بسم الله الرحمن الرحيم

Sudan University of Science &Technology

College of Graduate Studies

Faculty of computer science and information technology

Discovery patterns Influencing Academic performance of Information Technology students -University of Kassala

إكتشاف الأنما ط التي تؤثر على الأداء الأكاديمي لطلاب تقانة المعلومات- جامعة كسلا

Prepared by:

Madina Imam Idress Hamza

Supervisor:

Dr. Mohamed Adaney

2020

Introductive

قال تعالى: (يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا مِنْكُمْ وَ الَّذِينَ أُوتُوا الْعِلْمَ دَرَجَات)

سورة المجادلة الآية"11"

Acknowledgement

Praise be to Allah, my deep thanks, to my supervisor: Dr. Mohamed Adaney, I would like to express appreciation his guidance for this work.

I would like to express my sincere thanks to all the staff of faculty of Computer Science and Information Technology, University of kassala, for their valuable help, special and deep thanks are extended to Dr. Montaser Fadulalla Adam Allah and Ustaz muataze Dawod Salih, for their kind help, assistance and support.

Finally my deepest thanks are extended to my colleagues of the batch (9) in Master.

Abstract

Data Mining provides powerful techniques for various fields including education. The research in the educational field is rapidly increasing due to the massive amount of students' data which can be used to discover valuable patterns pertaining students' learning behavior. This research aim to providing guidance rules that implemented modern methods to help students enhance their academic performance and discover courses that more related with GPA and affect negatively or positively at GPA. This research applied on student's data from department of information system Faculty of computer science and information Technology University of Kassala (2012 -2018). The study applied two data mining techniques clustering algorithms (k-means) and association rules algorithms this techniques implemented in orange data mining tool to generating strong rules in each study year .The clustering algorithms results were evaluated regarding to high accuracy for each cluster and then applied Association rules algorithms for each cluster. The obtained results are strong rules that use to improve SAP (Student Academic Performance) and appear which courses are effect positively or negatively on accumulative GPA.

المستخلص

يوفرتنقيب البيانات تقنيات مفيده لمختلف المجالات بما في ذلك التعليم. يتزايد البحث في المجال التعليمي بسرعة بسبب الكم الهائل من بيانات الطلاب التي يمكن استخدامها لاكتشاف الانماط المتعلقة بسلوك تعلم الطلاب. هذاالبحث يهدف الى توفير قواعد ارشادية لمساعدة الطلاب على تحسين ادائهم الأكاديمي واكتشاف الكورسات الأكثر ارتباطاً ب المعدل وتؤثر سلبا او ايجاباً على المعدل التراكمي . تم تطبيق هذا البحث على بيانات الطلاب من قسم نظم المعلومات بكلية علوم الحاسوب وتقانة المعلومات بجامعة كسلا (2012 -2018) . طبقت الدراسة طريقتين من طرق التنقيب عن البيانات خوارزميات التجميع وخوارزميات قواعد الارتباط تم تنفيذهما على اداة التنقيب عن البيانات (اورانج) للحصول على قواعد قويه لكل سنة دراسية . تم تقييم نتائج خوارميات التجميع باعتبار الدقه العاليه لكل مجموعة ثم طبقت خوارزميات الارتباط لكل مجموعة . قدمت النتائج التي تم الحصول عليها قواعد قوية تستخدم لتحسين الاداء الأكاديمي وتظهر الكورسات التي لها تأثير إيجابي أو سلبي على المعدل التراكمي .

Table of Contents

Table of terms 4
CHAPTER ONE 5
1. Introduction 5
1.1 Research background 5
1.2 Problem statement 6
1.3 Objectives 6
1.4 Methodology 6
1.5 Scope of the research 7
1.6 contribution 8
1.7 Thesis organization 8
CHAPTER TWO 9
2.1 Introduction 9
2.2 DATA MINING 9
2.2.1 Data mining definition: 9
2.2.2 Data mining steps: 10
2.2.3 Clustering 11
2.2.4 Association Rules Mining 13
2.2.5 Data mining tool 13
2.3 Educational institutions management process 14
2.4 Education Data Mining (EDM) 14
2.5 Student academic performance (SAP) 19
2.6 There are many methods are used in the process of educational data mining in according to the application and the objectives for which used 19
2.6.1 Prediction SAP 19
2.6.2 Identify appropriate mechanisms to deal with similar student groups in learning style and social communication method (clustering) 21
2.5.3 Discover the weaknesses of the learners to improve them, Studying causal relationships in the educational process and discovering patterns of weakness to improve them (correlations Relationship) 22

CHAPTER THREE 24
Methodology........ 24
3.1 Introduction 24
3.2 Collect and describe data........ 25
3.3 Data preprocessing 25
3.3.1 Data integration 26
3.3.2 Handling the missing value 26
3.3.3 Feature sub set selection 27
3.3.4 Data transformation 27
3.4 Data mining tool........ 28
3.5 Clustering 28
3.6 Association Rules Mining 32
CHAPTER FOUR 39
DISCUSSION OF RESULTS ANDFUTUR WORK........ 39
4.1 Introduction 39
4.2 Discus the generated rules 39
4.2.1 The first year rules 39
4.2.2 The second year rules 41
4.3.3 The third year rules........ 43
4.3.4 The fourth year rules........ 45
4.3.5 The fifth year rules........ 47
4.4 Discussion guidance rules that help student to improve their performance 49
4.5 Conclusion........ 50
4.6 future works........ 50
References 50

Table of Contents

Table 1 Patterns of data extracted from educational information systems 15
Table 2 common methods between the educational data mining and analysis of the learning process 16

Table 3 the most important applications of educational data mining according to beneficiary . 19

Table 4 the discretization 28

Table 5 how k means cluster work to grouping student .... **Error! Bookmark not defined.**

Table of Figure

Figure 1 Data mining as a step in the process of knowledge discovery [4] 11

Figure 3 General framework procedural in Educational Data Mining.. 16

Figure 4.1 sample from the dataset.. 25

Figure 5 missing data of more than one 26

Figure 6 missing data of one course 27

Figure 7 orange continuize widget 27

Figure 8 the Silhouette score for first year 29

Figure 9 the Silhouette score for the second year 30

Figure 10 the Silhouette score for the third year 30

Figure 11 the Silhouette score for the fourth year... 31

Figure 12 the Silhouette score for the fifth year 31

Figure 13 generated strong rules on cluster number (1) in the first year 32

Figure 14 generated strong rules on cluster number (2) in the first year 33

Figure 15 generated strong rules on cluster number (3) in the first year 33

Figure 16 generated strong rules on cluster number (1) in the second year.............................. 34

Figure 17 generated strong rules on cluster number (2) in the second year.............................. 34

Figure 18 generated strong rules on cluster number (3) in the second year.............................. 34

Figure 19 generated strong rules on cluster number (4) in the second year.............................. 35

Figure 20 generated strong rules on cluster number (1) in the third year 35

Figure 21 generated strong rules on cluster number (2) in the third year 36

Figure 22 generated strong rules on cluster number (3) in the third year 36

Figure 23 generated strong rules on cluster number (1) in the fourth year 36

Figure 24 generated strong rules on cluster number (2) in the fourth year 37

Figure 25 generated strong rules on cluster number (3) in the fourth year 37

Figure 26 generated strong rules on cluster number (4) in the fourth year 37

Figure 27 generated strong rules on cluster number (1) in the fifth year................................... 38

Figure 28 generated strong rules on cluster number (2) in the fifth year................................... 38

Figure 29 generated strong rules on cluster number (3) in the fifth year................................... 38

Table of terms

SAP	Student Academic Performance
GPA	Grade Point Average
EDM	Educational Data Mining
KDD	Knowledge Discovery in Databases

CHAPTER ONE

1. Introduction

Introduces the current research with the background of the problem described first. After that, the problem statement, objective, methodology, scope, expected contribution, and the structure of this thesis are described respectively.

1.1 Research background

Educational Data Mining (EDM) is a new trend in the data mining and Knowledge Discovery in Databases (KDD) field which focuses in mining useful patterns and discovering useful knowledge from the educational information systems, such as, admissions systems, registration systems, course management systems (Moodle, blackboard, etc…), and any other systems dealing with students at different levels of education, from schools, to colleges and universities (Saa, 2016). Researchers in this field focus on discovering useful knowledge either to help the educational institutes manage their students better, or to help students to manage their education and deliverables better and enhance their performance.

Analyzing students' data and information to classify students, or to create decision trees or association rules, to make better decisions or to enhance student's performance is an interesting field of research, which mainly focuses on analyzing and understanding students' educational data that indicates their educational performance, and generates specific rules, classifications, and predictions to help students in their future educational performance.

Data mining and knowledge discovery applications have got a great attention due to its significance in decision making and it has become an essential component in various organizations including universities where educational data is mostly available. Moreover, knowledge extraction has got an additional opportunity since data mining techniques have been introduced into new fields of Statistics, Databases, Machine Learning, Pattern Reorganization, Artificial Intelligence and Computation capabilities, etc. (Ahmed and Elaraby, 2014)

The major motivation behind educational data mining in universities is that there are often information “hidden” in the data that are not readily evident, which may take enormous human effort, Educational Data Mining for Students’ Academic time and therefore cost to extract. Furthermore, with exponential increase in the processing power of machines now available today, it is possible for data mining search algorithms to quickly filter data, extracting significant and embedded information as required. The ability to extract important embedded information in data suffices in many situations, helping organizations, companies, and research analysts make significant progress on different problems and decisions that are based on more information.

By this we able to predict student performance and extract knowledge that describes pattern that influence of students’ academic performance

That help us to earlier identifying the students who need special attention and allow the teacher to provide appropriate advising/counseling or devising appropriate teaching methods.

1.2 Problem statement

The problem of research is, there is a large amount of data in the college database, has not been exploited yet to find students success and failure factors. Also sometimes the students get low grades in the courses and sometimes they fail which negatively affects the cumulative GPA and there is a lack of faculty academic guidance that implements modern methods to help students enhance their academic performance.

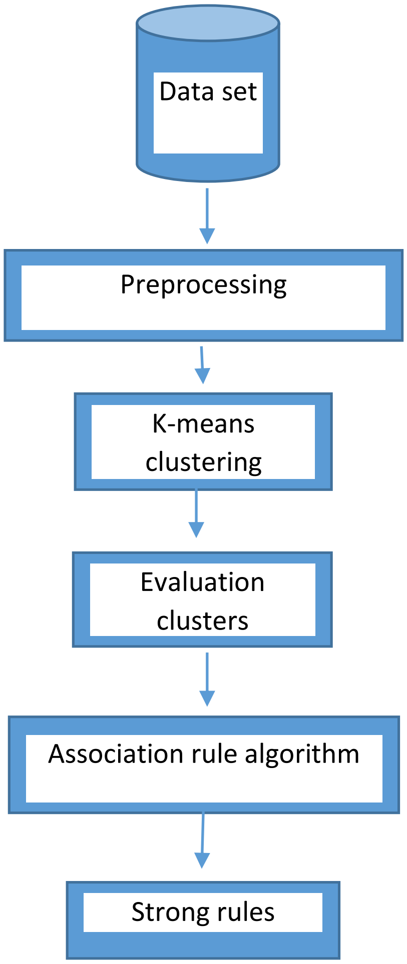
1.3 Objectives

- To providing guidance rule that implemented modern methods to help students enhance their academic performance
- To discover courses that more related with GPA and affect negatively and positively at GPA

1.4 Methodology

The method suggested in this research to know the pattern that affect in student performance and improve students’ academic performance by using Data Mining technique.

