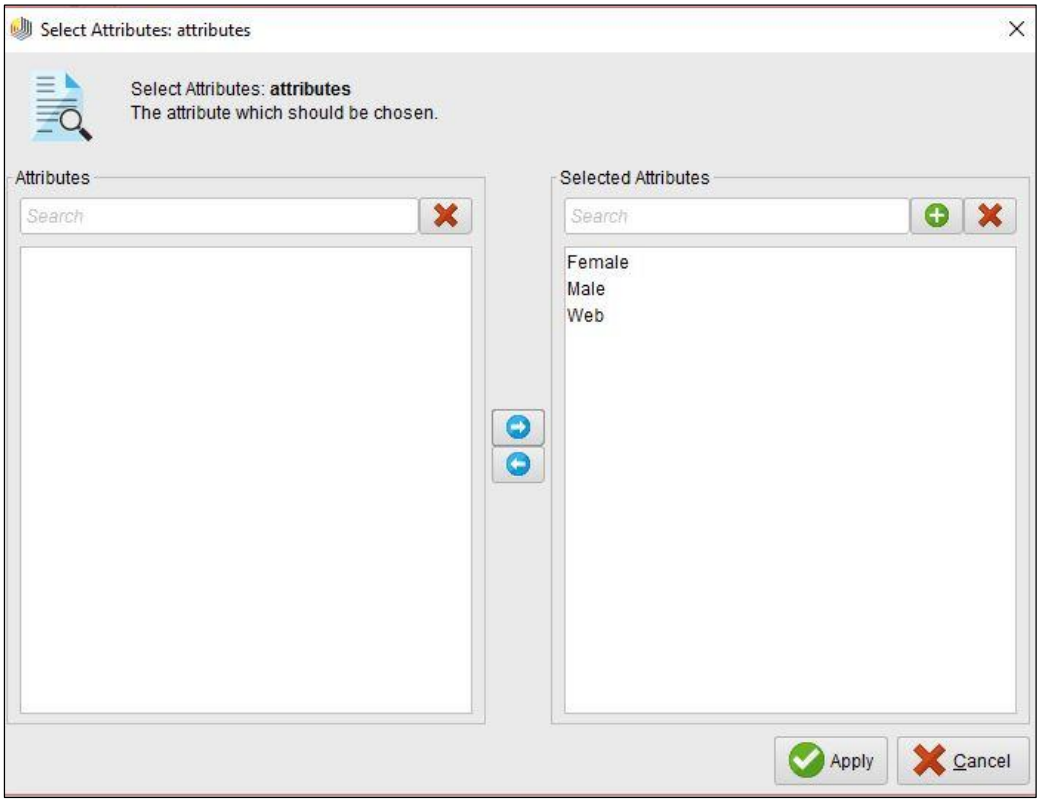


Figure 6.30 Select Attributes

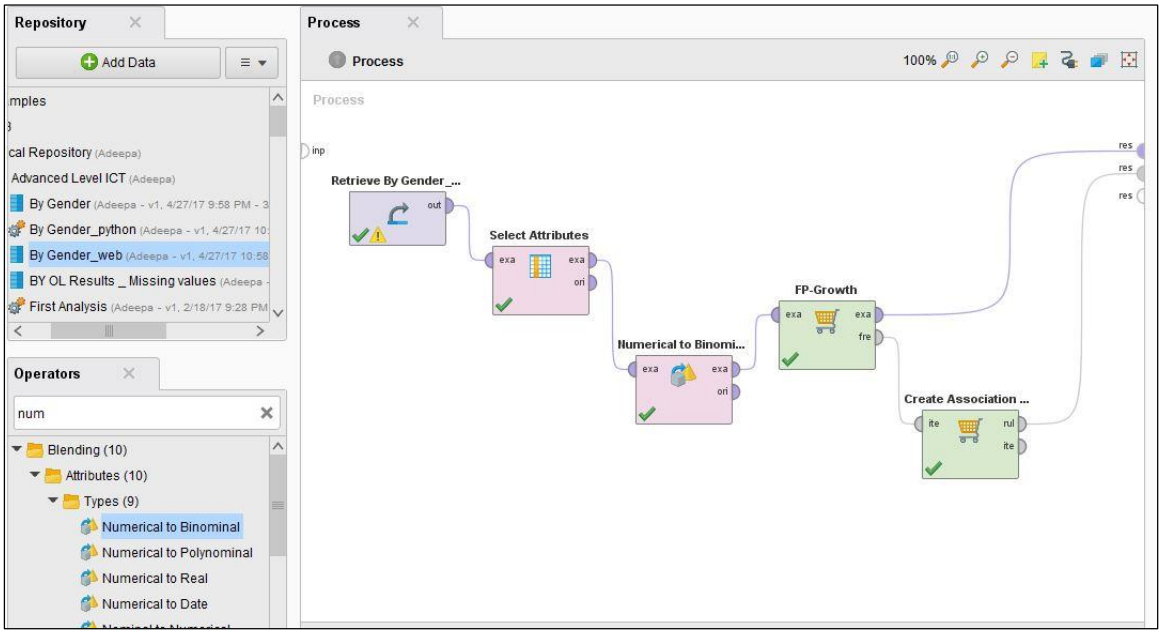
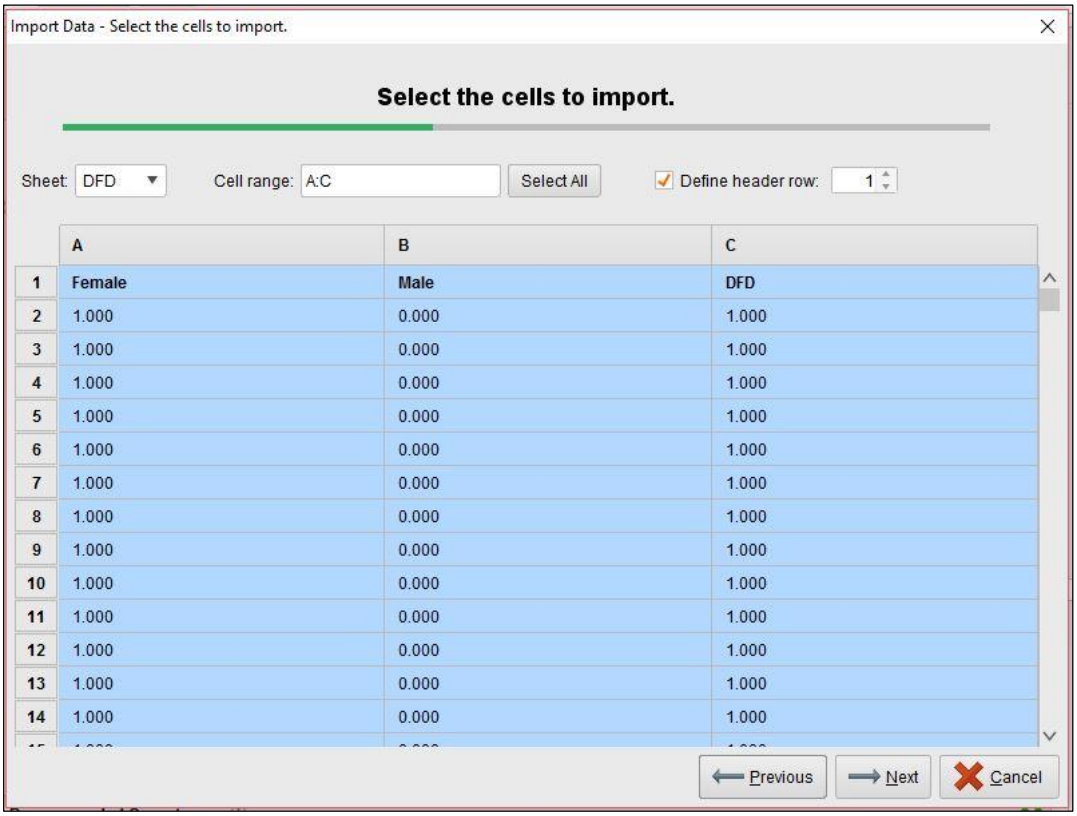


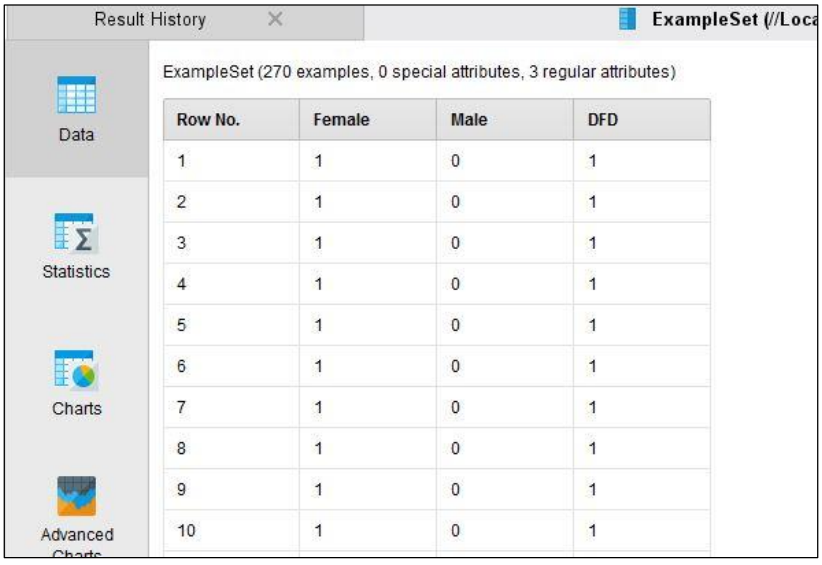
Figure 6.31 Process with operators

6.5.3.3 By DFD



	A	B	C
1	Female	Male	DFD
2	1.000	0.000	1.000
3	1.000	0.000	1.000
4	1.000	0.000	1.000
5	1.000	0.000	1.000
6	1.000	0.000	1.000
7	1.000	0.000	1.000
8	1.000	0.000	1.000
9	1.000	0.000	1.000
10	1.000	0.000	1.000
11	1.000	0.000	1.000
12	1.000	0.000	1.000
13	1.000	0.000	1.000
14	1.000	0.000	1.000

Figure 6.33 Select the cells to import



Row No.	Female	Male	DFD
1	1	0	1
2	1	0	1
3	1	0	1
4	1	0	1
5	1	0	1
6	1	0	1
7	1	0	1
8	1	0	1
9	1	0	1
10	1	0	1

Figure 6.32 Select data set

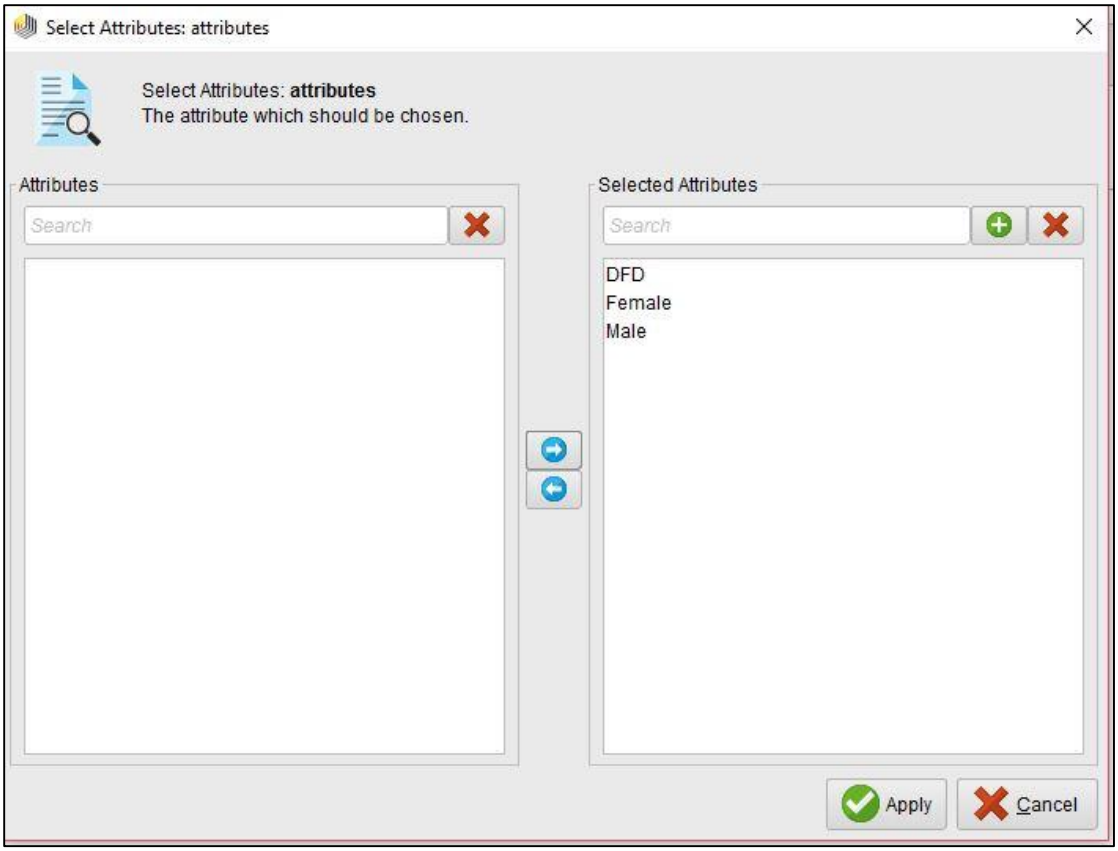


Figure 6.34 Select Attributes

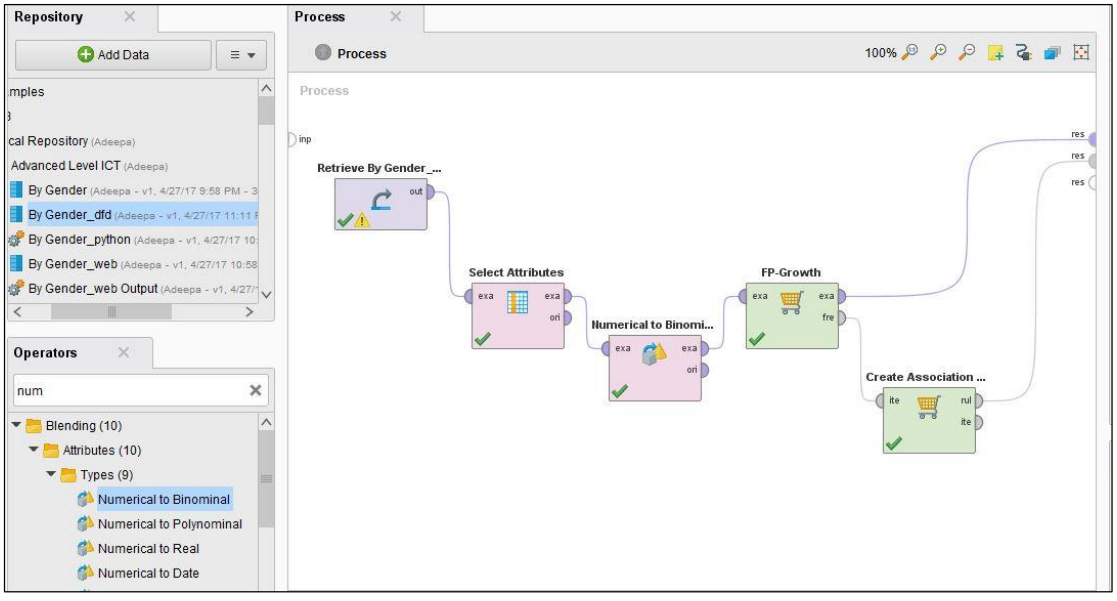


Figure 6.35 Process with operators

6.6 Analysis by the Stream sat for at the Advanced Level Exam

ICT subject can be combined with all five streams equally. Student can select ICT subject with any two subjects in each stream and sit for the university entrance exam. There are five main streams currently in the Sri Lankan school education system.

1. Commerce
2. Arts
3. Other
4. Engineering Technology
5. Bio Systems Technology

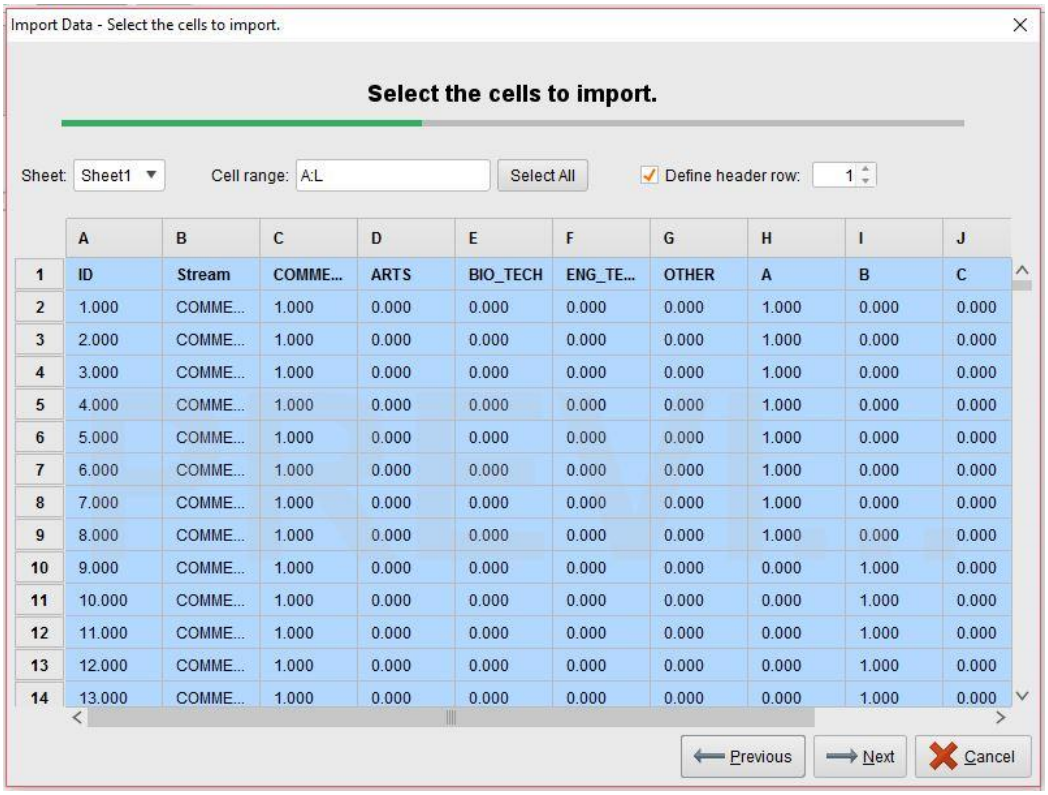
Total candidates sat for the ICT subject at the year 2016 are 18664. Among them, 2075 students were from commerce stream, 3951 students were from Arts stream, 2688 were from other stream, 8466 were from Engineering Technology stream and 1484 were from Bio System Stream.

Association Rule

Association rules are created by analyzing data under frequent if/then patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in the database. Confidence indicates the number of times the if/then statements have been found to be true. The frequent if/then patterns are mined using the operators like the FP-Growth operator. The Create Association Rules operator takes these frequent itemsets and generates association rules.

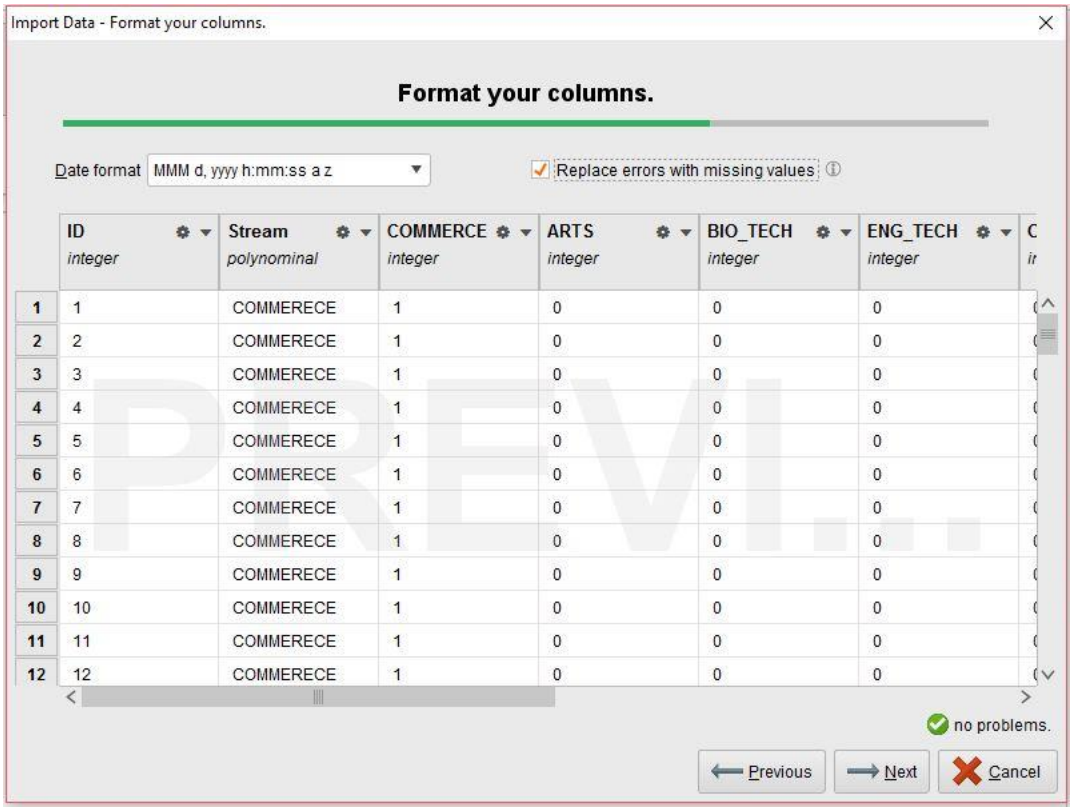
The "RESULTS BY STREAM_Association" data set is loaded using the Retrieve operator. Thus the FP-Growth operator cannot be applied on it directly because the FP-Growth operator requires all attributes to be binominal. We have to do some preprocessing to mold the ExampleSet into desired form. The Numerical to Binominal operator is applied to change these numerical attributes to binominal attributes. Finally, the FP-Growth operator is applied to generate frequent item sets. The frequent item sets generated from the FP-Growth operator are provided to the Create Association Rules operator. The resultant association rules can be viewed in the Results Workspace.

Figure 6.20 to **Error! Reference source not found.** shows this process.



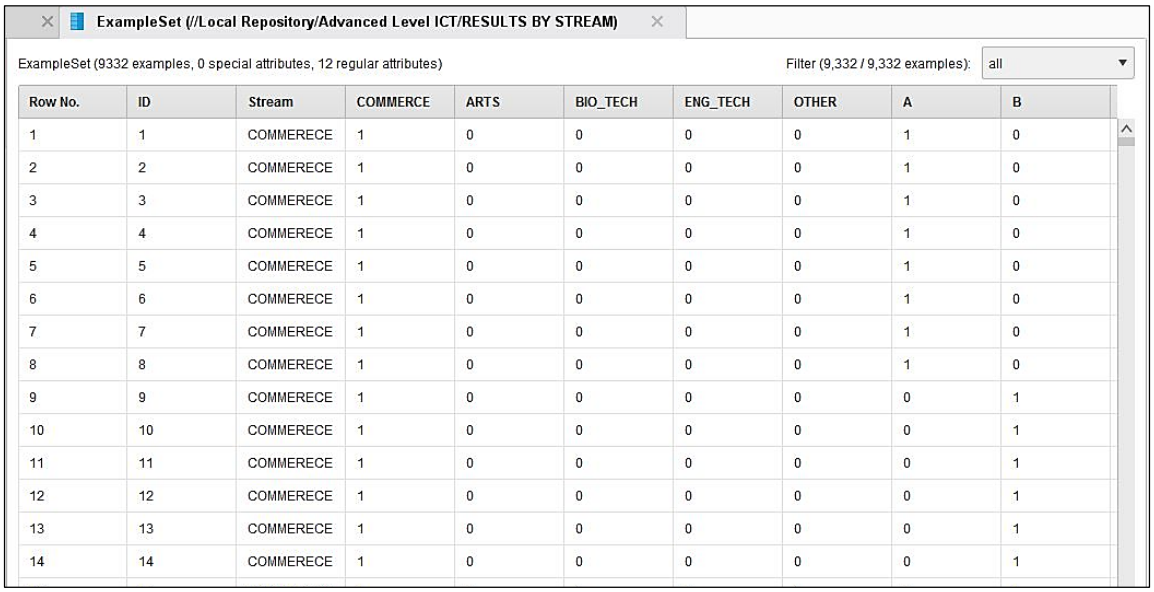
	A	B	C	D	E	F	G	H	I	J
1	ID	Stream	COMME...	ARTS	BIO_TECH	ENG_TE...	OTHER	A	B	C
2	1.000	COMME...	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
3	2.000	COMME...	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
4	3.000	COMME...	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
5	4.000	COMME...	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
6	5.000	COMME...	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
7	6.000	COMME...	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
8	7.000	COMME...	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
9	8.000	COMME...	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
10	9.000	COMME...	1.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
11	10.000	COMME...	1.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
12	11.000	COMME...	1.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
13	12.000	COMME...	1.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
14	13.000	COMME...	1.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000

Figure 6.36 Select cells to import



	ID *integer*	Stream *polynominal*	COMMERCE *integer*	ARTS *integer*	BIO_TECH *integer*	ENG_TECH *integer*
1	1	COMMERECE	1	0	0	0
2	2	COMMERECE	1	0	0	0
3	3	COMMERECE	1	0	0	0
4	4	COMMERECE	1	0	0	0
5	5	COMMERECE	1	0	0	0
6	6	COMMERECE	1	0	0	0
7	7	COMMERECE	1	0	0	0
8	8	COMMERECE	1	0	0	0
9	9	COMMERECE	1	0	0	0
10	10	COMMERECE	1	0	0	0
11	11	COMMERECE	1	0	0	0
12	12	COMMERECE	1	0	0	0

Figure 6.37 Format columns



ExampleSet (//Local Repository/Advanced Level ICT/RESULTS BY STREAM)

ExampleSet (9332 examples, 0 special attributes, 12 regular attributes) Filter (9,332 / 9,332 examples): all

Row No.	ID	Stream	COMMERCE	ARTS	BIO_TECH	ENG_TECH	OTHER	A	B
1	1	COMMERECE	1	0	0	0	0	1	0
2	2	COMMERECE	1	0	0	0	0	1	0
3	3	COMMERECE	1	0	0	0	0	1	0
4	4	COMMERECE	1	0	0	0	0	1	0
5	5	COMMERECE	1	0	0	0	0	1	0
6	6	COMMERECE	1	0	0	0	0	1	0
7	7	COMMERECE	1	0	0	0	0	1	0
8	8	COMMERECE	1	0	0	0	0	1	0
9	9	COMMERECE	1	0	0	0	0	0	1
10	10	COMMERECE	1	0	0	0	0	0	1
11	11	COMMERECE	1	0	0	0	0	0	1
12	12	COMMERECE	1	0	0	0	0	0	1
13	13	COMMERECE	1	0	0	0	0	0	1
14	14	COMMERECE	1	0	0	0	0	0	1