- Data collection is gathering all information available on students considering factors affect student performance.
- Preprocessing data is a necessary step for preparing the dataset before applying clustering.
- Clustering data into classes or clusters.
- Evaluate cluster
- Then generate association rules



1.5 Scope of the research

This research applied on student’s data from department of information system Faculty computer science and information Technology University of Kassala (2012-2018) to generate strong rule using data mining techniques

1.6 contribution

This research aim to improve students' academic performance. Collected student data since 2012 to 2018, merged it into five files, prepared the data to apply mining techniques and produced strong rules that help student to improve their academic performance.

1.7 Thesis organization

This research has four chapters organized as follows: Chapter one contains introduction. Chapter two discusses the literature review and related work. Chapter three describes the research methodology and the implementation. Chapter four presents the results and their discussion and Future work.

CHAPTER TWO

Literature review

2.1 Introduction

Education is an essential element for the betterment and progress of a country. Identifying the factors that influence student academic performance is essential to provide timely and effective support interventions. This chapter mainly describes the state of the art articles in analyzing the student performance using different techniques. Relevant information sources and related publications are mentioned. The first part presents data mining techniques and especially that was implemented in the educational section. Then describes educational institutions management process then education Data Mining (EDM) finally Student academic performance (SAP)

2.2 DATA MINING

2.2.1 Data mining definition:

Data mining refers to extracting or "mining" knowledge from large amounts of data. The term is actually a misnomer. Remember that the mining of gold from rocks or sand is referred to as gold mining rather than rock or sand mining. Thus, data mining should have been more appropriately named "knowledge mining from data," which is unfortunately somewhat long. "Knowledge mining," a shorter term may not reflect the emphasis on mining from large amounts of data. Others view data mining as simply an essential step in the process of knowledge discovery. (Zeynu and Patil, 2018)

2.2.2 Data mining steps:

Knowledge discovery as a process is depicted in *Figure 2.1* below and consists of an iterative sequence of the following steps

- Data cleaning (to remove noise and inconsistent data)
- Data integration (where multiple data sources may be combined)
- Data selection (where data relevant to the analysis task are retrieved from the database)
- Data transformation(where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance)
- Data mining (an essential process where intelligent methods are applied in order to extract data patterns)
- Pattern evaluation (to identify the truly interesting patterns representing knowledge Based on some interestingness measures)
- Knowledge presentation (where visualization and knowledge representation techniques are used to present the mined knowledge to the user)

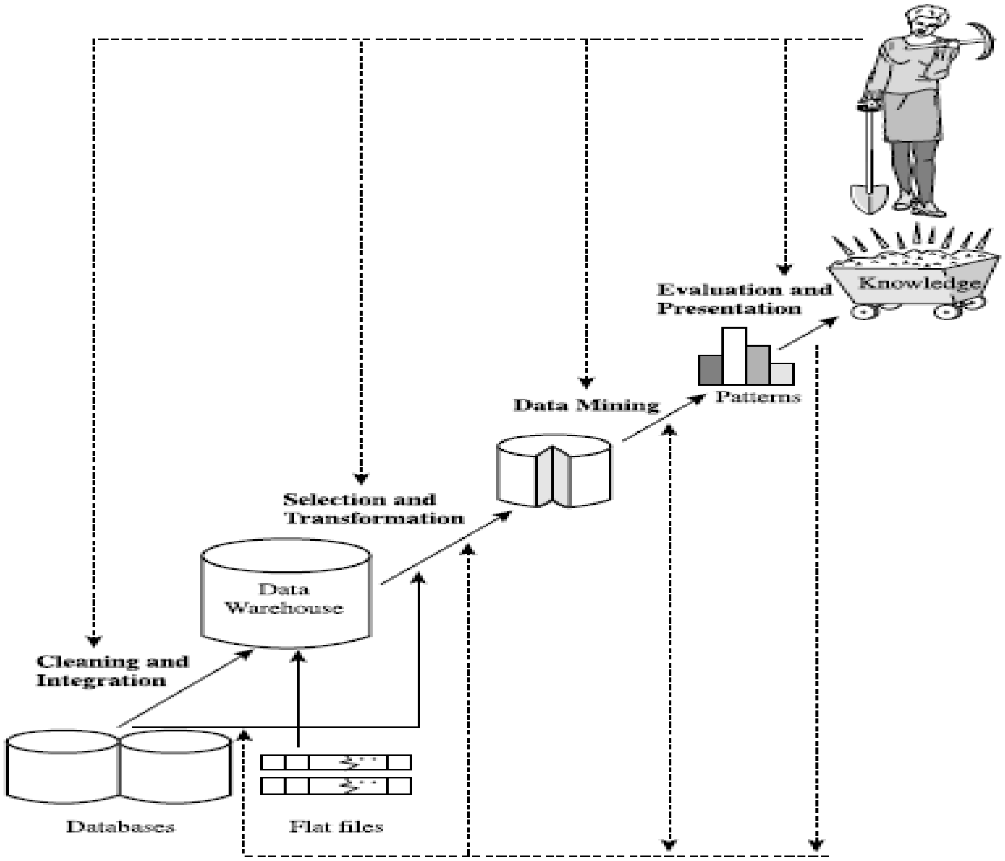
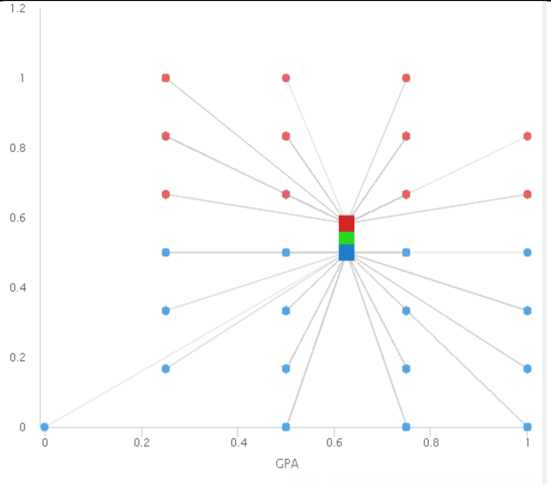
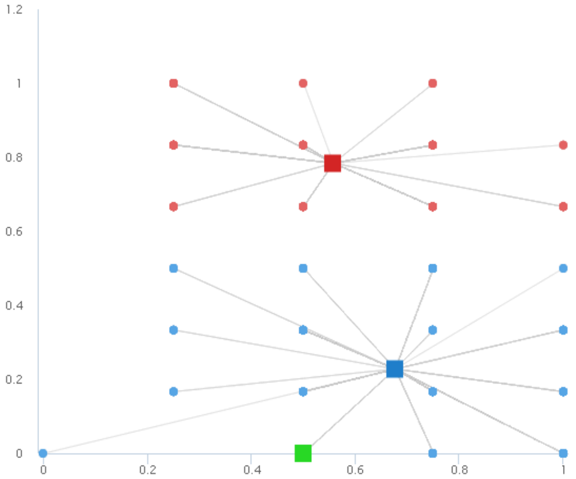
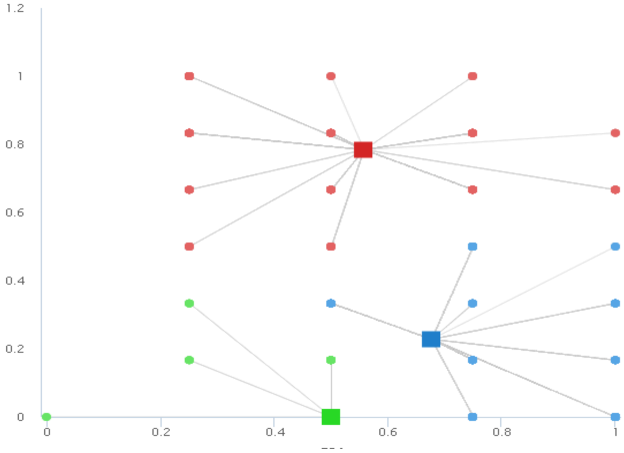
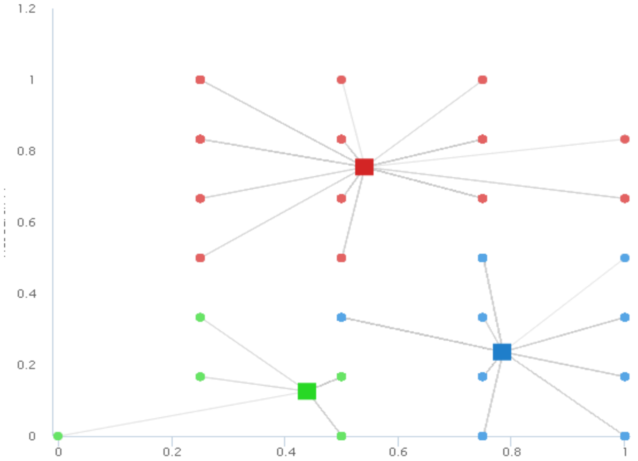


Figure 1 Data mining as a step in the process of knowledge discovery [4]

2.2.3 Clustering

K-mean clustering technique is used in this research. The main idea of the K-mean clustering is to define k centers, one for each cluster. The next step is to take each point belonging to a given dataset and associate it with the nearest centers. At this point, need to re-compute *k* new centers of the clusters resulting from the previous step. After having these *k* new centers, a new binding has to be done between the same data set points and the nearest new centers; a loop has been generated. As a result of this loop, may notice that the k centers change their location step by step until no more changes are done. (Mining, 2016)

Table 5 explain how k means cluster work to grouping student. Implemented on student data as example

 randomize center Chosen	 recomputed center and reassign position 1
 recomputed center and reassign position 2	 recomputed center and reassign position 3

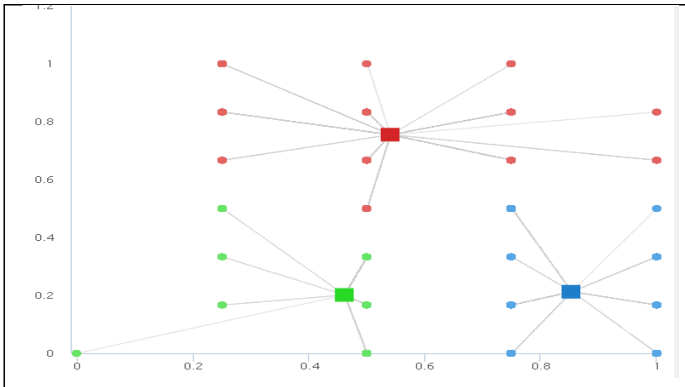
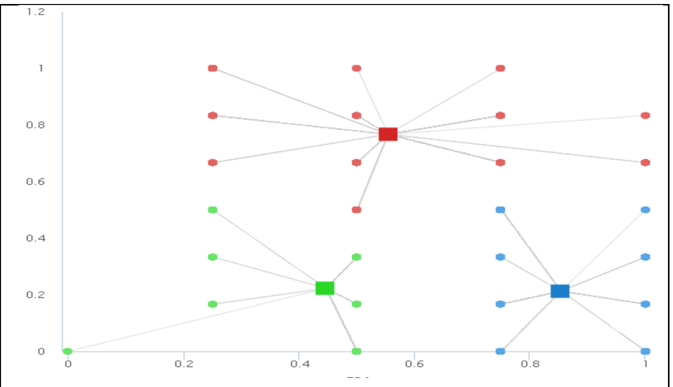
 recomputed center and reassign position 4	 recomputed center and reassign position 5

Table 1 how k means cluster work to grouping student

2.2.4 Association Rules Mining

Association rules are one of the most popular ways of the representing discovered knowledge and describe a close correlation between frequent items in a database. There are many association rule algorithms. The association rule algorithms require the user to set at least two thresholds, minimum support and minimum confidence. The support S of a rule $X \Rightarrow Y$ is defined as the probability that an entry has of satisfying both X and Y. Confidence is defined as the probability an entry has of satisfying Y when it satisfies X.