Figure 6.39 Selected data set

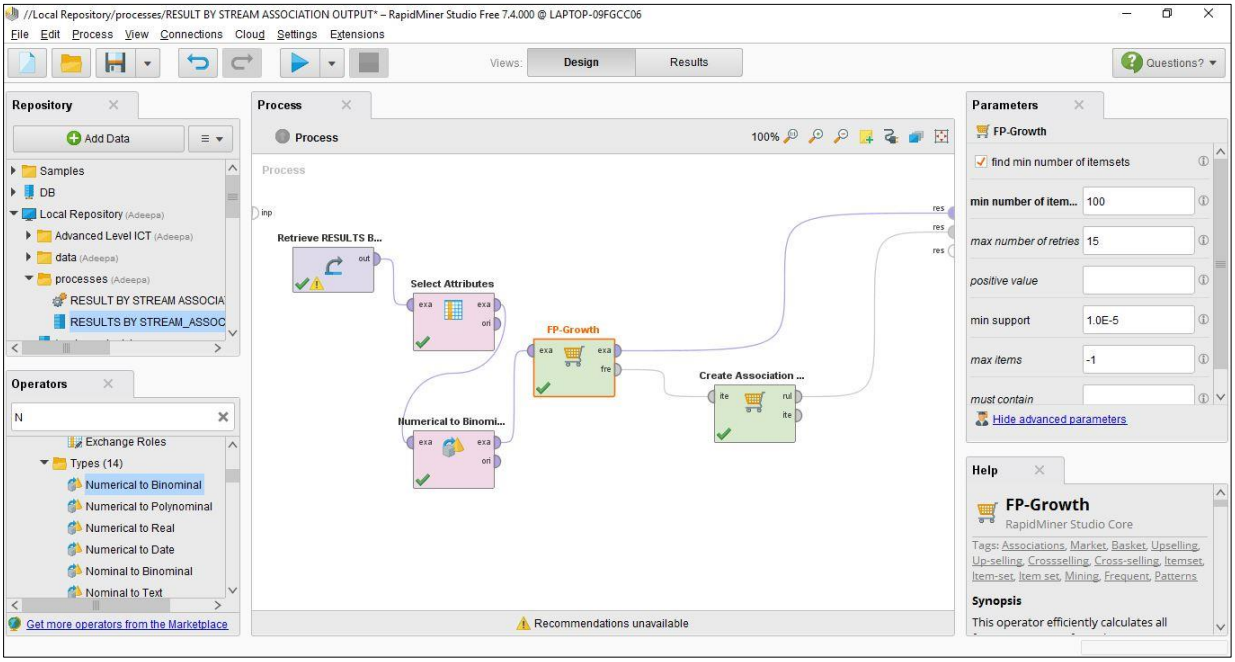


Figure 6.40 Process with operators

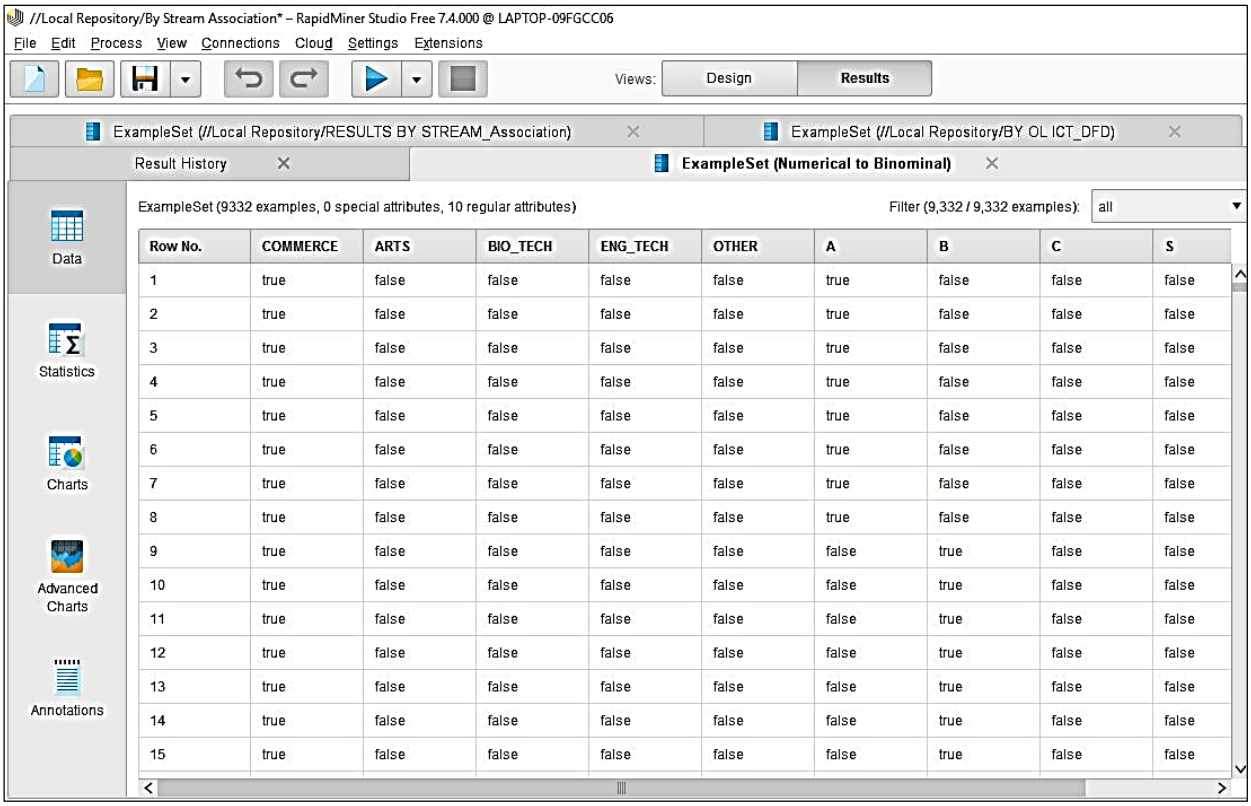


Figure 6.41 Numerical to Binominal converted data set

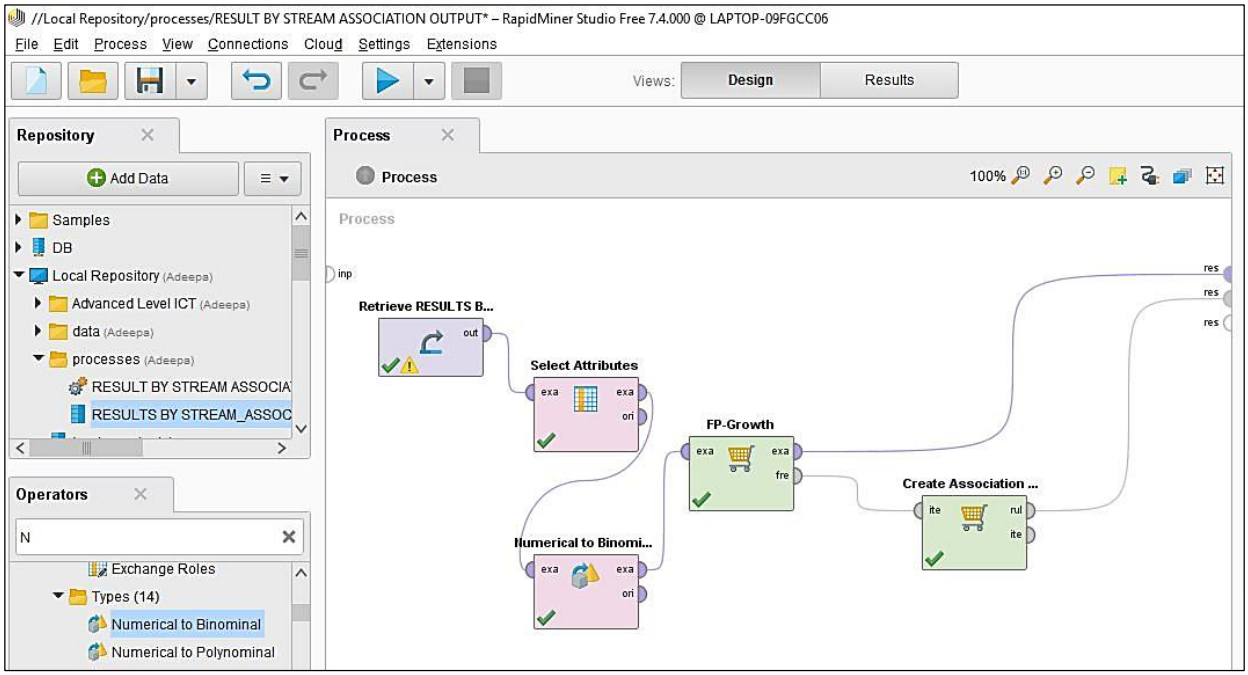


Figure 6.42 Process with operators

Chapter 7

7 Evaluation

7.1 Introduction

Previous chapter discussed the details on implementation of proposed areas of analysis that could have been analyzed the performance of the ICT results of year 2016. This chapter justifies and evaluates the results and the model generated for each module.

7.2 Analysis by Questions on Exam paper

From the present criteria it has been considered whether there is an impact on the level of difficulty from the grade obtained for the O\L Mathematics subject, whether there is an impact on O\L ICT followed or not followed and the impact on the Gender difference. To accomplish this task the performances for each question that each candidate achieved have considered. When considered 2016 G.C.E. Advanced level examination ICT question paper 85.07 candidates obtained only 0-25 marks for the Python programming question, 78.53 candidates obtained only 0-25 marks for the Web design question and 65.9 candidates obtained only 0-25 marks for the DFD question. Therefore it's more preferable to consider those three chapters to investigate the performances in ICT at Advanced level.

7.2.1 Evaluation on Grade obtained for Ordinary Level Mathematics.

Association Rules were generated to evaluate whether Grade obtained for ordinary level Mathematics has specific relationship with the G.C.E. Advanced Level ICT Results. Each association rule has a separate confidence value, lift value and number of occurrences.

7.2.1.1 By Python Chapter

500 students have been considered as the sample and 443 students had selected the python chapter as a difficult chapter. Figure 7.1 shows the set of association rules generated under above conditions using the 2016 G.C.E. Advanced level ICT results of the students.

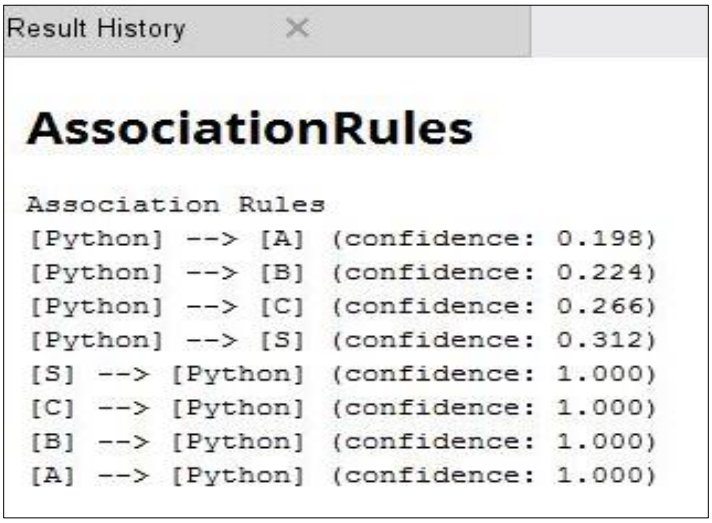


Figure 7.2 Association Rule result python by mathematics

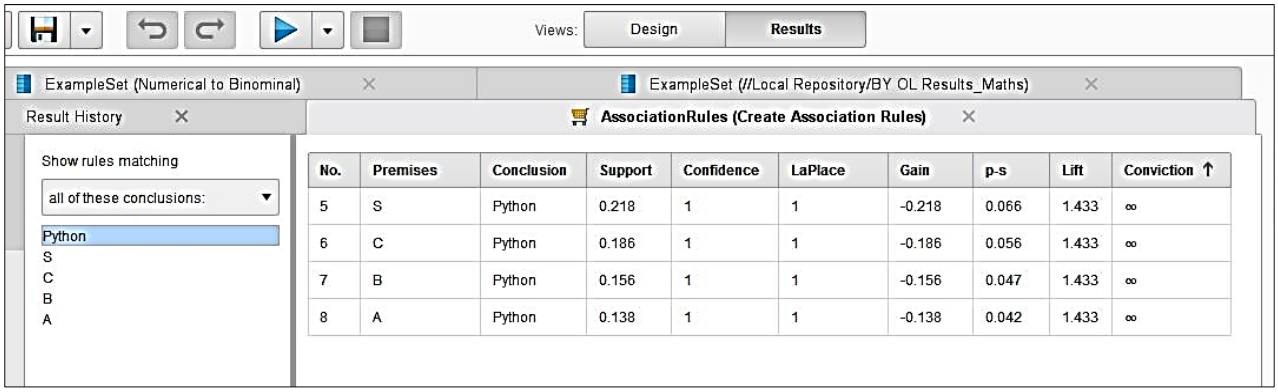


Figure 7.3 Association rule for python chapter by Mathematics

From 443 students who selected python as a difficult question, 57 students obtained A Grades for Mathematics at O\L. Therefore the support value is 0.138, confidence level is 1, and lift value is 1.433. From 443 student who selected python as difficult question 93 students obtained B Grades for Mathematics at O\L. Hence the support value is 0.156, confidence level is 1, and lift value is 1.433.

From 443 student who selected python as a difficult question 129 students obtained C Grades for Mathematics at O\L. Therefore the support value is 0.186, confidence level is 1, and lift value is 1.433.

From 443 student who selected python as a difficult question 164 students obtained S Grades for Mathematics at O\L. Hence the support value is 0.218, confidence level is 1, and lift value is 1.433.

After ordering the above support values in descending order most of the students who obtained S or C passes for mathematics at ordinary level took Python as a difficult chapter.

In accordance with the above association rules, outputs obtaining A or B grade for mathematics is more advisable to follow Advanced Level ICT subject according to the support values of the above results.

7.2.1.2 By Web Design Chapter

500 students have considered as the sample and 386 students had selected the web design chapter as difficult Figure 7.3 shows the set of association rules generated under above conditions using the 2016 G.C.E. Advanced level ICT results of the students.

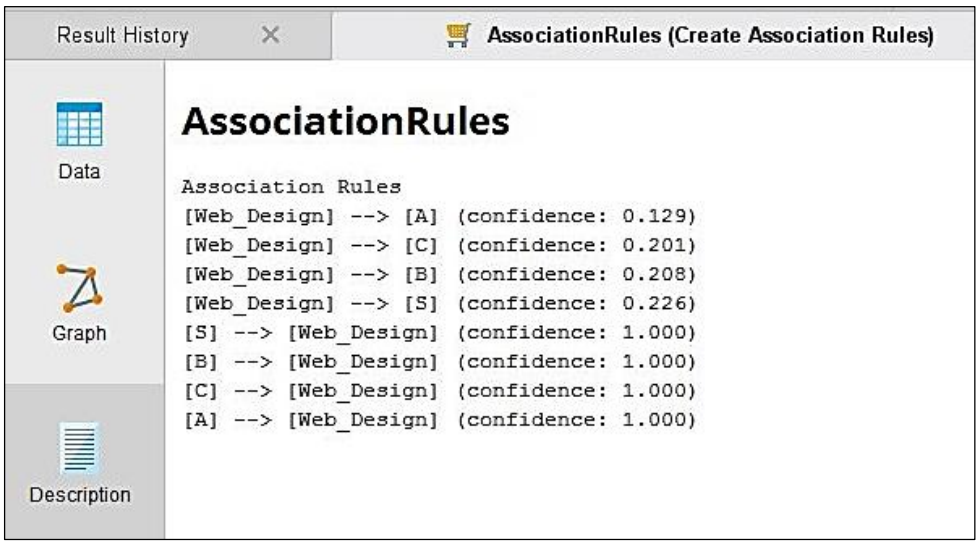
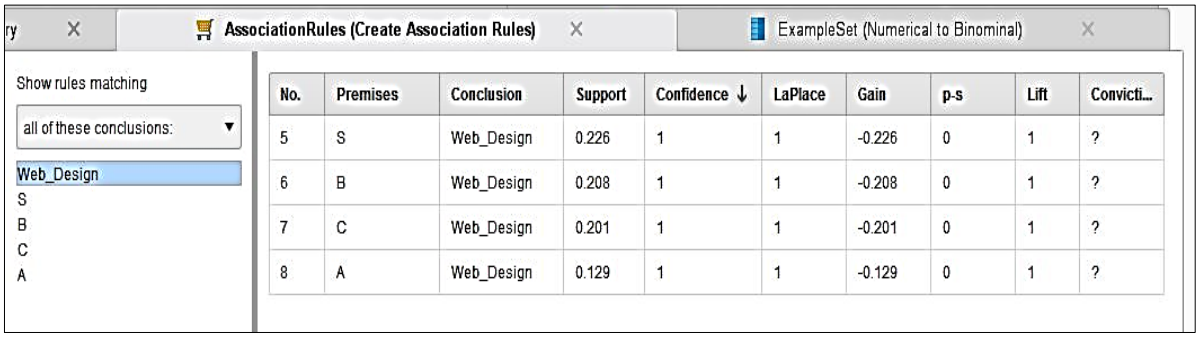


Figure 7.4 Association Rule Result Web Design and Mathematics



No.	Premises	Conclusion	Support	Confidence ↓	LaPlace	Gain	p-s	Lift	Convicti...
5	S	Web_Design	0.226	1	1	-0.226	0	1	?
6	B	Web_Design	0.208	1	1	-0.208	0	1	?
7	C	Web_Design	0.201	1	1	-0.201	0	1	?
8	A	Web_Design	0.129	1	1	-0.129	0	1	?

Figure 7.5 Association Rule for Web Design Chapter by Mathematics

From 386 students who selected Web Design as a difficult question 60 students obtained A passes for Mathematics at O\L. Therefore the support value is 0.226, confidence level is 1, and lift value is 1. From 386 student who selected Web Design as a difficult question, 94 students obtained B passes for Mathematics at O\L. Hence the support value is 0.208, confidence level is 1, and lift value is 1.

From 386 students who selected Web Design as a difficult question, 114 students obtained C passes for Mathematics at O\L. Therefore the support value is 0.201, confidence level is 1, and lift value is 1.

From 386 students who selected Web Design as a difficult question, 117 students obtained S passes for Mathematics at O\L. Hence the support value is 0.129, confidence level is 1, and lift value is 1.

After ordering above support values in descending order, more students who obtained S or C passes for mathematics at ordinary level took Web Design as a difficult chapter.

In accordance with the above association rules outputs obtaining A or B grade for mathematics is more advisable to follow Advanced Level ICT subject according to the support values of above results.

7.2.1.3 By DFD Design Chapter

500 students have considered as the sample and 389 students had selected the DFD chapter as difficult. Figure 7.5 shows the set of association rules generated under above conditions using the 2016 G.C.E. Advanced level ICT results of students.

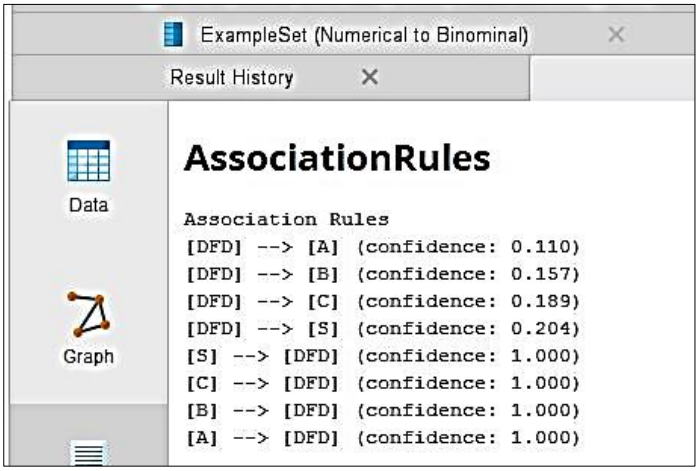
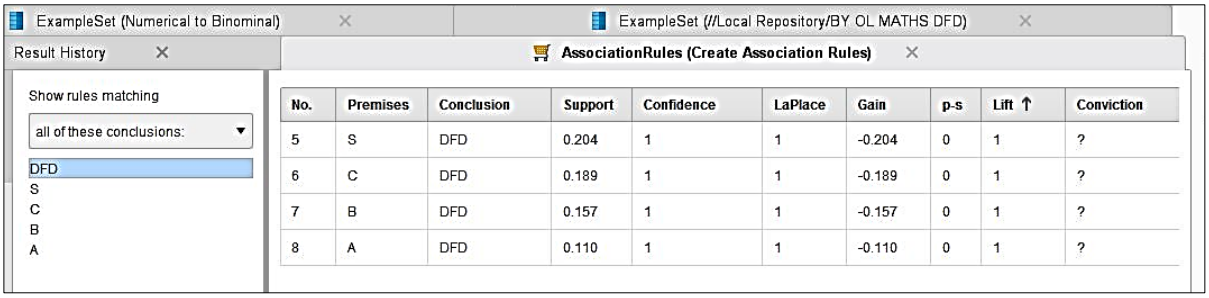


Figure 7.6 Association Rule Result DFD and Mathematics



No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Conviction
5	S	DFD	0.204	1	1	-0.204	0	1	?
6	C	DFD	0.189	1	1	-0.189	0	1	?
7	B	DFD	0.157	1	1	-0.157	0	1	?
8	A	DFD	0.110	1	1	-0.110	0	1	?

Figure 7.7 Association Rule Result for DFD Chapter

From 389 students who selected DFD as a difficult question 51 students obtained A passes for Mathematics at O\L. Therefore the support value is 0.110, confidence level is 1, and lift value is 1. From 271 students who selected DFD as a difficult question 101 students obtained B passes for Mathematics at O\L. Hence the support value is 0.157, confidence level is 1, and lift value is 1.

From 389 student who selected DFD as a difficult question 110 students obtained C passes for Mathematics at O\L. Therefore the support value is 0.189, confidence level is 1, and lift value is 1.

From 389 student who selected DFD as a difficult question 124 students obtained S passes for Mathematics at O\L. Hence the support value is 0.204, confidence level is 1, and lift value is 1.

After ordering above support values in descending order according to the highest, more students who obtained S or C passes for mathematics at ordinary level took DFD as a difficult chapter.

In accordance with the above association rules outputs obtaining A or B grade for mathematics is more advisable to follow Advanced Level ICT subject according to the support values of above results.

7.2.2 Evaluation on whether followed ICT for Ordinary Level or Not followed.

Association Rules were generated to evaluate whether there is an impact on G.C.E. Advanced Level ICT Results based on the fact that O\L ICT was followed or not. Each association rule has a separate confidence value, lift value and number of occurrences.

7.2.2.1 BY Python Chapter

500 students have considered as the sample and 443 students had selected the Python chapter as difficult Figure 7.8 shows the set of association rules generated under above conditions using the 2016 G.C.E. Advanced level ICT results of the students.

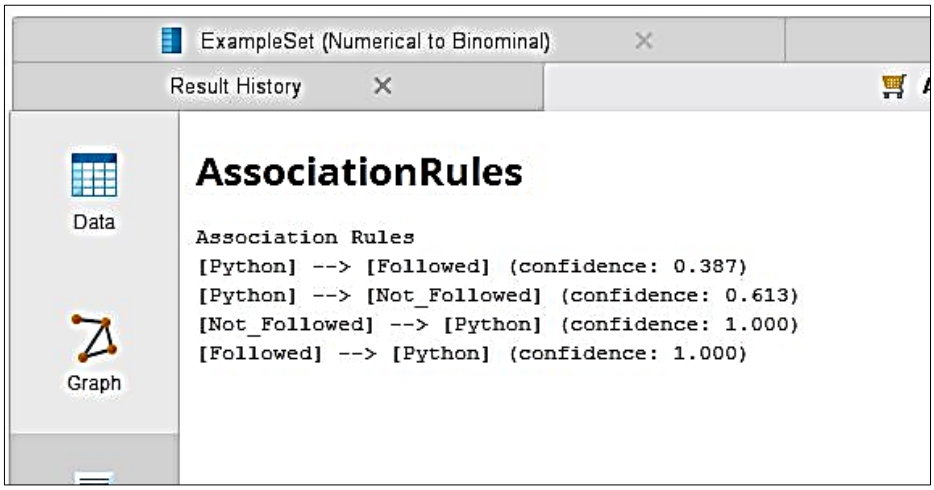


Figure 7.8 Association Rule Result IT Subject

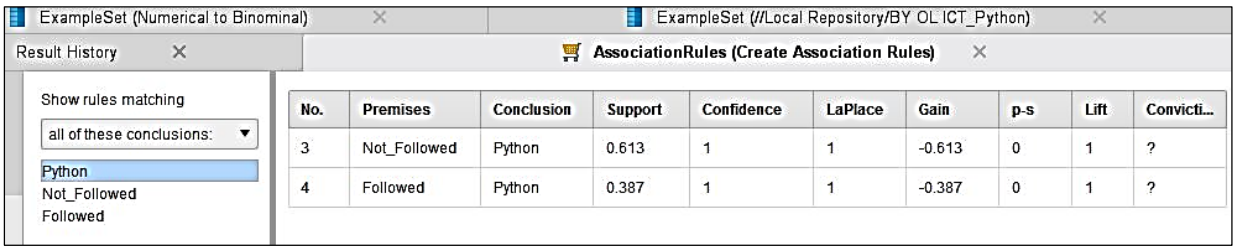


Figure 7.9 Association Rule Python IT subject

From 443 students who selected Python as a difficult question, 173 students followed ICT for O\L. Therefore the support value is 0.387, confidence level is 1, and lift value is 1. And 270 student from the sample who Not followed ICT at O\L selected Python as difficult question. Hence the support value 0.613, confidence level 1, and lift value is 1.

As per the above association rules only 38% (Support value of 0.387) students selected python as a difficult chapter after they following ICT for Ordinary Level and 61% (Support value of 0.613) student who not followed ICT for Ordinary level select Python as a difficult chapter. Therefore following ICT for Ordinary level examination is provide certain support to face the advanced level ICT paper in a successful way.

7.2.2.2 By Web Design Chapter

500 students have considered as the sample and 386 students had selected the Web Design chapter as difficult Figure 7.9 shows the set of association rules generated under above conditions using the 2016 G.C.E. Advanced level ICT results of the students.

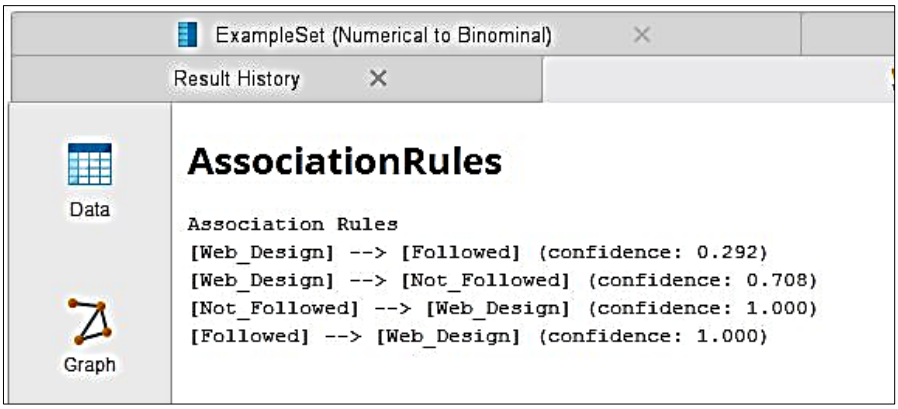
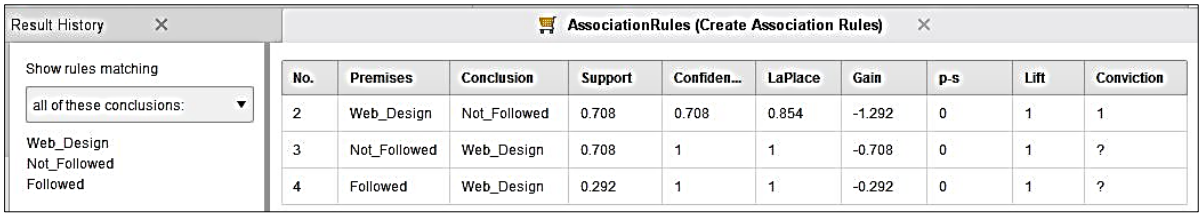


Figure 7.10 Association Rule result Web Design IT Subject



No.	Premises	Conclusion	Support	Confiden...	LaPlace	Gain	p-s	Lift	Conviction
2	Web_Design	Not_Followed	0.708	0.708	0.854	-1.292	0	1	1
3	Not_Followed	Web_Design	0.708	1	1	-0.708	0	1	?
4	Followed	Web_Design	0.292	1	1	-0.292	0	1	?

Figure 7.11 Association Rule Web Design for IT subject

From 386 student who selected Web Design as difficult question 87 students followed ICT for O\L. Therefore the support value 0.292, confidence level 1, and lift value 1. And 298 student from the sample who Not followed ICT at O\L selected Web Design as difficult question. Hence the support value is 0.708, confidence level is 1, and lift value is 1.

As per the above association rules only 23% (Support value of 0.292) students selected Web Design as difficult chapter after they following ICT for Ordinary Level and, 77 % (Support value of 0.708) student who did not followed ICT for Ordinary level select Web Design as a difficult chapter. Therefore following ICT for Ordinary level examination is provide absolute support to face the advanced level ICT paper in an effective way.