2.2.5 Data mining tool

Orange is an open-source data mining and analytics program that offers opportunities such as data preparation, exploratory data analysis and data modeling. With Orange, data exploration can be achieved by visual programming or by Python scripts. It has components for machine learning and plug-ins for bioinformatics and text mining. Orange brings data analysis functions in it. It contains modules for visually rich and flexible programming, analysis and user friendly data visualization. Orange has Python libraries for connecting and coding. It provides complete set of components such as data preprocessing, rating and filtering functionality, modeling, evaluation of models and exploration techniques (orange, 2018)

2.3 Educational institutions management process

The educational institutions management process is one of the difficulties faced by the supervisors because of the large size and complexity of its structure and the multiple sources of data. Therefore, the educational institution faces several problems during the management of the educational process, including academic, financial and administrative. These problems need to be studied, conclusions and recommendations that contribute to the decision-making process that facilitates the process of education based on an information system that is built in advance in a modern scientific way. In the data of the institutions on students, graduates and faculty members, and the correlation of all this with the most important indicators of performance such as student achievement, survival rate or leakage and quality of performance of faculty members. One of the biggest problems affecting in the performance of the educational process in universities is Student cannot decide about their field of study before they are enrolled in specific field of study.(Gulati and Sharma, 2012)

2.4 Education Data Mining (EDM)

Education Data Mining (EDM) is growing at a very fast pace. The main aim of EDM is to develop methods to explore the unique type of data that comes from educational institutes and to use those methods to understand the students and their learning environments. EDM deals with mining of large data sets of educational data to answer educational research questions. These data sets may come from learning management systems, interactive learning environments, intelligent tutoring systems, or any system used in a learning context.

The educational institutions management process is one of the difficulties faced by the supervisors because of the large size and complexity of its structure and the multiple sources of data. Therefore, the educational institution faces several problems during the management of the educational process, including academic, financial and administrative. These problems need to be studied, conclusions and recommendations that contribute to the decision-making process that facilitates the process of education based on an information system that is built in advance in a modern scientific way. In the data of the institutions on students, graduates and faculty members, and the correlation of all this with the most important indicators of performance such as student

achievement, survival rate or leakage and quality of performance of faculty members. (Zeynu and Patil, 2018)

The spread of the use of educational information systems in the institutions of higher education and the emergence of new concepts in teaching and learning, such as e-Learning and distance learning to the availability of a large amount of data mining from these systems, which led to the search for ways to extract information to improve the performance of students and teachers. There are different methods of data mining depending on the types of data applied, as data patterns can be classified into the following basic categories:

Table -1 Patterns of data extracted from educational information systems

Table 2 Patterns of data extracted from educational information systems

Data Pattern	**Methods used in data mining**
Structured organizational data extracted from relational databases	Basic exploration methods are used, such as classification, clustering, correlation and prediction relationships
Historical data expressed in time series	Special prospecting methods applied to time series such as prediction, Expectation and correlation study
Texts	Methods of textual exploration
Multimedia data such as images, audio and video	Data mining methods of multimedia
Data generated by web applications	Data manipulation methods provided by Web applications are three different forms: 1 - Prospecting in the content of pages. 2. Prospecting in the structure of pages. 3-Excavation in the records of the course.

The methods used in educational data mining, they are same methods used in the traditional methods of data, the need to understand the environment that will be handled and collect data to clean, arrange and select the techniques to be applied and finally interpretation of the results as well as verify the validity of the techniques being applied, taking into account the different methods; Using results from the privacy of educational environment and purpose of prospecting (Zeynu and Patil, 2018)The procedure of mining in educational data can be summarized as follows:

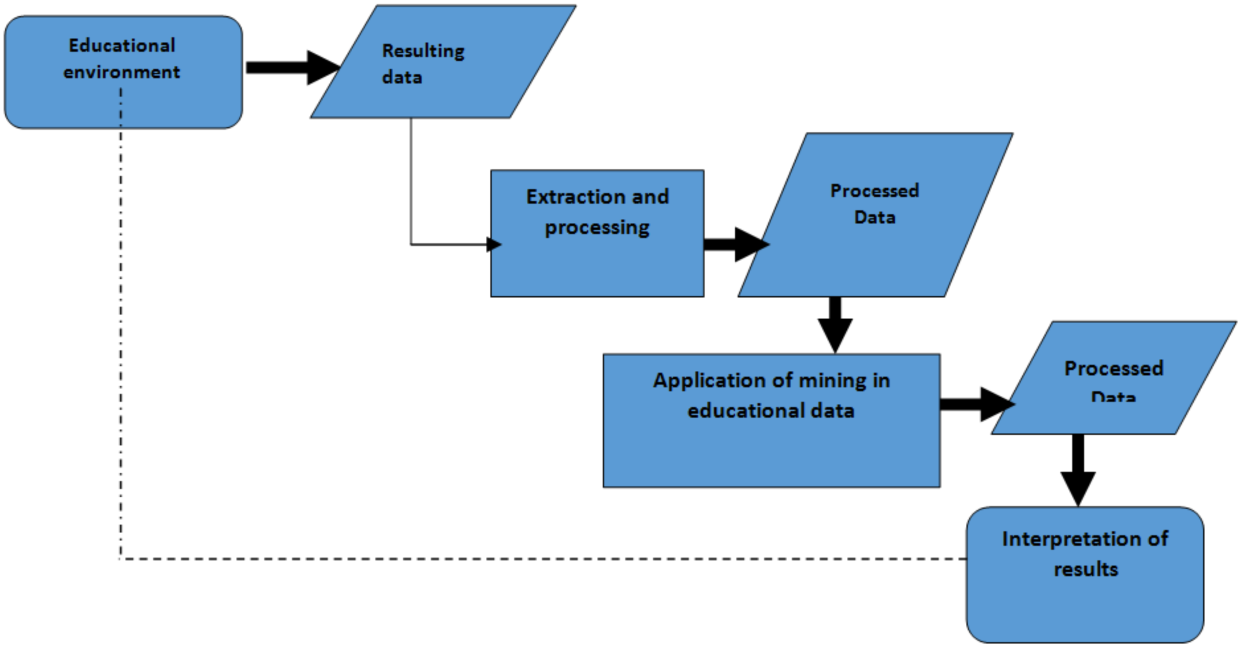

Educational environment
Resulting data
Extraction and processing
Processed Data
Application of mining in educational data
Processed Data
Interpretation of results

Figure 2 General framework procedural in Educational Data Mining

There are many methods are used in the process of educational data mining in according to the application and the objectives for which used; the following is review of the most important of these general methods in addition to the most important applications:

Table-2. The common methods between the educational data mining and analysis of the learning process

Table 3 common methods between the educational data mining and analysis of the learning process

Methods	**Objectives**	**Basic applications**

Prediction	Predict the values of variables based on the values of other variables being filtered. Forecasting uses the following basic methods: decision tree, classifications, aggregation, and neural networks	There are many applications to be used to predict a student result, and accordingly correcting student behavior to get the best expected performance.
Clustering	Assemble students in homogeneous and similar groups	Identify appropriate mechanisms to deal with similar student groups in learning style and social communication method.
correlations Relationship	Find causal relationships between variables and the most important methods in use.	Discover the weaknesses of the learners to improve them. Studying causal relationships in the educational process and discovering patterns of weakness to improve them.
Text Mining	Extract valuable information from text.	Analyze student conversations in forums to discover problems. Analyzing the file of the movements resulting from the student wandering in the educational system in order to follow him and extract useful information about his interests.
Social Network Analysis	Discover and analyses relationships through social networks	Analyzing the nature of relationships and interaction in communication networks;

		interactive tools in order to discover the student's educational style and discover weaknesses and preferences and other applications.

Table 4 the most important applications of educational data mining according to beneficiary

Beneficiary	**Applications of Educational Data Mining**
Student	To discover the weaknesses of the student and suggest educational resources and educational activities help in improving his level. To discover the student's learning style in order allocates a learning session for each student.
Teacher	To help to obtain objective analysis and feedback on the method of education in order to improve it. To identify students who are in need of support. To expect student performance for direction guidance To categorize students according to their levels or educational. To identify activities of the most active students in delivering knowledge. To improve the allocation of educational content.
Designers of methods, programs and study plans	To evaluate and improve curricula in terms of content. To evaluate and improve study plans. To identify the teacher's educational model; and

	the student's model as well as design the study programs accordingly.
Higher administration of educational institutions	To Improve the decision-making process at the level of senior management after studying the indicators resulting from the use of mining methods
Managers of educational systems	To determine the best way to display and design electronic educational content. To choose the best design for distance learning. To Identify and preserve the value of indicators that should be studied to improve the quality of education. Best artistic design.

Table 5 the most important applications of educational data mining according to beneficiary

2.5 Student academic performance (SAP)

The SAP prediction on will allow IHL to study what features of a model are important for prediction and to get the hidden information in students' data.