7.2.2.3 By DFD Design Chapter

500 students have considered as the sample and 389 students have selected the DFD chapter as difficult Figure 7.11 shows the set of association rules generated under above conditions using the 2016 G.C.E. Advanced level ICT results of students.

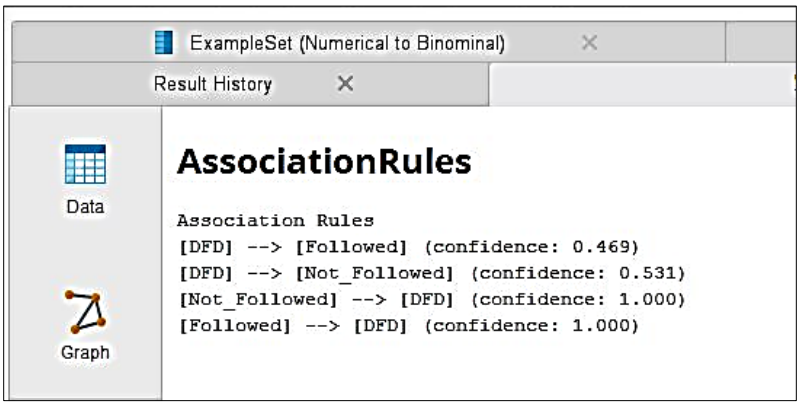
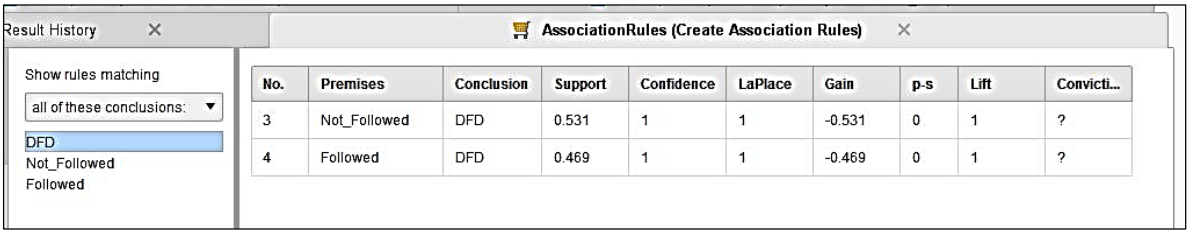


Figure 7.13 Association Rule Result DFD IT subject



No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convicti...
3	Not_Followed	DFD	0.531	1	1	-0.531	0	1	?
4	Followed	DFD	0.469	1	1	-0.469	0	1	?

Figure 7.14 Association Rule DFD question by IT subject

From 389 students who selected DFD as a difficult question 187 students followed ICT for O\L. Therefore the support value is 0.469, confidence level is 1, and lift value is 1. And 200 students from the sample who did Not followed ICT at O\L selected DFD as a difficult question. Hence the support value is 0.531, confidence level is 1, and lift value is 1.

As per the above association rules only 48% (Support value of 0.292) students selected DFD as a difficult question after following ICT for Ordinary Level and 51 % (Support value of 0.708) student who if not follow ICT for Ordinary level selected DFD as a difficult question. Therefore following ICT for Ordinary level examination provides support to some extend to face the advanced level ICT paper in an effective way.

This depicts that approximately a half of those who followed ICT for the G.C.E. Ordinary level examination, had selected the DFD question as difficult.

Thus by considering all the above facts, it can be concluded as even following ICT for G.C.E. Ordinary level examination does not guarantee 100% success in G.C.E . Advanced level results.

This reveals the importance of further study along these lines. Hence this study focused on identifying the factors which have an impact on the G.C.E. Advanced level result of a student.

7.2.3 Evaluation on Gender.

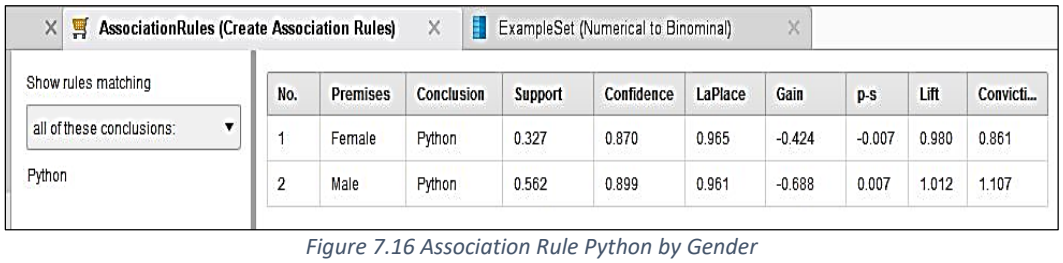
500 students have considered as the sample and 315 (63%) of the students were male and 185 (37%) of students were female. Association Rules were generated to evaluate whether there is an impact on G.C.E. Advanced Level ICT Results based on Gender differences. Each association rule has a separate confidence value, lift value and number of occurrences.

7.2.3.1 By Python Question

Out of the 500 Students of the sample 443 of students had selected the Python question as difficult.. Out of 163 female students (87%) of students and from male students 282 (89%) of male students have selected python as a difficult question.

Figure 7.13 shows the set of association rules generated under above conditions using the 2016 G.C.E. Advanced level ICT results of the students.

Figure 7.15 Association Rule Result Python by Gender



No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convicti...
1	Female	Python	0.327	0.870	0.965	-0.424	-0.007	0.980	0.861
2	Male	Python	0.562	0.899	0.961	-0.688	0.007	1.012	1.107

Figure 7.16 Association Rule Python by Gender

Therefore for Male students' support value is 0.562, while the confidence level is 0.899, and lift value is 1.012. And female students' support value is 0.327, while the confidence level 0.870, and lift value is 0.980. Hence according to the above association rules, female students show a significant support value than male student in python question. But in general majority of both male and female students had selected python as a difficult question. (Out of 350 students 310 of students (88%) are female 87% and male 89%). In this way only 12% of candidates faced the python question with confidence.

This reveals the importance of further study along these lines. Hence this study focused on identifying the factors which have an impact on the G.C.E. Advanced level ICT subject results and performances of the student.

7.2.3.2 By Web Design question

From 500 Students of the sample 386 students had selected the Web Design question as difficult. Out of 185 female students 140 (75%) of students and from 315 male students 246 (78%) of male students have selected Web Design as a difficult question.

Figure 7.15 shows the set of association rules generated under above conditions using the 2016 G.C.E. Advanced level ICT results of students.

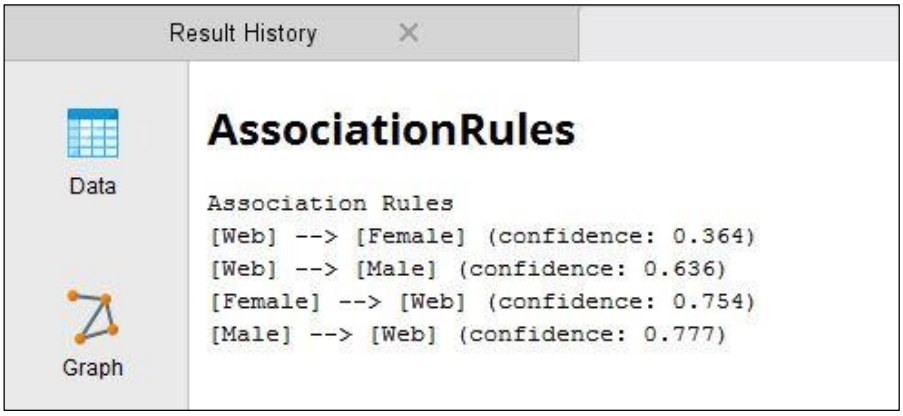


Figure 7.17 Association Rule Result Web Design by Gender

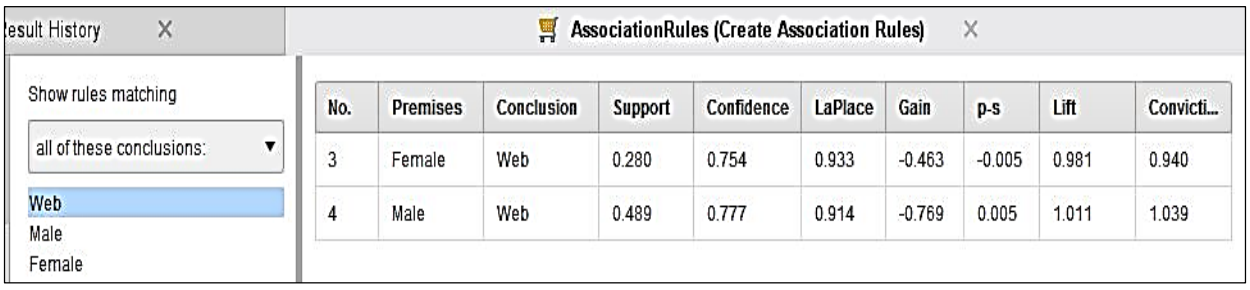


Figure 7.18 Association Rule Web Design by Gender

Therefore for Male students support value is 0.489, while the confidence level is 0.777, and lift value 1.011. And female students' support value is 0.280, while the confidence level is 0.754, and lift value 0.981. Accordingly the above association rules of female student show a significant support value than male student in Web design question. But for the most part of both male and female students have selected web design as a difficult question. (Out of 500 students 386 of students (77%) female 75% and male 78%). In this way only 23% of candidates face the web design question in a successful way.

This reveals the importance of further study along these lines.

7.2.3.3 By DFD Question

From 500 Students of the sample 389 of students had selected the DFD question as difficult. Out of 185 female students 110 (60%) of students and from 315 male students 277 (88%) of male students had selected DFD as a difficult question.

Figure 7.17 shows the set of association rules generated under above conditions using the 2016 G.C.E. Advanced level ICT results of the students.

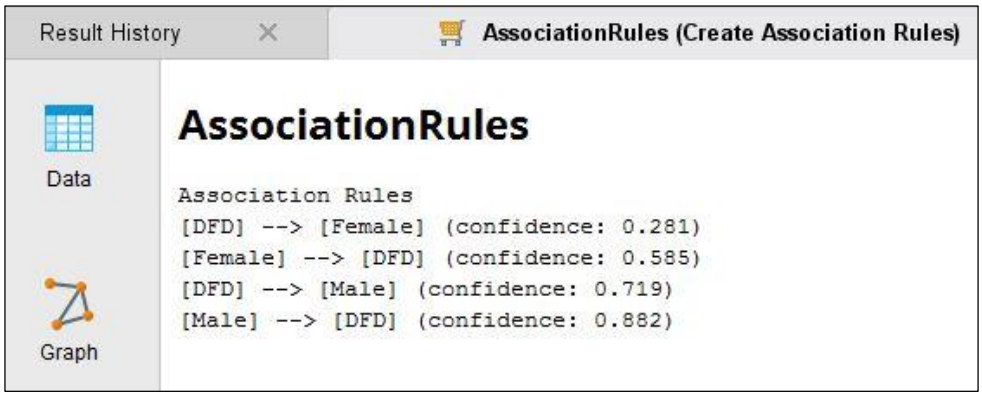
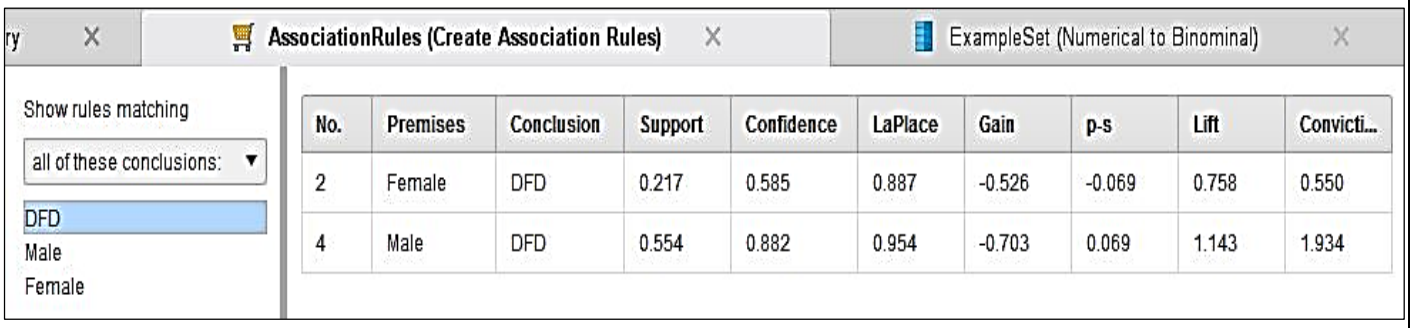


Figure 7.19 Association rule Result DFD by gender



No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convicti...
2	Female	DFD	0.217	0.585	0.887	-0.526	-0.069	0.758	0.550
4	Male	DFD	0.554	0.882	0.954	-0.703	0.069	1.143	1.934

Figure 7.20 Association Rule DFD by Gender

Therefore for Male students' support value is 0.554, while confidence level is 0.882, and lift value is 1.143. And female students' support value is 0.217, confidence level 0.585, and lift value 0.758. Hence as per the above association rules, female students show a significant support value than male students in DFD question as same as the previous two questions. But most of the male and female students have selected web design as a difficult question. (Out of 500 students 389 of students (78%) as female 60% and male 88%). In this way only 22% of candidates faced the DFD question in a successful way.

This reveals the importance of further study along these lines.

7.3 Analysis by the Stream sat for the G.C.E Advanced Level Examination

From the present criteria it was considered whether there is an impact on the Subject stream sat for the G.C.E Advanced Level Examination. 9332 students have considered as the sample and among those, 1038 students were from commerce stream, 1976 students were from Arts stream, 1344 from other streams, 4233 from Engineering Technology stream and 742 from Bio System Stream. Association Rules were generated to evaluate whether stream sat for the G.C.E. Advanced Level Examination has a specific relationship with the G.C.E. Advanced Level ICT Results. Each association rule had a separate confidence value, lift value and number of occurrences.

7.3.1 Commerce stream

1038 students who sat for the Advanced level examination in commerce have been considered as the sample. Among them 8 students (0.72%) obtained A grades, 34 students (3.23%) obtained B grades, 286 students (27.52%) obtained C grades, 411 students (39.61%) obtained S grades and 300 students (28.92%) obtained F grades. Figure 7.19 shows the set of association rules generated under the above conditions using the 2016 G.C.E. Advanced level ICT results of the students.

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convicti...
27	A	COMMERCE	0.001	0.174	0.996	-0.009	0.000	1.564	1.076
20	B	COMMERCE	0.004	0.119	0.975	-0.056	0.000	1.071	1.009
21	C	COMMERCE	0.031	0.120	0.821	-0.480	0.002	1.079	1.010
17	F	COMMERCE	0.032	0.099	0.779	-0.617	-0.004	0.891	0.987
19	S	COMMERCE	0.044	0.114	0.754	-0.727	0.001	1.027	1.003

Figure 7.21 Association Rule By commerce

According to the above results, out of the 9332 students who sat for the Advanced Level ICT subject 8 students (0.08%) obtained 'A' grades for ICT from commerce stream. And from 1038 students who sat for the ICT paper from commerce stream 8 students (0.72%) obtained 'A' grades for ICT. Therefore the support value is 0.001, confidence level is 0.174, and lift value is 1.564.

From 9332 students who sat for the Advanced Level ICT subject 34 students obtained B grades for the ICT from commerce stream. And from 1038 students who sat for the ICT paper from commerce stream 34 students obtained B grades for ICT. Therefore the support value is 0.004, confidence level is 0.119, and lift value 1.701.

From 9332 students who sat for the Advanced Level ICT subject 286 students obtained C grades for the ICT from commerce stream. And from 1038 students who sat for the ICT paper from commerce stream 286 students obtained C grades for ICT. Therefore the support value is 0.031, confidence level is 0.120, and lift value is 1.079.

From 9332 students who sat for the Advanced Level ICT subject 411 students obtained S grades for the ICT from commerce stream. And from 1038 students who sat for the ICT paper from commerce stream 411 students obtained S grades for ICT. Therefore the support value is 0.044, confidence level is 0.114, and lift value is 1.027.

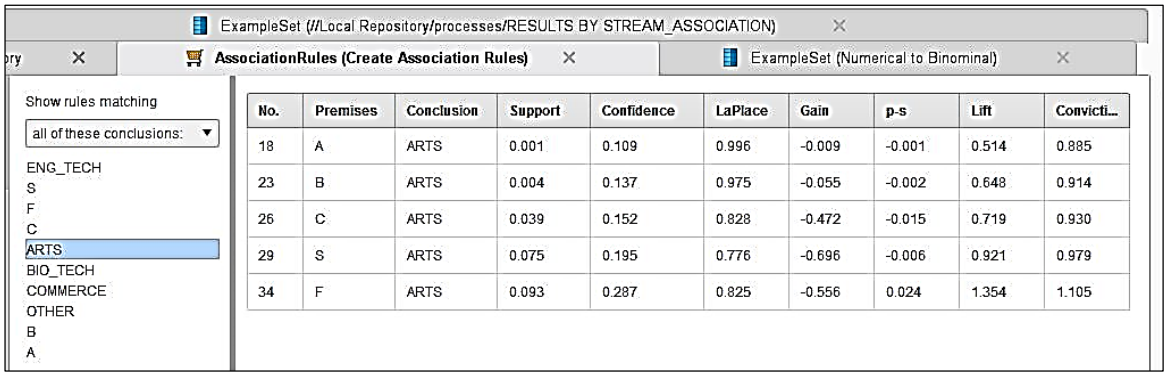
From 9332 students who sat for the Advanced Level ICT subject 300 students obtained F grades for the ICT from commerce stream. And from 1038 students who sat for the ICT paper from commerce stream 300 students obtained F grades for ICT. Therefore the support value is 0.032, confidence level is 0.099, and lift value is 0.891.

After ordering above support values of each grades it was evident that, in commerce stream, most of the students obtained S grades for ICT. And least number of students obtained A grades for ICT.

7.3.2 Arts stream

1976 students who sat for the Advanced level examination in Arts stream have been considered as the sample. Among them 6 students have obtained A grades, 38 students obtained B grades, 363 students obtained C grades, 701 students obtained S grades and 869

students obtained F grades. Figure 7.20 shows the set of association rules generated under above conditions using the 2016 G.C.E. Advanced level ICT results of the students.



No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convicti...
18	A	ARTS	0.001	0.109	0.996	-0.009	-0.001	0.514	0.885
23	B	ARTS	0.004	0.137	0.975	-0.055	-0.002	0.648	0.914
26	C	ARTS	0.039	0.152	0.828	-0.472	-0.015	0.719	0.930
29	S	ARTS	0.075	0.195	0.776	-0.696	-0.006	0.921	0.979
34	F	ARTS	0.093	0.287	0.825	-0.556	0.024	1.354	1.105

Figure 7.22 Association rule by Arts Stream

According to the above results From 9332 students who sat for the Advanced Level ICT subject, 6 students obtained A grades for the ICT from Arts stream. And from 1976 students who sat for the ICT paper from Arts stream 6 students obtained A grades for ICT. Therefore the support value is 0.001, confidence level is 0.109, and lift value is 0.514.

From 9332 students who sat for the Advanced Level ICT subject 38 students obtained B grades for the ICT from Arts stream. And from 1976 students who sat for the ICT paper from Arts stream 38 students obtained B grades for ICT. Therefore the support value is 0.004, confidence level is 0.137, and lift value is 0.648.

From 9332 students who sat for the Advanced Level ICT subject, 363 students obtained C grades for the ICT from Arts stream. And from 1976 students who sat for the ICT paper from Arts stream 363 students obtained C grades for ICT. Therefore the support value is 0.039, confidence level is 0.152, and lift value is 0.719

From 9332 students who sat for the Advanced Level ICT subject, 701 students obtained S grades for the ICT from Arts stream. And from 1976 students who sat for the ICT paper from Arts stream 701 students obtained S grades for ICT. Therefore the support value is 0.075, confidence level is 0.195, and lift value is 0.921.

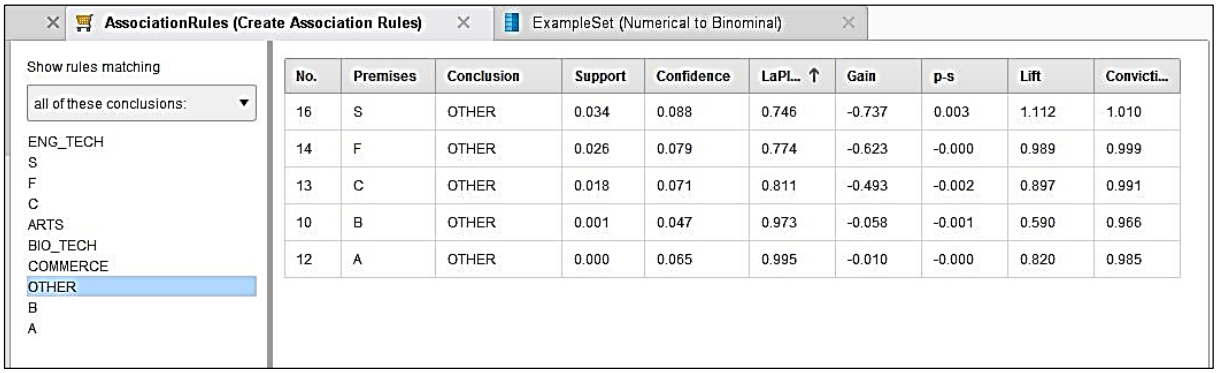
From 9332 students who sat for the Advanced Level ICT subject, 869 students obtained F grades for the ICT from Arts stream. And from 1976 students who sat for the ICT paper

from Arts stream 869 students obtained F grades for ICT. Therefore the support value is 0.093, confidence level is 0.287, and lift value is 0.024.

After ordering above support values of each grades in descending order, in Arts stream, most of the students obtained F grades for ICT. And the least number of students obtained A grades for ICT.

7.3.3 Other stream

1976 students who sat for the Advanced level examination in other streams have been considered as the sample. Among them 23 students have obtained A grades, 111 students obtained B grades, 584 students obtained C grades, 440 students obtained S grades and 188 students obtained F grades. Figure 7.21 shows the set of association rules generated under above conditions using the 2016 G.C.E. Advanced level ICT results of the students.



No.	Premises	Conclusion	Support	Confidence	LaPl...	Gain	p-s	Lift	Convicti...
16	S	OTHER	0.034	0.088	0.746	-0.737	0.003	1.112	1.010
14	F	OTHER	0.026	0.079	0.774	-0.623	-0.000	0.989	0.999
13	C	OTHER	0.018	0.071	0.811	-0.493	-0.002	0.897	0.991
10	B	OTHER	0.001	0.047	0.973	-0.058	-0.001	0.590	0.966
12	A	OTHER	0.000	0.065	0.995	-0.010	-0.000	0.820	0.985

Figure 7.23 Association Rule By other stream

According to the above results From 9332 students who sat for the Advanced Level ICT subject, 23 students obtained A grades for the ICT from Other stream. And from 1344 students who sat for the ICT paper from Other stream 23 students obtained A grades for ICT. Therefore the support value is 0.000, confidence level is 0.065, and lift value is 0.820.

From 9332 students who sat for the Advanced Level ICT subject, 111 students obtained B grades for the ICT from Other stream. And from 1344 students who sat for the ICT paper from Other stream 111 students obtained B grades for ICT. Therefore the support value is 0.001, confidence level is 0.047, and lift value is 0.590.

From 9332 students who sat for the Advanced Level ICT subject, 584 students obtained C grades for the ICT from Other stream. And from 1344 students who sat for the ICT paper

from Other stream 584 students obtained C grades for ICT. Therefore the support value is 0.018, confidence level is 0.071, and lift value is 0.897.

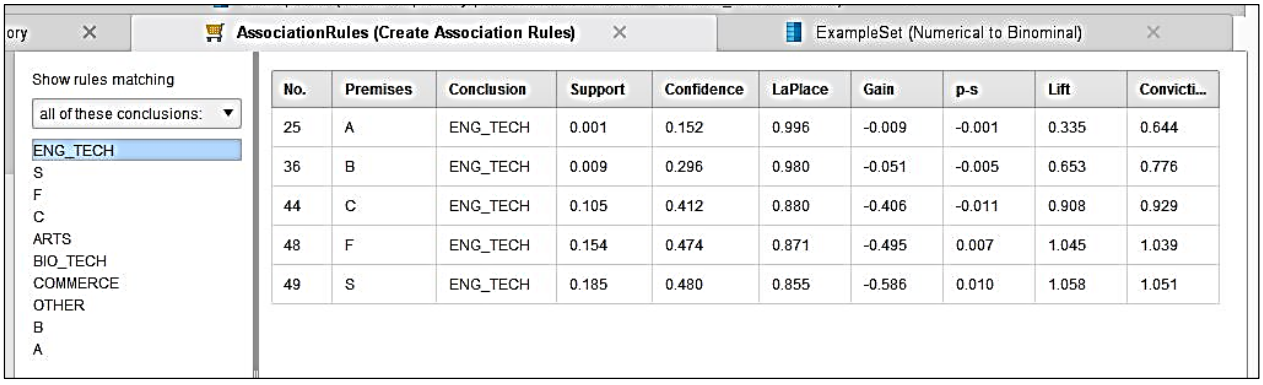
From 9332 students who sat for the Advanced Level ICT subject, 440 students obtained S grades for the ICT from Arts stream. And from 1344 students who sat for the ICT paper from Other stream 440 students obtained S grades for ICT. Therefore the support value is 0.034, confidence level is 0.088, and lift value is 1.112.

From 9332 students who sat for the Advanced Level ICT subject, 188 students obtained F grades for the ICT from Other stream. And from 1344 students who sat for the ICT paper from Other stream 188 students obtained F grades for ICT. Therefore the support value is 0.026, confidence level is 0.079, and lift value is 0.989.

After ordering the above support values of each grades according to the descending order, in Other streams, most of the students obtained S grades for ICT. And the least number of students obtained A grades for ICT.

7.3.4 Engineering Technology stream

4233 students who sat for the Advanced level examination in Engineering Technology stream have been considered as the sample. Among them 7 students have obtained A grades, 83 students obtained B grades, 982 students obtained C grades, 1727 students obtained S grades and 1436 students obtained F grades. Figure 7.22 shows the set of association rules generated under above conditions using the 2016 G.C.E. Advanced level ICT results of the students.



No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convicti...
25	A	ENG_TECH	0.001	0.152	0.996	-0.009	-0.001	0.335	0.644
36	B	ENG_TECH	0.009	0.296	0.980	-0.051	-0.005	0.653	0.776
44	C	ENG_TECH	0.105	0.412	0.880	-0.406	-0.011	0.908	0.929
48	F	ENG_TECH	0.154	0.474	0.871	-0.495	0.007	1.045	1.039
49	S	ENG_TECH	0.185	0.480	0.855	-0.586	0.010	1.058	1.051

Figure 7.24 Association Rule by Eng_Tech stream

According to the above results from 9332 students who sat for the Advanced Level ICT subject, 7 students obtained A grades for the ICT from Engineering Technology stream. And from 4233 students who sat for the ICT paper from Engineering Technology stream 7 students obtained A grades for ICT. Therefore the support value is 0.001, confidence level is 0.152, and lift value is 0.335.

From 9332 students who sat for the Advanced Level ICT subject, 83 students obtained B grades for the ICT from Engineering Technology stream. And from 4233 students who sat for the ICT paper from Engineering Technology stream 83 students obtained B grades for ICT. Therefore the support value is 0.009, confidence level is 0.296, and lift value is 0.653.

From 9332 students who sat for the Advanced Level ICT subject, 982 students obtained C grades for the ICT from Engineering Technology stream. And from 4233 students who sat for the ICT paper from Engineering Technology stream 982 students obtained C grades for ICT. Therefore the support value is 0.105, confidence level is 0.412, and lift value is 0.908.

From 9332 students who sat for the Advanced Level ICT subject, 1727 students obtained S grades for the ICT from Engineering Technology stream. And from 4233 students who sat for the ICT paper from Engineering Technology stream 1727 students obtained S grades for ICT. Therefore the support value is 0.185, confidence level is 0.480, and lift value is 1.058.