2.6 There are many methods are used in the process of educational data mining in according to the application and the objectives for which used

2.6.1 Prediction SAP

There are a lot of researches conducted to develop an SAP prediction model for particular courses or subjects. (Golding, Facey-Shaw and Tennant, 2006) (Chen, 2018) (Sembiring *et al.*, 2011) (Ahmad, Ismail and Aziz, 2015) (Ahmed and Elaraby, 2014) (Shovon and Haque, 2012) (Talley and Scherer, 2013) ,One of these studies is a study presents an applied study in data mining and knowledge discovery. It aims at discovering patterns within historical students' academic and financial data at UST (University of Science and Technology) from the year 1993 to 2005 in order to contribute improving academic performance at UST. Results show that these rules concentrate on three main issues, students' academic achievements (successes and failures), students' drop out, and students' financial behavior. Clustering (by K-means algorithm), association rules (by Apriori algorithm) and decision trees by (J48 and Id3 algorithms) techniques have been used to build the data model. Results have been discussed and analyses comprehensively and then well evaluated by experts in terms of some criteria such as validity, reality, utility, and originality. In addition, practical evaluation using SQL queries have been applied to test the accuracy of produced model (rules), the shortcomings of this study are as follows:

a) Not included of Associate student data, educational staff and other university branches.

b) Not included in scholarship data and lots of personal data for students.

c) Not included attendance and absence data for students. (Al-shargabi and Nusari, 2010)

Another study was conducted in Ethiopia In Debre_Markos University study has shown that data mining techniques can be applied by higher education institutions or universities in determining student failure/success rate so that managing students' enrolment at the beginning of the year, assist students before they reached risk of failure, effective resource utilization and cost minimization, helping and guiding administrative officers to be successful in management and decision making. The study applied data mining technology to the data of university students for the purpose of forecasting the success or failure of students, the study used CRISP methodology the analysis was carried out by the WEKA program and the forecast model was built the study found the main class, number of courses given in a semester, and field of study are the major factors affecting the student performances. (Asif *et al.*, 2017)

Another research considered that one of the common tools to evaluate instructors' performance is the course evaluation questionnaire to evaluate based on students' perception. In this study, classification algorithm of Naïve Bayes and C5.0 are used to build classifier models. Their performances are compared over a dataset composed of answer of students to a real course evaluation questionnaire using accuracy, precision, recall, and specificity performance metrics. Although all the classifier models show comparably high classification performances, Naïve Bayes classifier is the best with respect to accuracy, precision, and specificity. In addition, an analysis of the variable importance for each classifier model is done. This research describes the performances of classification algorithms used in building a model does not necessarily indicate that the one that used the least time is the best model to use. Some Algorithms can take the least time but may not produce the best result in term of accuracy. This research used classification algorithms and data mining techniques such as, Naïve Bayes classifier, C5.0 as well as data from universities. Naïve Bayes classifier is the best with respect to accuracy, precision, and specificity. (Patil, 2017)

Another study looks at and compare well performing algorithms such as Naïve Bayes, decision tree (J48), Random Forest, Naïve Bayes Multiple Nominal, K-star and IBk. And it mentions Educational Data mining is a relatively new field and has a lot of potential to help society if used in the proper manner. The study compared six algorithms J48 (Decision Tree), Random Forest, Naive Bayes, Naive Bayes Multinomial, K-star, IBk. In the comparative study of all these algorithms can see that the closest we got in terms of getting an accurate prediction was the Random Forest Technique which narrowly edged the J48 to claim the top spot. This was that was done on a relatively larger dataset hence random forest becomes more accurate with the number of entries but all algorithms need modification if they can ever be used because the current amount of accuracy is low for this to be implemented on a large scale in the present state (Kapur, Ahluwalia and Sathyaraj, 2017)

Another study explored the opportunities of the Education data mining for improving students' performance. Educational data mining is used to study the data available in the educational field and bring out the hidden knowledge from it. The study used Classification methods like decision trees, Bayesian network etc. which can be applied on the educational data for predicting the student's performance in examination. This prediction will help to identify the weak students and help them to score better marks. The C4.5, ID3 and CART decision tree algorithms are applied on engineering student's data to predict their performance in the final exam. The results of this study provided predicted the number of students who are likely to pass, fail or promoted to next year, steps to improve the performance of the students who were predicted to fail or promoted and the comparative analysis of the results states that the prediction has helped the weaker students to improve and brought out betterment in the result (Yadav and Pal, 2012)

2.6.2 Identify appropriate mechanisms to deal with similar student groups in learning style and social communication method (clustering)

Using data mining to grouped student into similar group there are many researches grouping student using clustering one of these research is Using K-means clustering which used for pattern recognized classification and clustered students according to their class performance, sessional and attendance record. Using K-Means Clustering clustered the students based on their Class Performance, sessional and Attendance in class. Centroids are calculated from the educational data set taking Kclusters. This study is helpful to notify the students with less attendance and slow performance in sessional but also enhances the decision-making approach to monitor the performance of students. Also, on increasing the value of K, the accuracy becomes better with huge dataset and Kmeans can find the better grouping of the data. The results obtained help to cluster those students who need special attention (Guleria and Sood, 2014)

2.5.3 Discover the weaknesses of the learners to improve them, Studying causal relationships in the educational process and discovering patterns of weakness to improve them (correlations Relationship)

In further study, the researchers predicted the performance of students based on their academic levels in several areas; using the detection of the rules of correlation to ensure that the factors affecting the final outcome of the student linked to each other, having calculate the correlation coefficient while the student attributes were shown and the result was different from the resulting relationship using the detection of the rules of correlation. (Borkar and Rajeswari, 2013)

Other research analyze collected students' information through a survey, a survey was constructed that has targeted university students and collected multiple personal, social, and academic data related to them. The study explores multiple factors theoretically assumed to affect students' performance in higher education, and finds a qualitative model which best classifies and predicts the students' performance based on related personal and social factors , Four decision tree algorithms have been implemented, as well as, with the Naïve Bayes algorithm. It was slightly found that the student's performance is not totally dependent on their academic efforts (Saa, 2016) Also other study the researcher focuses on predicting performance of student at an early stage of the degree program, in order to help the university not only to focus more on bright students but also to initially identify students with low academic achievement and find ways to support them. Data set of 5729 students' record in this study was obtained from five university's 2014/2015 entry student records found in Amhara Regional state, Ethiopia. A model is built using C4.5 Decision tree learning algorithm – generates five classification rule set classifiers (predictors) in an experiment. The experiment using a test data set produces 81.4% accuracy the findings of the study have important implications for educators, teachers, counselors, university curriculum designers, students and other decision makers. (Tegegne and Alemu, 2018)

Table 4: Summary of Literatures

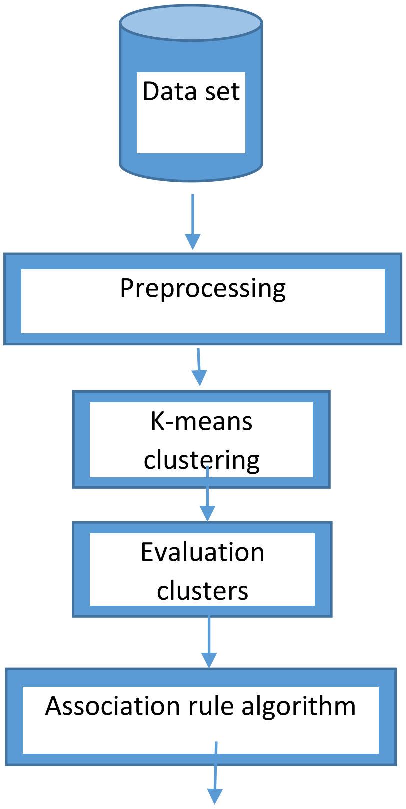
Author	**objective**	**techniques**	**result**
(Patil 2017)	Comparative between naive bays and c5	naive bays and c5	Naïve bays best
(Amjad 2016)	Prediction SAP	Naïve bays	Predictive model
(Borkar 2013)	Discover factor affect at final student out come	Aproiri algorithm	Set of rule
(yadav 2012)	Prediction SAP	ID3	model
(Guleria 2014)	Grouping student into similar group	k. means	Identify student that need special attention
(Al-shargabi & Nusari 2010)	Discovery pattern	Apriori algorithm	Rule
(asif & Ali 2017)	Predict SAP	CRISP Methodology	Forecast model
(kapur 2017)	Compare between Naive bays and J48	Naive bays and J48	Naïve bays
(Tegegne &Alemu 2018)	Predict SAP in first year study	C4.5	model

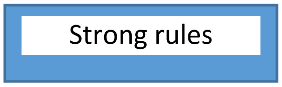
CHAPTER THREE

Methodology

3.1 Introduction

Present the structured methodology used in this research. The methodology was carried on five phases in order to achieve the objective of this research. The first phase explains the collecting data and describes it. The second phase explains data preprocessing technique that applied to the data set. The third phase is Appling clustering algorithm to grouping students into similar group. The fourth phase evaluation clusters and last phase implements the association rule algorithm.

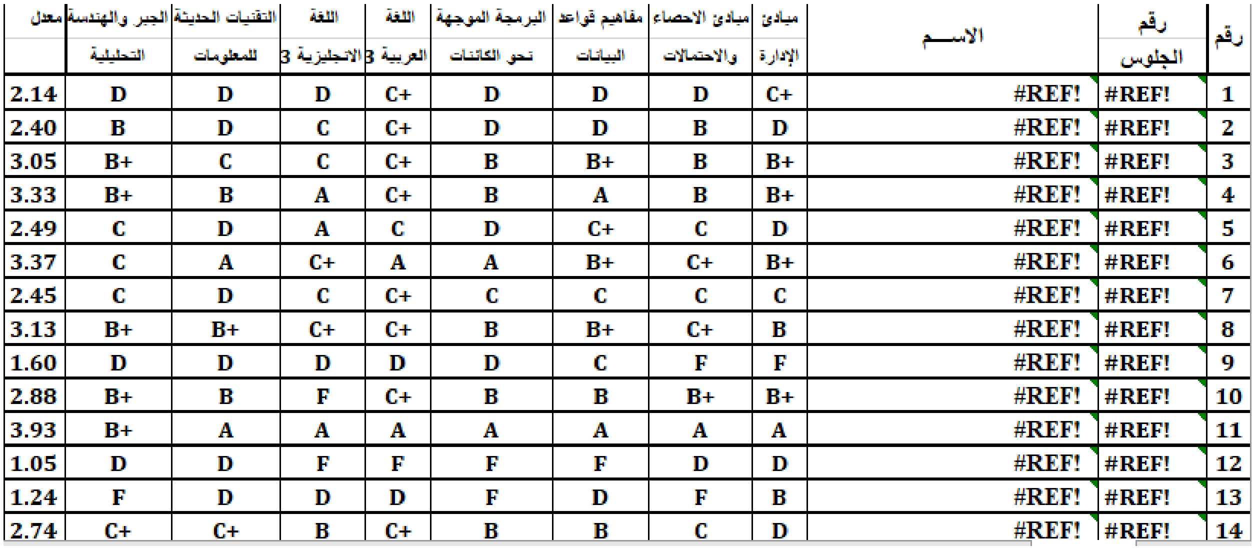




3.2 Collect and describe data

The data were collected from the Registrar's Office and the Information Systems Department at University of Kassala Faculty of Computer Science and Information Technology, the data contains 1700 records corresponding to students enrolled through the year’s study 2012 to 2018. The dataset contains the number of the index, student name, student number, courses, and average of courses. Should hide the name and number of student due to privacy.

The number of courses in faculty equals sixty-six courses affording to ten semesters, the structure of study in faculty has been distributed into five-year study equivalent to ten semesters. Each year study included two semesters. Figure 3-1 explains the sample of data.