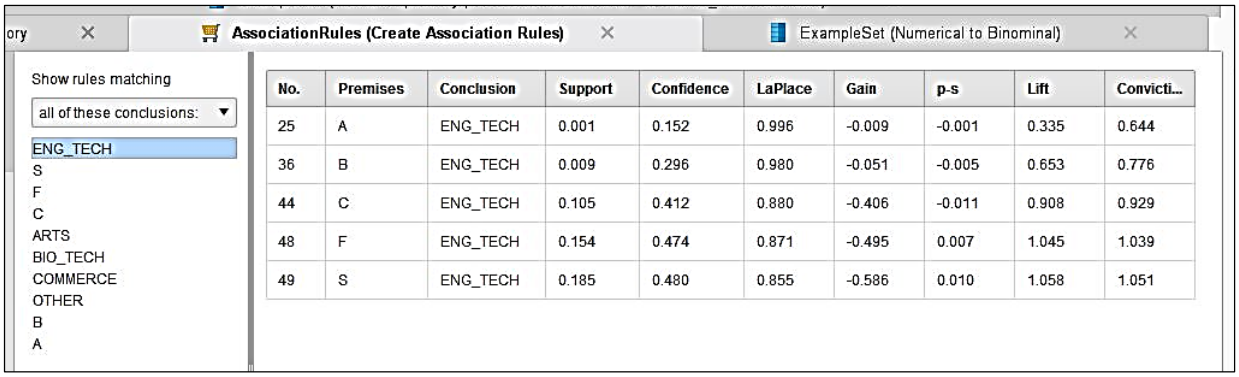
From 9332 students who sat for the Advanced Level ICT subject, 1436 students obtained F grades for the ICT from Engineering Technology stream. And from 4233 students who sat for the ICT paper from Engineering Technology stream 1436 students obtained F grades for ICT. Therefore the support value is 0.154, confidence level is 0.474, and lift value is 1.045.

After ordering the above support values of each grades in descending order, in Engineering Technology stream, most of the students obtained S grades for ICT. And the least number of students obtained ‘A’ grades for ICT.

7.3.5 Bio-System Technology stream

742 students who sat for the Advanced level examination in Bio-System Technology stream have been considered as the sample. Among them 3 students have obtained A

grades, 14 students obtained B grades, 170 students obtained C grades, 318 students obtained S grades and 239 students obtained F grades. Figure 7.23 shows the set of association rules generated under above conditions using the 2016 G.C.E. Advanced level ICT results of the students.



No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convicti...
25	A	ENG_TECH	0.001	0.152	0.996	-0.009	-0.001	0.335	0.644
36	B	ENG_TECH	0.009	0.296	0.980	-0.051	-0.005	0.653	0.776
44	C	ENG_TECH	0.105	0.412	0.880	-0.406	-0.011	0.908	0.929
48	F	ENG_TECH	0.154	0.474	0.871	-0.495	0.007	1.045	1.039
49	S	ENG_TECH	0.185	0.480	0.855	-0.586	0.010	1.058	1.051

Figure 7.25 Association Rule by Grade

According to the above results from 9332 students who sat for the Advanced Level ICT subject, 3 students obtained A grades for the ICT from Bio-System Technology stream. And from 742 students who sat for the ICT paper from Bio-System Technology stream 3 students obtained A grades for ICT. Therefore the support value is 0.002, confidence level is 0.500, and lift value is 3.472.

From 9332 students who sat for the Advanced Level ICT subject, 14 students obtained B grades for the ICT from Bio-System Technology stream. And from 742 students who sat for the ICT paper from Bio-System Technology stream 14 students obtained B grades for ICT. Therefore the support value is 0.012, confidence level is 0.401, and lift value is 2.782.

From 9332 students who sat for the Advanced Level ICT subject, 170 students obtained C grades for the ICT from Bio-System Technology stream. And from 742 students who sat for the ICT paper from Bio-System Technology stream 170 students obtained C grades for ICT. Therefore the support value is 0.062, confidence level is 0.245, and lift value is 1.698.

From 9332 students who sat for the Advanced Level ICT subject, 318 students obtained S grades for the ICT from Bio-System Technology stream. And from 742 students who sat for the ICT paper from Bio-System Technology stream 318 students obtained S grades for ICT. Therefore the support value is 0.047, confidence level is 0.122, and lift value is 0.849.

From 9332 students who sat for the Advanced Level ICT subject, 239 students obtained F grades for the ICT from Bio-System Technology stream. And from 742 students who sat for the ICT paper from Bio-System Technology stream 239 students obtained F grades for ICT. Therefore the support value is 0.020, confidence level is 0.062, and lift value is 0.429.

After ordering the above support values of each grades from to the highest value, in Bio-System Technology stream most of the students obtained S grades for ICT. And the least number of students obtained 'A' grades for ICT.

Following Figure 7.25 shows the set of association rules generated under the above conditions using the 2016 G.C.E. Advanced level ICT results of the students.

Association Rules Figure 7.25

[ENG_TECH] --> [A] (confidence: 0.002)
[ARTS] --> [A] (confidence: 0.003)
[OTHER] --> [A] (confidence: 0.004)
[COMMERCE] --> [A] (confidence: 0.008)
[BIO_TECH] --> [A] (confidence: 0.017)
[OTHER] --> [B] (confidence: 0.018)
[ARTS] --> [B] (confidence: 0.019)
[ENG_TECH] --> [B] (confidence: 0.019)
[COMMERCE] --> [B] (confidence: 0.032)
[B] --> [OTHER] (confidence: 0.047)
[F] --> [BIO_TECH] (confidence: 0.062)
[A] --> [OTHER] (confidence: 0.065)
[C] --> [OTHER] (confidence: 0.071)
[F] --> [OTHER] (confidence: 0.079)
[BIO_TECH] --> [B] (confidence: 0.083)
[S] --> [OTHER] (confidence: 0.088)
[F] --> [COMMERCE] (confidence: 0.099)
[A] --> [ARTS] (confidence: 0.109)
[S] --> [COMMERCE] (confidence: 0.114)
[B] --> [COMMERCE] (confidence: 0.119)
[C] --> [COMMERCE] (confidence: 0.120)
[S] --> [BIO_TECH] (confidence: 0.122)
[B] --> [ARTS] (confidence: 0.137)
[BIO_TECH] --> [F] (confidence: 0.139)

[A] --> [ENG_TECH] (confidence: 0.152)

[C] --> [ARTS] (confidence: 0.152)

[A] --> [COMMERCE] (confidence: 0.174)

[ARTS] --> [C] (confidence: 0.184)

[S] --> [ARTS] (confidence: 0.195)

[OTHER] --> [C] (confidence: 0.229)

[ENG_TECH] --> [C] (confidence: 0.232)

[C] --> [BIO_TECH] (confidence: 0.245)

[COMMERCE] --> [C] (confidence: 0.276)

[F] --> [ARTS] (confidence: 0.287)

[COMMERCE] --> [F] (confidence: 0.289)

[B] --> [ENG_TECH] (confidence: 0.296)

[OTHER] --> [F] (confidence: 0.321)

[BIO_TECH] --> [S] (confidence: 0.327)

[ENG_TECH] --> [F] (confidence: 0.339)

[ARTS] --> [S] (confidence: 0.355)

[COMMERCE] --> [S] (confidence: 0.396)

[B] --> [BIO_TECH] (confidence: 0.401)

[ENG_TECH] --> [S] (confidence: 0.408)

[C] --> [ENG_TECH] (confidence: 0.412)

[OTHER] --> [S] (confidence: 0.429)

[BIO_TECH] --> [C] (confidence: 0.434)

[ARTS] --> [F] (confidence: 0.439)

[F] --> [ENG_TECH] (confidence: 0.474)

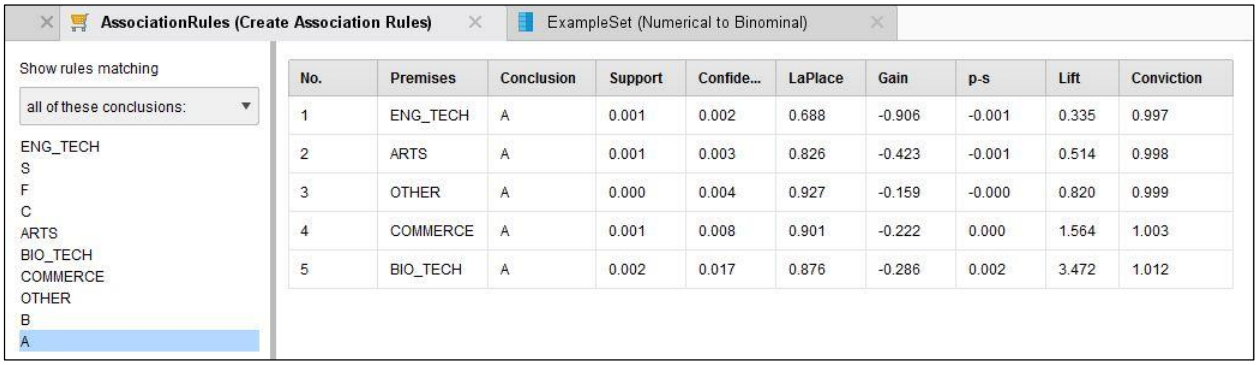
[S] --> [ENG_TECH] (confidence: 0.480)

[A] --> [BIO_TECH] (confidence: 0.500)

Figure 7.26 Association Rule Generated using conditions

7.4 Analysis by the Grades obtained by each stream

7.4.1 Grade A

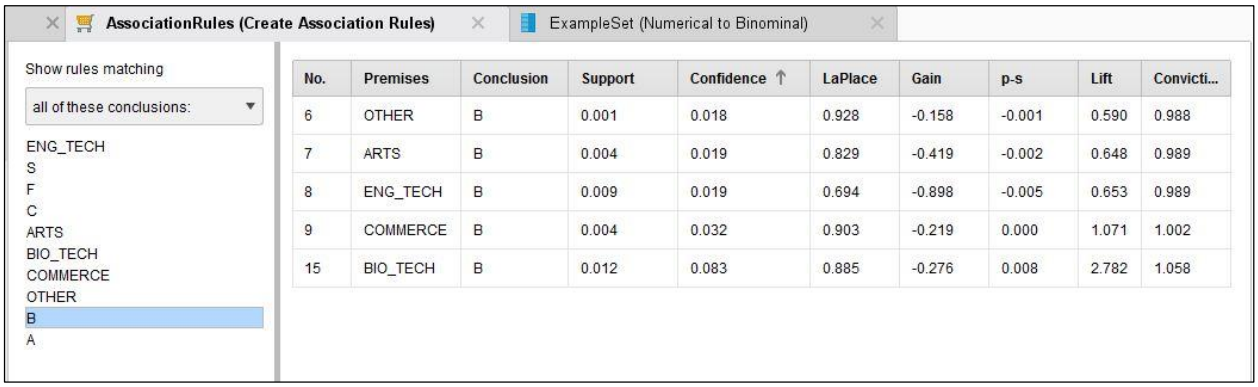


No.	Premises	Conclusion	Support	Confide...	LaPlace	Gain	p-s	Lift	Conviction
1	ENG_TECH	A	0.001	0.002	0.688	-0.906	-0.001	0.335	0.997
2	ARTS	A	0.001	0.003	0.826	-0.423	-0.001	0.514	0.998
3	OTHER	A	0.000	0.004	0.927	-0.159	-0.000	0.820	0.999
4	COMMERCE	A	0.001	0.008	0.901	-0.222	0.000	1.564	1.003
5	BIO_TECH	A	0.002	0.017	0.876	-0.286	0.002	3.472	1.012

Figure 7.27 Association Rule By A grades

As per the above association rules, when considering A grades, Bio-Technology stream has the highest confidence values 0.017 and the commerce stream has the next highest confidence value of 0.08. Therefore the following students who sat for the exam in Bio-Technology and commerce streams had scored A grades for G.C.E Advanced Level ICT than the students of other three streams.

7.4.2 Grade B



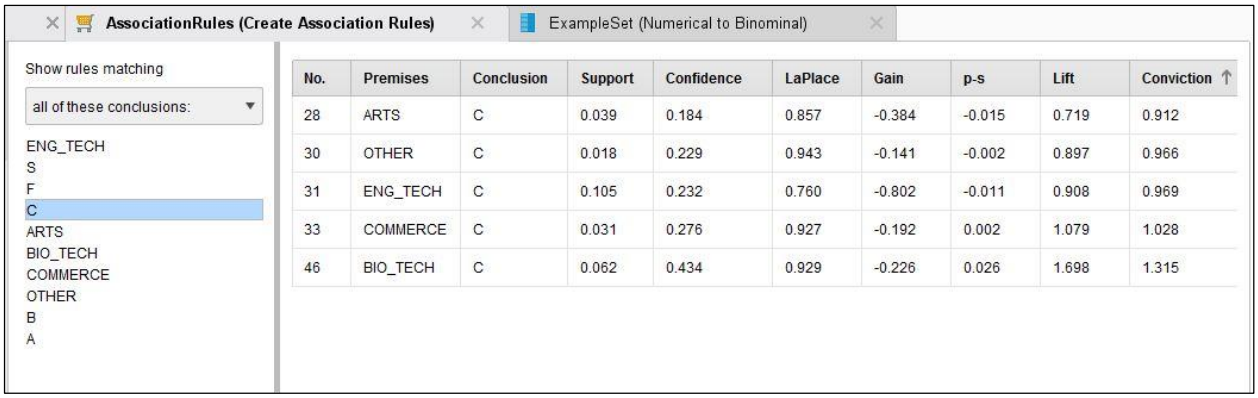
No.	Premises	Conclusion	Support	Confidence ↑	LaPlace	Gain	p-s	Lift	Convicti...
6	OTHER	B	0.001	0.018	0.928	-0.158	-0.001	0.590	0.988
7	ARTS	B	0.004	0.019	0.829	-0.419	-0.002	0.648	0.989
8	ENG_TECH	B	0.009	0.019	0.694	-0.898	-0.005	0.653	0.989
9	COMMERCE	B	0.004	0.032	0.903	-0.219	0.000	1.071	1.002
15	BIO_TECH	B	0.012	0.083	0.885	-0.276	0.008	2.782	1.058

Figure 7.28 Association rule by B gardes

As per the above association rules, when considering B grades, Bio-Technology stream has the highest confidence values 0.083 and the commerce stream has the next highest confidence value of 0.032. Therefore the following students who sat for the exam in Bio-

Technology and commerce streams had scored B grades for G.C.E Advanced Level ICT than the students of other three streams.