معدل	الجبر والهندسة	التقنيات الحديثة	اللغة	اللغة	البرمجة الموجهة	مفاهيم قواعد	مبادئ الاحصاء	مبادئ	الاســـم	رقم	رقم
	التحليلية	للمعلومات	الانجليزية 3	العربية 3	نحو الكائنات	البيانات	والاحتمالات	الإدارة		الجلوس	
2.14	D	D	D	C+	D	D	D	C+	#REF!	#REF!	1
2.40	B	D	C	C+	D	D	B	D	#REF!	#REF!	2
3.05	B+	C	C	C+	B	B+	B	B+	#REF!	#REF!	3
3.33	B+	B	A	C+	B	A	B	B+	#REF!	#REF!	4
2.49	C	D	A	C	D	C+	C	D	#REF!	#REF!	5
3.37	C	A	C+	A	A	B+	C+	B+	#REF!	#REF!	6
2.45	C	D	C	C+	C	C	C	C	#REF!	#REF!	7
3.13	B+	B+	C+	C+	B	B+	C+	B	#REF!	#REF!	8
1.60	D	D	D	D	D	C	F	F	#REF!	#REF!	9
2.88	B+	B	F	C+	B	B	B+	B+	#REF!	#REF!	10
3.93	B+	A	A	A	A	A	A	A	#REF!	#REF!	11
1.05	D	D	F	F	F	F	D	D	#REF!	#REF!	12
1.24	F	D	D	D	F	D	F	B	#REF!	#REF!	13
2.74	C+	C+	B	C+	B	B	C	D	#REF!	#REF!	14

Figure 3.1 sample from the dataset

3.3 Data preprocessing

Some general tasks of the data preprocessing have to be performed on the dataset, such as data integration, data cleaning, data reduction, data transformation. The first task of the data preprocessing is

3.3.1 Data integration

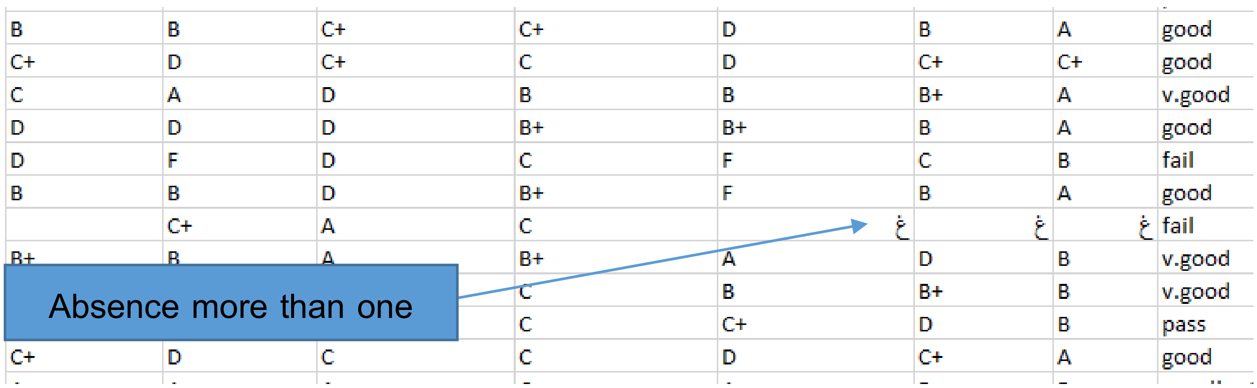
Merged student’s data into five files. The first semesters and second semesters to one file (first year), third semesters and fourth semesters to one file (second year), fifth semesters and sixth semesters to one file (third year), seventh semesters and eight semesters to one file (fourth years), ninth semesters and tenth semesters to one file (fifth years).

3.3.2 Handling the missing value

Missing data generally arises due to the student absence from examination, may be absence one exam and may more and we have more than one way to handling missing value

3.3.2.1 Ignoring the tuple

When the student is absented of the more than one exam Figure (4) showed this.

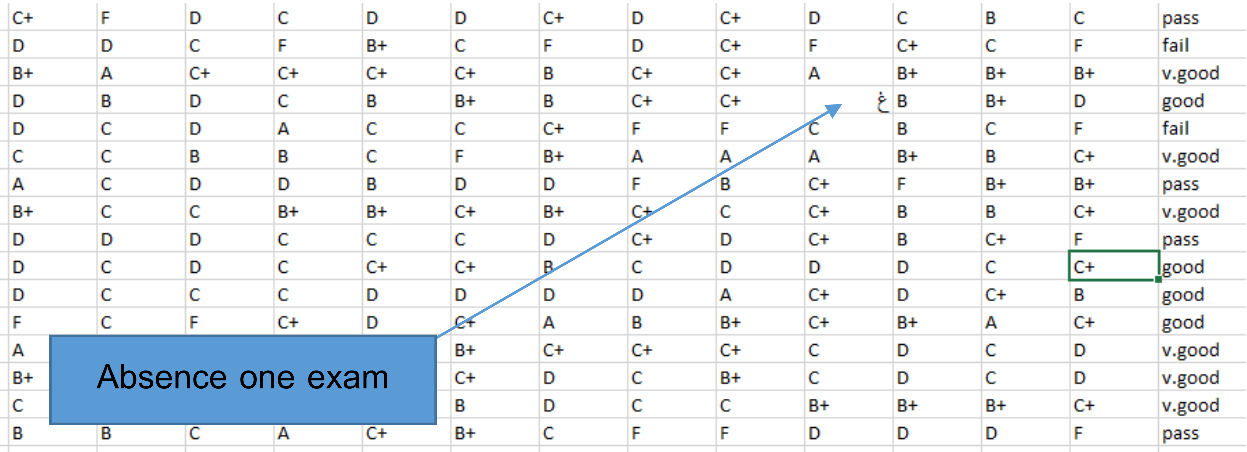


B	B	C+	C+	D	B	A	good
C+	D	C+	C	D	C+	C+	good
C	A	D	B	B	B+	A	v.good
D	D	D	B+	B+	B	A	good
D	F	D	C	F	C	B	fail
B	B	D	B+	F	B	A	good
	C+	A	C	غ	غ	غ	fail
B+	B	A	B+	A	D	B	v.good
			C	B	B+	B	v.good
			C	C+	D	B	pass
C+	D	C	C	D	C+	A	good

Figure 4 missing data of more than one

3.3.2.2 *handling by impute average*

When the student absent from one exam Figure (5) showed it



C+	F	D	C	D	D	C+	D	C+	D	C	B	C	pass
D	D	C	F	B+	C	F	D	C+	F	C+	C	F	fail
B+	A	C+	C+	C+	C+	B	C+	C+	A	B+	B+	B+	v.good
D	B	D	C	B	B+	B	C+	C+	غ	B	B+	D	good
D	C	D	A	C	C	C+	F	F	C	B	C	F	fail
C	C	B	B	C	F	B+	A	A	A	B+	B	C+	v.good
A	C	D	D	B	D	D	F	B	C+	F	B+	B+	pass
B+	C	C	B+	B+	C+	B+	C+	C	C+	B	B	C+	v.good
D	D	D	C	C	C	D	C+	D	C+	B	C+	F	pass
D	C	D	C	C+	C+	B	C	D	D	D	C	C+	good
D	C	C	C	D	D	D	D	A	C+	D	C+	B	good
F	C	F	C+	D	C+	A	B	B+	C+	B+	A	C+	good
A					B+	C+	C+	C+	C	D	C	D	v.good
B+					C+	D	C	B+	C	D	C	D	v.good
C					B	D	C	C	B+	B+	B+	C+	v.good
B	B	C	A	C+	B+	C	F	F	D	D	D	F	pass

Absence one exam

Figure 5 missing data of one course

3.3.3 Feature sub set selection

The feature selection is one of the important and frequently used techniques in data preprocessing for data mining. Feature selection is a process of identifying and selecting a useful subset from original features. In this study will be select names of courses and corresponding student's degrees and the GPA of students.

3.3.4 Data transformation

The data should transform into forms appropriate for mining need. Data transformations involve many techniques. In this study normalizing the courses marks cause the k means cluster use to grouping student into similar group. Using orange continuize widget to normalized data Figure 6 show orange continuize widget

Figure 6 orange continuize widget

And use discretization technique to discretize student GPA since the performance of students is compared by association rule table 4 show the discretization.

Range of student GPA	Discretized GPA
4.00-3.50	excellent
3.49-3.00	v.good
2.99-2.50	good
2.49-2.00	pass
1.99-0	fail

Table 6 the discretization

3.4 Data mining tool

This study use orange is an open-source data mining and analytics program that offers opportunities such as data preparation, exploratory data analysis and data modeling. With Orange, data exploration can be achieved by visual programming or by Python scripts. It has components for machine learning and plug-ins for bioinformatics and text mining. Orange brings data analysis functions in it. It contains modules for visually rich and flexible programming, analysis and user friendly data visualization. Orange has Python libraries for connecting and coding. It provides complete set of components such as data preprocessing, rating and filtering functionality, modeling, evaluation of models and exploration techniques

3.5 Clustering

K-mean clustering technique is used in this research. Orange evaluate cluster by introduce Silhouette scoring reports .Silhouette contrasts average distance to elements in the same cluster with the average distance to elements in other clusters. k-Means use the silhouette score and guess the best number of cluster.

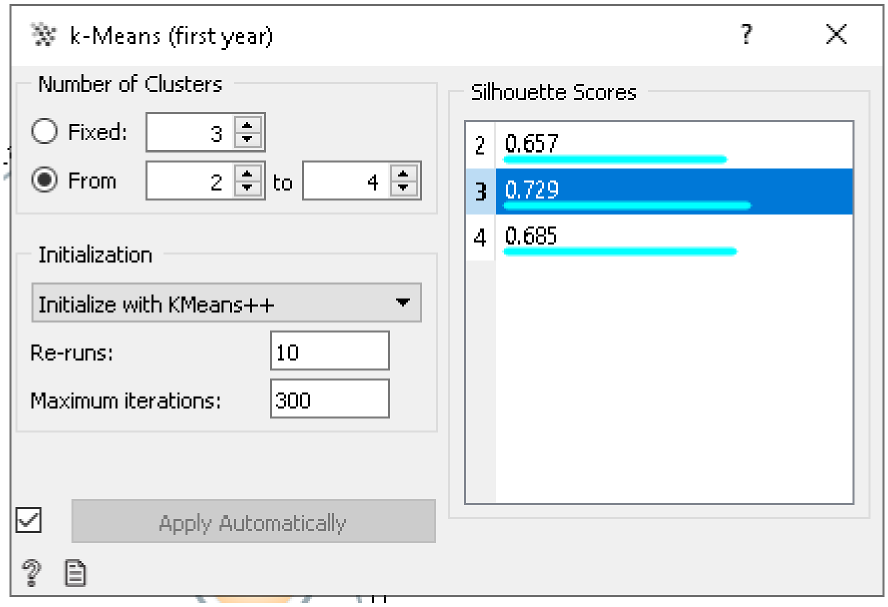
Explain k-Means widget in orange

Select the number of clusters have two option.

- **Fixed**: algorithm clusters data in a specified number of clusters you use this option if you know the best number of k.

• **Optimized**: widget shows clustering scores for the selected cluster range. Use this option if you want orange suggests the best number of k orange use Silhouette score to suggests the best number of clusters, Silhouette (contrasts average distance to elements in the same cluster with the average distance to elements in other clusters)

The datasets are deployed in orange tool and then k means clustering algorithms are applied to the datasets. The datasets equivalent to five files that mentioned in section 3.3.1 in this chapter. Each file will be deploying separately. Figure 7 show the Silhouette score for first year



k-Means (first year)

Number of Clusters

○ Fixed: 3

● From 2 to 4

Initialization

Initialize with KMeans++

Re-runs: 10

Maximum iterations: 300

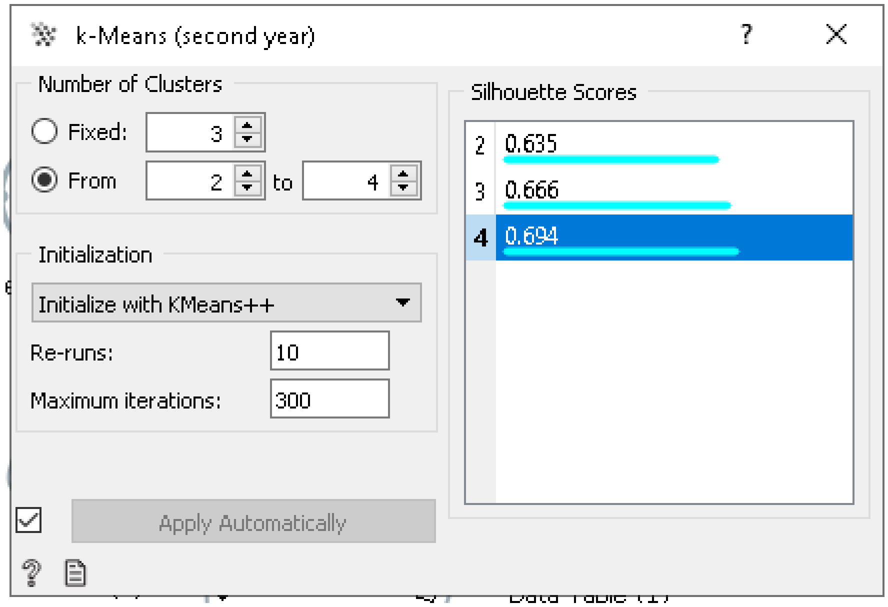
☑ Apply Automatically

Silhouette Scores

2	0.657
3	0.729
4	0.685

Figure 7 the Silhouette score for first year

That mean the best cluster when number of cluster equal 3.   Student data in first year divided to 3 cluster because the number of student who got an excellent grade and who failed few. In the same way applying k means on all files. The results of the best cluster in each year as shows in below figures. figure 8 show the best cluster in the second year when number of cluster equal 4.figure 9 show the best cluster in the third year when number of cluster equal 4 figure 10 show the best cluster in the fourth year when number of cluster equal 4.figure 11 show the best cluster in the fifth year when number of cluster equal 3.



k-Means (second year)

Number of Clusters

○ Fixed: 3

● From 2 to 4

Initialization

Initialize with KMeans++

Re-runs: 10

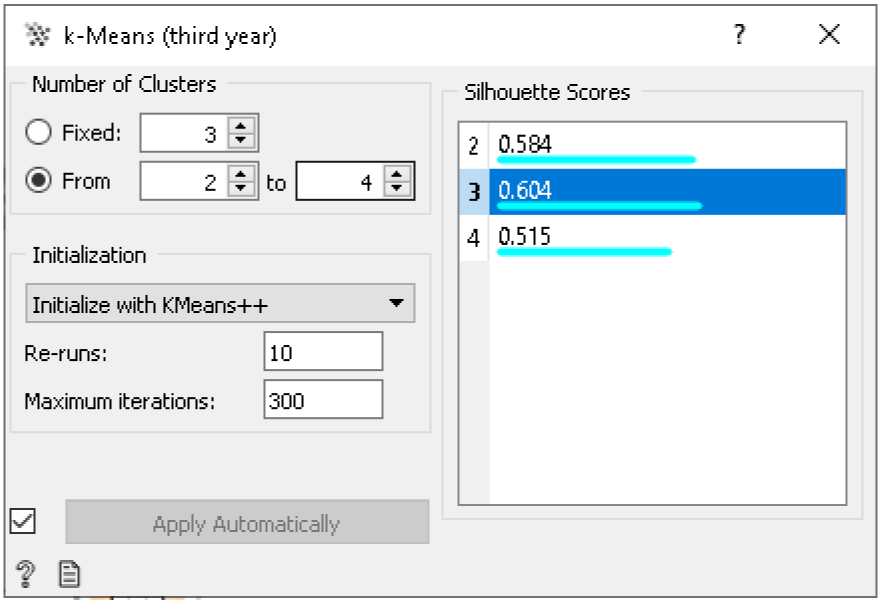
Maximum iterations: 300

☑ Apply Automatically

Silhouette Scores

2	0.635
3	0.666
4	0.694

Figure 8 the Silhouette score for the second year



k-Means (third year)

Number of Clusters

○ Fixed: 3

● From 2 to 4

Initialization

Initialize with KMeans++

Re-runs: 10

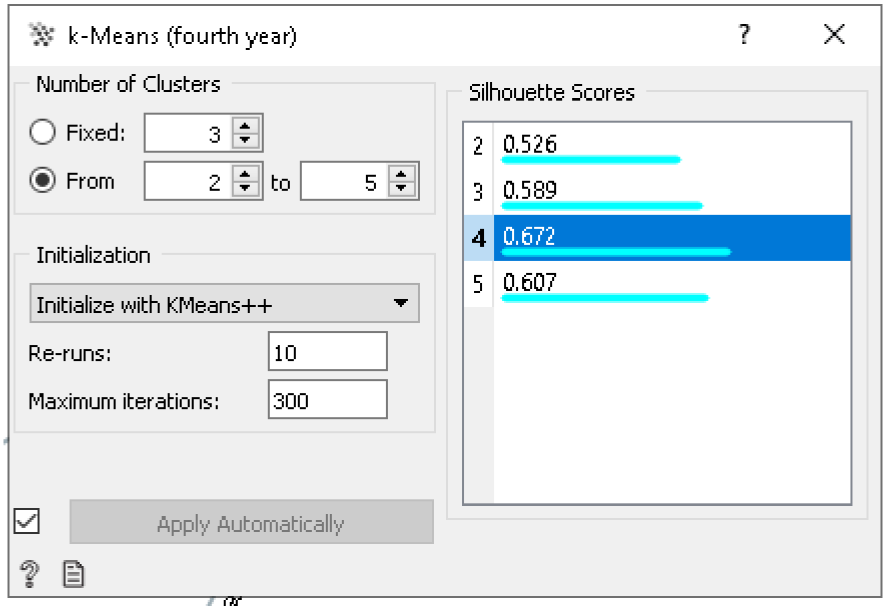
Maximum iterations: 300

☑ Apply Automatically

Silhouette Scores

2	0.584
3	0.604
4	0.515

Figure 9 the Silhouette score for the third year



k-Means (fourth year)

Number of Clusters

○ Fixed: 3

● From 2 to 5

Initialization

Initialize with KMeans++

Re-runs: 10

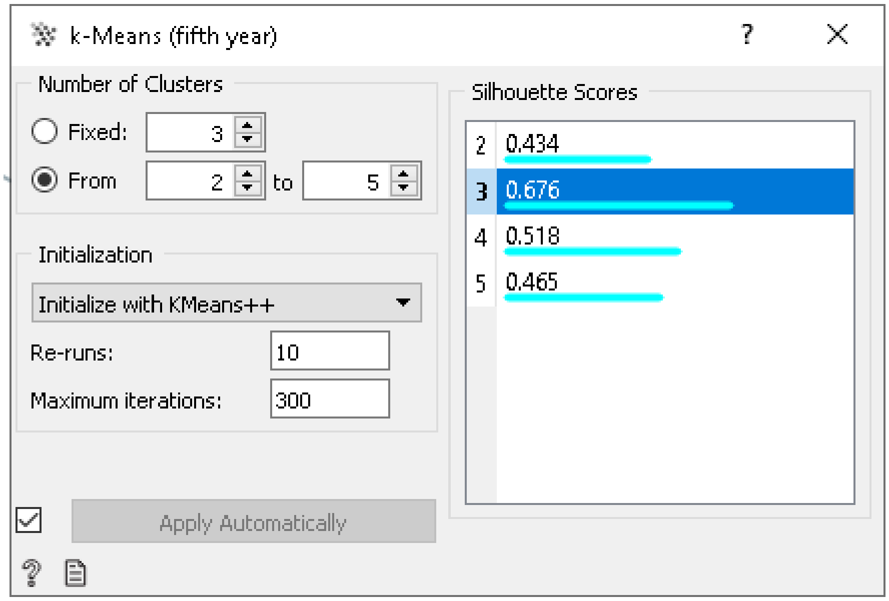
Maximum iterations: 300

☑ Apply Automatically

Silhouette Scores

2	0.526
3	0.589
4	0.672
5	0.607

Figure 10 the Silhouette score for the fourth year



k-Means (fifth year)

Number of Clusters

○ Fixed: 3

● From 2 to 5

Initialization

Initialize with KMeans++

Re-runs: 10

Maximum iterations: 300

☑ Apply Automatically

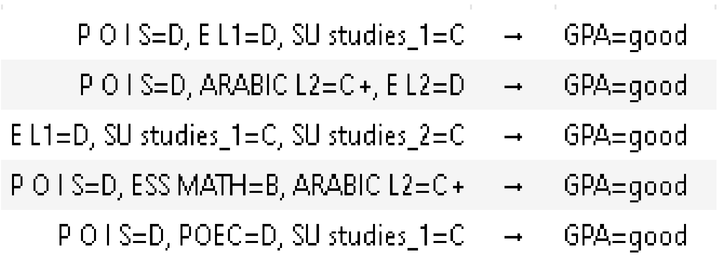
Silhouette Scores

2	0.434
3	0.676
4	0.518
5	0.465

Figure 11 the Silhouette score for the fifth year

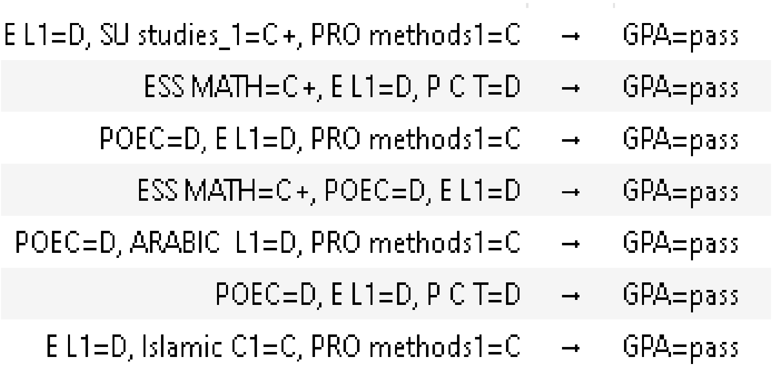
3.6 Association Rules Mining

this study applied tow data mining techniques to get best rule like applied association rule on results of correct clusters there are generated in the previous sections. in this research use orange association rule widget this widget implements FP-growth algorithm. In the first implement association rule on the data of the first year which is contents three clusters. The first cluster observed was generated strong rule with (Support 12% and Confidence 92%) figur12 show the strong generated rules on cluster number (1) in the first year. The second cluster observed was generated strong rule with (Support 14% and Confidence 88%) figure 13 show the strong generated rules on cluster number (2) in the first year. The third cluster observed was generated strong rule with (Support 16% and Confidence 89%) figur14 show the strong generated rules on cluster number (3) in the first year.