7.4.3 Grade C

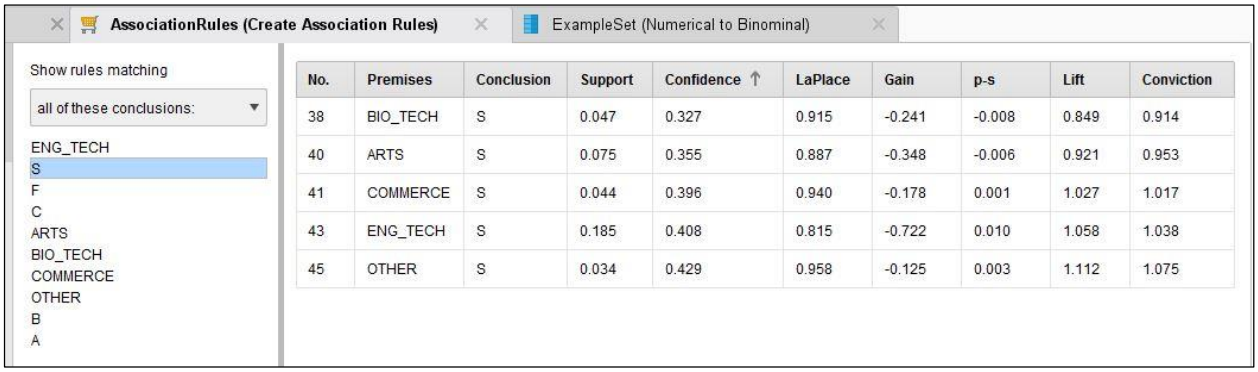


No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Conviction
28	ARTS	C	0.039	0.184	0.857	-0.384	-0.015	0.719	0.912
30	OTHER	C	0.018	0.229	0.943	-0.141	-0.002	0.897	0.966
31	ENG_TECH	C	0.105	0.232	0.760	-0.802	-0.011	0.908	0.969
33	COMMERCE	C	0.031	0.276	0.927	-0.192	0.002	1.079	1.028
46	BIO_TECH	C	0.062	0.434	0.929	-0.226	0.026	1.698	1.315

Figure 7.29 Association Rule by C grades

As per the above the association rules, when considering C grades, Bio-Technology stream has the highest confidence values 0.434 and the commerce stream has the next highest confidence value of 0.276. Therefore the following students who sat for the exam in Bio-Technology and commerce streams had scored C grades for G.C.E Advanced Level ICT than the students of other three streams.

7.4.4 Grade S



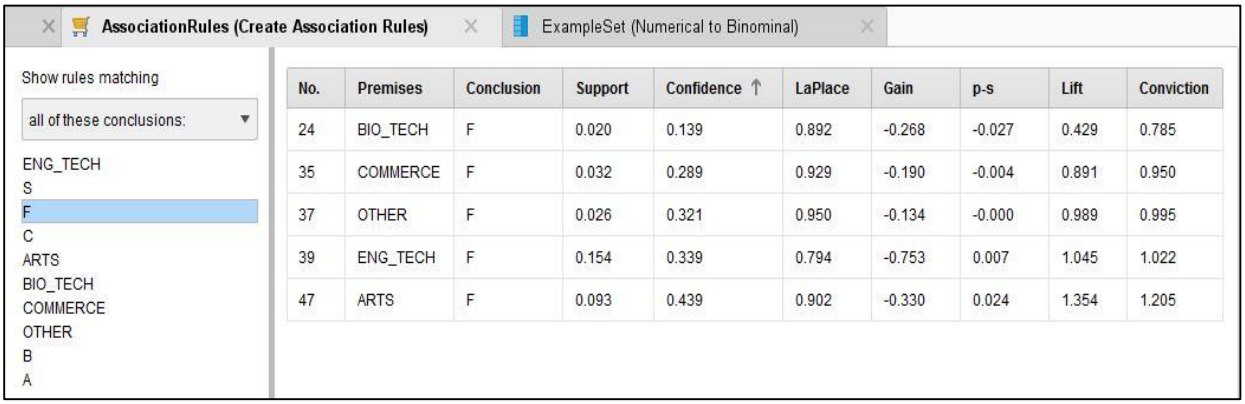
No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Conviction
38	BIO_TECH	S	0.047	0.327	0.915	-0.241	-0.008	0.849	0.914
40	ARTS	S	0.075	0.355	0.887	-0.348	-0.006	0.921	0.953
41	COMMERCE	S	0.044	0.396	0.940	-0.178	0.001	1.027	1.017
43	ENG_TECH	S	0.185	0.408	0.815	-0.722	0.010	1.058	1.038
45	OTHER	S	0.034	0.429	0.958	-0.125	0.003	1.112	1.075

Figure 7.30 Association Rule by S gardes

As per the above association rules, when considering S grades, Other stream has the highest confidence values 0.429 and the Engineering Technology stream has the next highest confidence value of 0.408. Therefore the following students who sat for the exam in Other

and Engineering Technology streams had scored S grades for G.C.E Advanced Level ICT than the students of other three streams.

7.4.5 Grade F



No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Conviction
24	BIO_TECH	F	0.020	0.139	0.892	-0.268	-0.027	0.429	0.785
35	COMMERCE	F	0.032	0.289	0.929	-0.190	-0.004	0.891	0.950
37	OTHER	F	0.026	0.321	0.950	-0.134	-0.000	0.989	0.995
39	ENG_TECH	F	0.154	0.339	0.794	-0.753	0.007	1.045	1.022
47	ARTS	F	0.093	0.439	0.902	-0.330	0.024	1.354	1.205

Figure 7.31 Association Rule by F grades

As per the above association rules, when considering F grades, Arts stream has the highest confidence values of 0.439 and the Engineering Technology stream has the next highest confidence value of 0.339. And the Bio-Technology stream has the lowest confidence value of 0.139. Therefore the students who sat for the exam in Arts and Engineering Technology streams have the highest rate of failure at the Advanced Level ICT subject than the students of other streams and the Bio-Technology stream has the lowest rate of failure when compared with other streams.

Chapter 8

8 Conclusion and Future work

8.1 Introduction

As the final chapter of this thesis, this will identify and interpret the findings that have been received on evaluation chapter. And further discusses some limitations and future improvements which can be proposed for further work.

8.2 Conclusion

This study has been conducted with the intention of identifying problems related to the G.C.E Advanced Level Examination ICT subject and providing suggestions to enhance the performance of the education system. Therefore the study is leads to discover any prevailing relationships between a student's G.C.E. Ordinary level Mathematics results, on the fact that he/she followed or not followed ICT for Ordinary level examination, the Stream sat for the Advanced Level examination and his/her G.C.E. Advanced level result in Information and Communication Technology subject.

Initially, the study focused on identifying the difficulty level of each question on exam paper by using the answers provided on the questionnaire. Based on selected questions the study was initiated. Association Rule technique of data mining was used to identify certain patterns in the examination results.

From the analysis carried out, it was observed that the G.C.E. Ordinary level Mathematics result of a student has a direct impact on his/her G.C.E. Advanced level ICT subject result. Further it was proved that following ICT subject at G.C.E. Ordinary level examination also has a direct impact on his/her G.C.E. Advanced level ICT subject result. In other words, better the performance in these two subjects of mathematics and ICT, higher the probability of performing well at G.C.E. Advanced level examination ICT subject. This was clear from the computed confidence and support values too. Other analysis was to find whether there is an impact on gender difference for his/her G.C.E. Advanced level ICT subject result. It was found that when considering overall factors, female students performed well in the exam than male students.

In final analysis done to find the impact of the subject Stream sat for the examination. It was evident that the students who followed Bio System Technology had obtained more A grades and it had the lowest 'F' failure grades compared with other streams, while and Arts stream has the highest rate of 'F' failure grades among the other streams.

From this study, the fact that G.C.E. Ordinary level Mathematics and ICT were the most important subjects behind the success at the G.C.E. Advanced level ICT was highlighted. Although the analysis revealed the above results, it was observed that some students with A grade for Mathematics and followed ICT at the G.C.E. Ordinary level examination had not performed well at the G.C.E. Advanced level. Moreover it was observed that apart from the selected questions some students had not performed well in some other questions too like Logic Gates and Networking. Though certain male students obtained good results at G.C.E. Advanced level ICT subject it was observed that female students performed better than male students in pointed questions here. This may be due to some other factors which cannot be captured through an analysis which was based on quantitative factors.

There were also a quite a few students who have performed extremely well at the G.C.E. Advanced level examination but those had not satisfied the above criterion. However, from the study, it can be concluded that a student who fulfills the identified requirements has a higher chance of performing well at ICT subject at the G.C.E. Advanced level examination.

8.3 Limitations

A few limitations had to overcome during this study. The major limitation was the lack of knowledge on other factors that influence the performance at the G.C.E. Advanced level examination. In this study the main focus was on identifying any relationship between G.C.E. Ordinary level examination result on Mathematics, based on the fact that followed ICT for G.C.E. Ordinary level examination or Not, Impact on Gender difference and the level of difficulty of each question of ICT paper at G.C.E. Advanced level examination. Further it was analyzed the impact on the stream sat for the G.C.E. Advanced level examination and result of students. However there could be other external factors that affect a student's performance at the G.C.E. Advanced level examination. It should be further pointed out that quality of teaching and the method of learning can also be vary according to the person. Similarly the standard of the Ordinary level and Advanced level

examinations can vary from year to year. Hence one may argue that comparing the results of students who sat the examination in different years is not practical.

8.4 Future Developments

As a future extension of this study a model using same candidate details with results of G.C.E Advanced Level ICT will be developed. And other factors that may have an effect on the G.C.E. Advanced level results of the students could be considered.

8.5 Summary

The chapter contains the review activities that have been carried throughout the period of study. And how it realized the aims and objectives of the study. Limitations that identified during the period of study and the proposals for future enhancements have described as the final chapter.

Chapter 9

9 References

[1] (Samaranayake & Caldera, 2012b)

[2] Castells, M. (1996). The Rise of the Network Society (vol. 2). Oxford: Blackwell Publishers.

[3] Irvin R. KATZ Educational Testing Service Princeton, NJ 08541 USA , Alexius Smith MACKLIN Purdue University West Lafayette, IN 47907 USA

[4] ICT Curriculum in Sri Lankan Schools: A Critical Review Chandima H. de Silva Department of Statistics & Computer Science University of Kelaniya, Sri Lanka

[5]"EducationalDataMining.org". 2013. Retrieved 2013-07-15

[6]Baker, R.S.; Yacef, K (2009). "The state of educational data mining in 2009: A review and future visions". JEDM-Journal of Educational Data Mining. 1 (1): 2017.

[7]Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. Expert Systems with Applications, 33(1), 135–146. http://doi.org/10.1016/j.eswa.2006.04.005

[8] Nandeshwar, A., & Chaudhari, S. (2009). Enrollment prediction models using data mining. Retrieved January, 10, 2010.

[9] G - Prof.-David-Kember-publication-PDF.pdf. (n.d.). Retrieved September 16, 2016, from http://educationconference.co/2016/wp-content/uploads/2015/10/Prof.-David-Kember-publication-PDF.pdf

[10]Sarker, F., Tiropanis, T., & Davis, H. C. (2013). Students' performance prediction by using institutional internal and external open data sources. Retrieved from http://eprints.soton.ac.uk/353532/1/Students'%20mark%20prediction%20model.pdf

[11]Yuan, W., Deng, C., Zhu, H., & Li, J. (2012). The Statistical Analysis and Evaluation of Examination Results of Materials Research Methods Course.

[12] El-Halees, A. (2008) ‘Mining Students Data to Analyze Learning Behavior: A Case Study’, The 2008 international Arab Conference of Information Technology (ACIT2008) – Conference Proceedings, University of Sfax, Tunisia, Dec 15- 18.

[13]Baradwaj, B. and Pal, S. (2011) ‘Mining Educational Data to Analyze Student s’ Performance’, International Journal of Advanced Computer Science and Applications, vol. 2, no. 6, pp. 63-69.

[14]Chandra, E. and Nandhini, K. (2010) ‘Knowledge Mining from Student Data’, European Journal of Scientific Research, vol. 47, no. 1, pp. 156-163.

[15] Arora, Rakesh Kumar, and Dharmendra Badal. "Mining Association Rules to Improve Academic Performance." (2014).

[16] Ayesha, Shaeela, Tasleem Mustafa, Ahsan Raza Sattar, and M. Inayat Khan. "Data mining model for higher education system." Europen Journal of Scientific Research 43, no. 1 (2010): 24-29.

[17] Ramasubramanian, P., K. Iyakutti, and P. Thangavelu. "Enhanced data mining analysis in higher educational system using rough set theory." African Journal of Mathematics and Computer Science Research 2, no. 9 (2009): 184-188.
*
[18] Yadav, Surjeet Kumar, and Saurabh Pal. "Data Mining Application in Enrollment Management: A Case Study." International Journal of Computer Applications (IJCA) 41.5 (2012): 1-6.

[19] Kovacic, Zlatko. "Early prediction of student success: Mining students' enrolment data." (2010).

[20] Wu, X., Zhang, H., & Zhang, H. (2010, October). Study of comprehensive evaluation method of undergraduates based on data mining. In Intelligent Computing and Integrated Systems (ICISS), 2010 International Conference on(pp. 541-543). IEEE.

[21] Srimani, P. K., and Malini M. Patil. "A Classification Model for Edu-Mining." PSRC-ICICS Conference Proceedings. 2012.

[22] Nghe, Nguyen Thai, Paul Janecek, and Peter Haddawy. "A comparative analysis of techniques for predicting academic performance." Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07. 37th Annual. IEEE, 2007.

.

[23]Al-Radaideh, Q., Al-Shawakfa, E. and Al-Najjar, M. (2006) 'Mining Student Data Using Decision Trees', The 2006 International Arab Conference on Information Technology (ACIT'2006) – Conference Proceedings.

[24] Ayesha, S. , Mustafa, T. , Sattar, A. and Khan, I. (2010) 'Data Mining Model for Higher Education System', European Journal of Scientific Research, vol. 43, no. 1, pp. 24-29.

[25] Johina et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015, 2928-2930

[26]Mrs. Bharati M. Ramageri "DATA MINING TECHNIQUES AND APPLICATIONS" Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305

[27]Amandeep Kaur Mann, Navneet Kaur" SURVEY PAPER ON CLUSTERING TECHNIQUES" International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 4, April 2013

10 Appendices