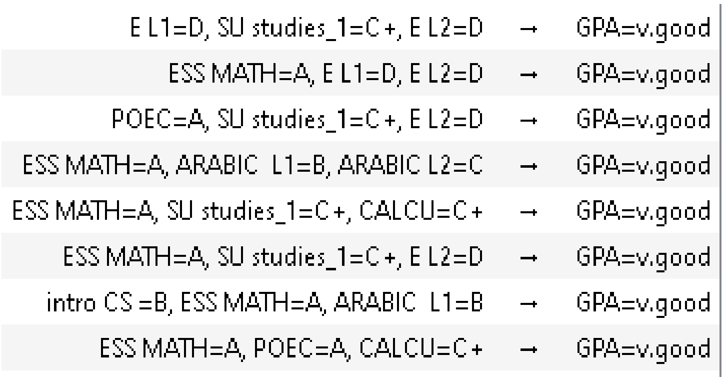
P O I S=D, E L1=D, SU studies_1=C	→	GPA=good
P O I S=D, ARABIC L2=C+, E L2=D	→	GPA=good
E L1=D, SU studies_1=C, SU studies_2=C	→	GPA=good
P O I S=D, ESS MATH=B, ARABIC L2=C+	→	GPA=good
P O I S=D, POEC=D, SU studies_1=C	→	GPA=good

Figure 12 generated strong rules on cluster number (1) in the first year



E L1=D, SU studies_1=C+, PRO methods1=C	→	GPA=pass
ESS MATH=C+, E L1=D, P C T=D	→	GPA=pass
POEC=D, E L1=D, PRO methods1=C	→	GPA=pass
ESS MATH=C+, POEC=D, E L1=D	→	GPA=pass
POEC=D, ARABIC L1=D, PRO methods1=C	→	GPA=pass
POEC=D, E L1=D, P C T=D	→	GPA=pass
E L1=D, Islamic C1=C, PRO methods1=C	→	GPA=pass

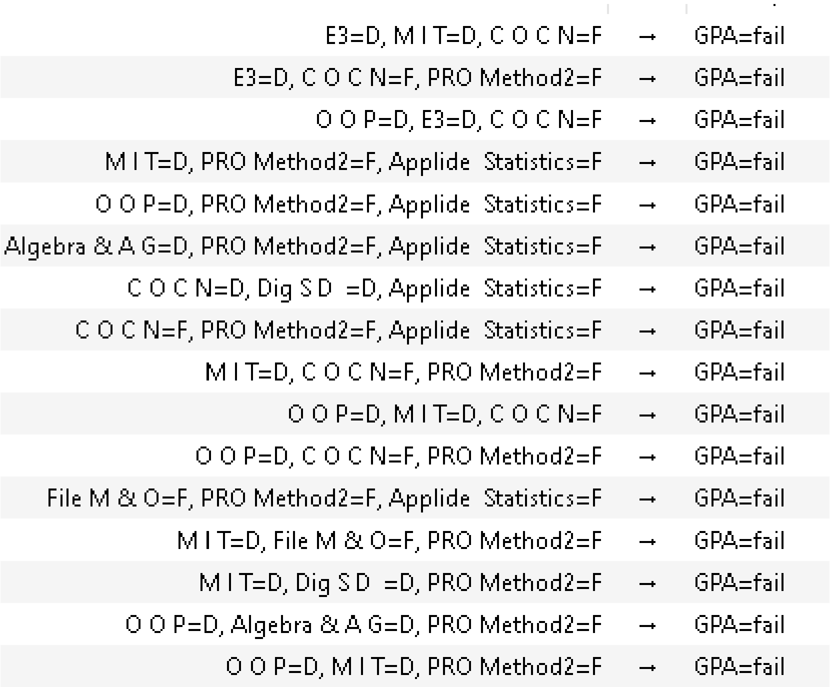
Figure 13 generated strong rules on cluster number (2) in the first year



E L1=D, SU studies_1=C+, E L2=D	→	GPA=v.good
ESS MATH=A, E L1=D, E L2=D	→	GPA=v.good
POEC=A, SU studies_1=C+, E L2=D	→	GPA=v.good
ESS MATH=A, ARABIC L1=B, ARABIC L2=C	→	GPA=v.good
ESS MATH=A, SU studies_1=C+, CALCU=C+	→	GPA=v.good
ESS MATH=A, SU studies_1=C+, E L2=D	→	GPA=v.good
intro CS =B, ESS MATH=A, ARABIC L1=B	→	GPA=v.good
ESS MATH=A, POEC=A, CALCU=C+	→	GPA=v.good

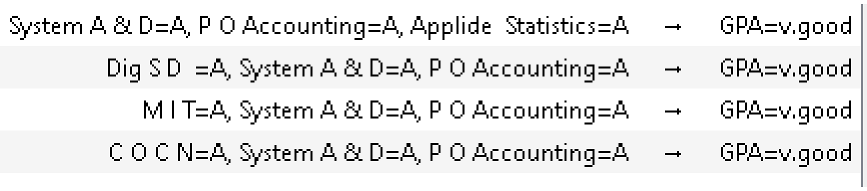
Figure 14 generated strong rules on cluster number (3) in the first year

As the similar way implement association rule on the data of the second year which is contents four clusters. The first cluster observed was generated strong rule with (Support 21% and Confidence 79%) fegur15 show the strong generated rules on cluster number (1) in the second year. The second cluster observed was generated strong rule with (Support 10% and Confidence 83%) fegur16 show the strong generated rules on cluster number (2) in the second year. The third cluster observed was generated strong rule with (Support 8 % and Confidence 87%) fegur17 show the strong generated rules on cluster number (3) in the second year. The fourth cluster observed was generated strong rule with (Support 11 % and Confidence 89%) fegur18 show the strong generated rules on cluster number (4) in the second year.



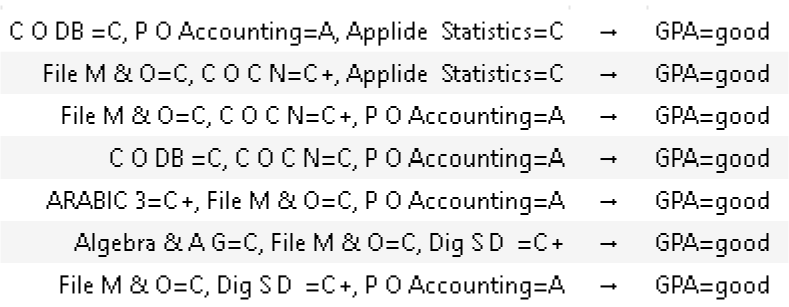
E3=D, M I T=D, C O C N=F	→	GPA=fail
E3=D, C O C N=F, PRO Method2=F	→	GPA=fail
O O P=D, E3=D, C O C N=F	→	GPA=fail
M I T=D, PRO Method2=F, Applide Statistics=F	→	GPA=fail
O O P=D, PRO Method2=F, Applide Statistics=F	→	GPA=fail
Algebra & A G=D, PRO Method2=F, Applide Statistics=F	→	GPA=fail
C O C N=D, Dig S D =D, Applide Statistics=F	→	GPA=fail
C O C N=F, PRO Method2=F, Applide Statistics=F	→	GPA=fail
M I T=D, C O C N=F, PRO Method2=F	→	GPA=fail
O O P=D, M I T=D, C O C N=F	→	GPA=fail
O O P=D, C O C N=F, PRO Method2=F	→	GPA=fail
File M & O=F, PRO Method2=F, Applide Statistics=F	→	GPA=fail
M I T=D, File M & O=F, PRO Method2=F	→	GPA=fail
M I T=D, Dig S D =D, PRO Method2=F	→	GPA=fail
O O P=D, Algebra & A G=D, PRO Method2=F	→	GPA=fail
O O P=D, M I T=D, PRO Method2=F	→	GPA=fail

Figure 15 generated strong rules on cluster number (1) in the second year



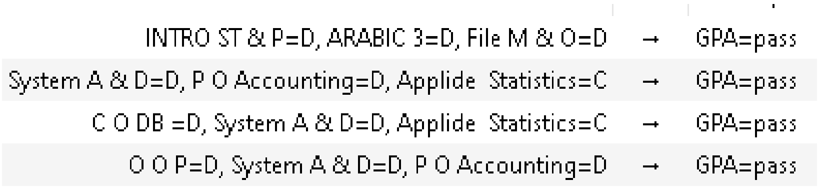
System A & D=A, P O Accounting=A, Applide Statistics=A	→	GPA=v.good
Dig S D =A, System A & D=A, P O Accounting=A	→	GPA=v.good
M I T=A, System A & D=A, P O Accounting=A	→	GPA=v.good
C O C N=A, System A & D=A, P O Accounting=A	→	GPA=v.good

Figure 16 generated strong rules on cluster number (2) in the second year



C O DB =C, P O Accounting=A, Applide Statistics=C	→	GPA=good
File M & O=C, C O C N=C+, Applide Statistics=C	→	GPA=good
File M & O=C, C O C N=C+, P O Accounting=A	→	GPA=good
C O DB =C, C O C N=C, P O Accounting=A	→	GPA=good
ARABIC 3=C+, File M & O=C, P O Accounting=A	→	GPA=good
Algebra & A G=C, File M & O=C, Dig S D =C+	→	GPA=good
File M & O=C, Dig S D =C+, P O Accounting=A	→	GPA=good

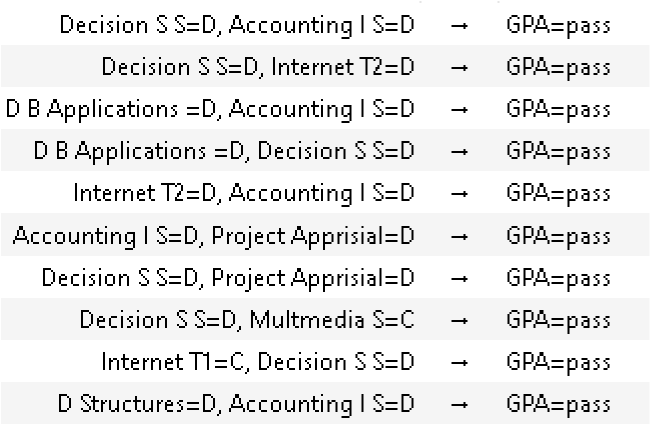
Figure 17 generated strong rules on cluster number (3) in the second year



INTRO ST & P=D, ARABIC 3=D, File M & O=D	→	GPA=pass
System A & D=D, P O Accounting=D, Applide Statistics=C	→	GPA=pass
C O DB =D, System A & D=D, Applide Statistics=C	→	GPA=pass
O O P=D, System A & D=D, P O Accounting=D	→	GPA=pass

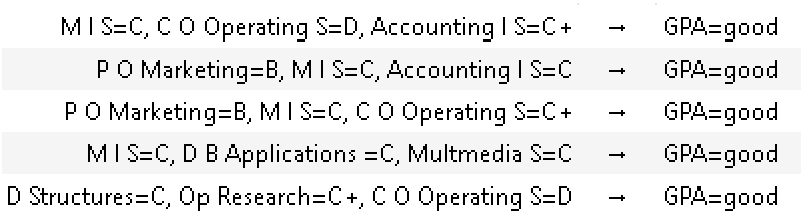
Figure 18 generated strong rules on cluster number (4) in the second year

Implement association rule on the data of the third year which is contents three clusters. The first cluster observed was generated strong rule with (Support 25% and Confidence 86%) fegur19 show the strong generated rules on cluster number (1) in the third year. The second cluster observed was generated strong rule with (Support 10% and Confidence 79%) fegur20 show the strong generated rules on cluster number (2) in the third year. The third cluster observed was generated strong rule with (Support 17 % and Confidence 81%) fegur21 show the strong generated rules on cluster number (3) in the third year.



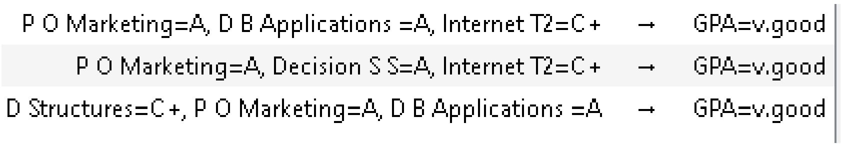
Decision S S=D, Accounting I S=D	→	GPA=pass
Decision S S=D, Internet T2=D	→	GPA=pass
D B Applications =D, Accounting I S=D	→	GPA=pass
D B Applications =D, Decision S S=D	→	GPA=pass
Internet T2=D, Accounting I S=D	→	GPA=pass
Accounting I S=D, Project Apprisial=D	→	GPA=pass
Decision S S=D, Project Apprisial=D	→	GPA=pass
Decision S S=D, Multmedia S=C	→	GPA=pass
Internet T1=C, Decision S S=D	→	GPA=pass
D Structures=D, Accounting I S=D	→	GPA=pass

Figure 19 generated strong rules on cluster number (1) in the third year



M I S=C, C O Operating S=D, Accounting I S=C+	→	GPA=good
P O Marketing=B, M I S=C, Accounting I S=C	→	GPA=good
P O Marketing=B, M I S=C, C O Operating S=C+	→	GPA=good
M I S=C, D B Applications =C, Multmedia S=C	→	GPA=good
D Structures=C, Op Research=C+, C O Operating S=D	→	GPA=good

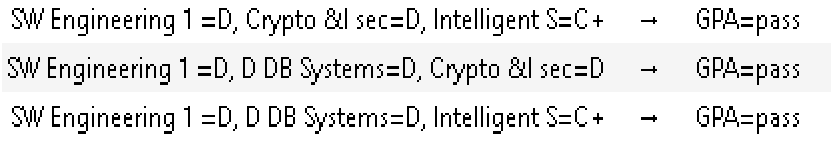
Figure 20 generated strong rules on cluster number (2) in the third year



P O Marketing=A, D B Applications =A, Internet T2=C+	→	GPA=v.good
P O Marketing=A, Decision S S=A, Internet T2=C+	→	GPA=v.good
D Structures=C+, P O Marketing=A, D B Applications =A	→	GPA=v.good

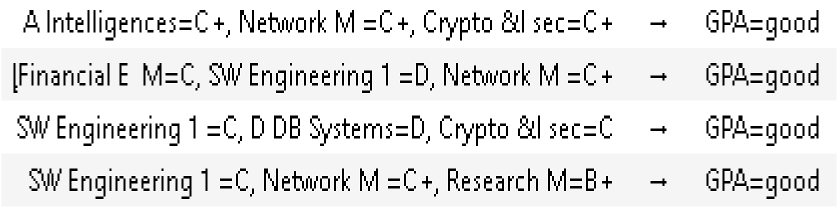
Figure 21 generated strong rules on cluster number (3) in the third year

Implement association rule on the data of the fourth year which is contents four clusters. The first cluster observed was generated strong rule with (Support 15% and Confidence 90%) figure 22 show the strong generated rules on cluster number (1) in the fourth year. The second cluster observed was generated strong rule with (Support 7% and Confidence 84%) figure 23 show the strong generated rules on cluster number (2) in the fourth year. The third cluster observed was generated strong rule with (Support 29 % and Confidence 77%) figure 24 show the strong generated rules on cluster number (3) in the fourth year. The fourth cluster observed was generated strong rule with (Support 17 % and Confidence 88%) figure 25 show the strong generated rules on cluster number (4) in the fourth year.



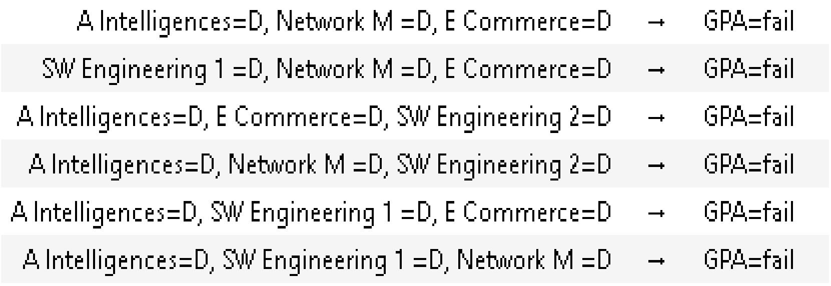
SW Engineering 1 =D, Crypto &I sec=D, Intelligent S=C+	→	GPA=pass
SW Engineering 1 =D, D DB Systems=D, Crypto &I sec=D	→	GPA=pass
SW Engineering 1 =D, D DB Systems=D, Intelligent S=C+	→	GPA=pass

Figure 22 generated strong rules on cluster number (1) in the fourth year



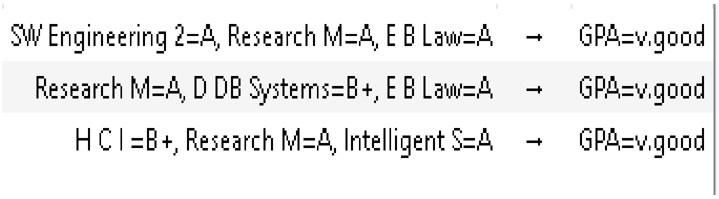
A Intelligences=C+, Network M =C+, Crypto &I sec=C+	→	GPA=good
|Financial E M=C, SW Engineering 1 =D, Network M =C+	→	GPA=good
SW Engineering 1 =C, D DB Systems=D, Crypto &I sec=C	→	GPA=good
SW Engineering 1 =C, Network M =C+, Research M=B+	→	GPA=good

Figure 23 generated strong rules on cluster number (2) in the fourth year



A Intelligences=D, Network M =D, E Commerce=D	→	GPA=fail
SW Engineering 1 =D, Network M =D, E Commerce=D	→	GPA=fail
A Intelligences=D, E Commerce=D, SW Engineering 2=D	→	GPA=fail
A Intelligences=D, Network M =D, SW Engineering 2=D	→	GPA=fail
A Intelligences=D, SW Engineering 1 =D, E Commerce=D	→	GPA=fail
A Intelligences=D, SW Engineering 1 =D, Network M =D	→	GPA=fail

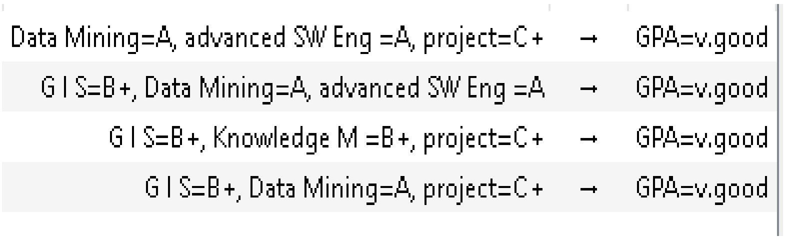
Figure 24 generated strong rules on cluster number (3) in the fourth year



SW Engineering 2=A, Research M=A, E B Law=A	→	GPA=v.good
Research M=A, D DB Systems=B+, E B Law=A	→	GPA=v.good
H C I =B+, Research M=A, Intelligent S=A	→	GPA=v.good

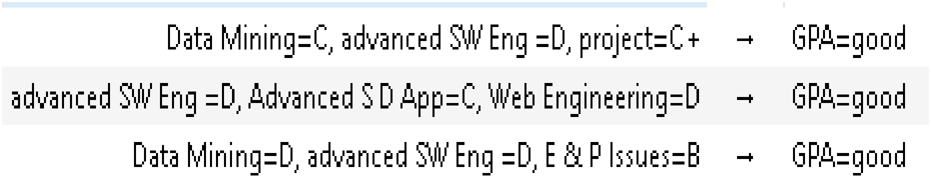
Figure 25 generated strong rules on cluster number (4) in the fourth year

Implement association rule on the data of the fifth year which is contents three clusters. The first cluster observed was generated strong rule with (Support 13% and Confidence 79%) figure 26 show the strong generated rules on cluster number (1) in the fifth year. The second cluster observed was generated strong rule with (Support 9% and Confidence 93%) figure 27 shows the strong generated rules on cluster number (2) in the fifth year. The third cluster observed was generated strong rule with (Support 20 % and Confidence 83%) figure 28 shows the strong generated rules on cluster number (3) in the fifth year.



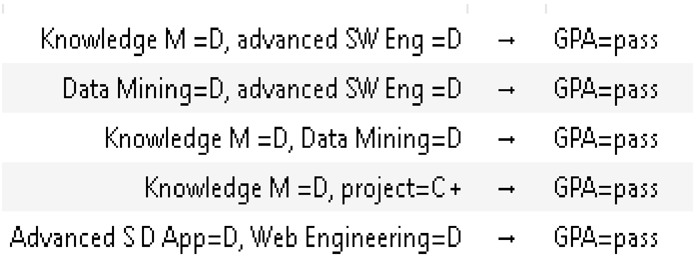
Data Mining=A, advanced SW Eng =A, project=C+	→	GPA=v.good
G I S=B+, Data Mining=A, advanced SW Eng =A	→	GPA=v.good
G I S=B+, Knowledge M =B+, project=C+	→	GPA=v.good
G I S=B+, Data Mining=A, project=C+	→	GPA=v.good

Figure 26 generated strong rules on cluster number (1) in the fifth year



Data Mining=C, advanced SW Eng =D, project=C+	→	GPA=good
advanced SW Eng =D, Advanced S D App=C, Web Engineering=D	→	GPA=good
Data Mining=D, advanced SW Eng =D, E & P Issues=B	→	GPA=good

Figure 27 generated strong rules on cluster number (2) in the fifth year



Knowledge M =D, advanced SW Eng =D	→	GPA=pass
Data Mining=D, advanced SW Eng =D	→	GPA=pass
Knowledge M =D, Data Mining=D	→	GPA=pass
Knowledge M =D, project=C+	→	GPA=pass
Advanced S D App=D, Web Engineering=D	→	GPA=pass

Figure 28 generated strong rules on cluster number (3) in the fifth year

CHAPTER FOUR

DISCUSSION OF RESULTS ANDFUTUR WORK

4.1 Introduction

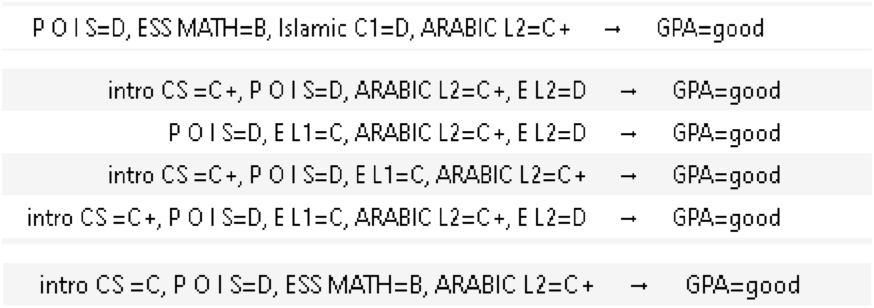
This chapter discussed the results that obtained in chapter three. explain and discus the strong rules were generated in chapter three. And suggested future work

4.2 Discus the generated rules

In this point discus the strong rules were generated in chapter three. The strong rule use as guidance rule to help to improving their academic performance for each year study

4.2.1 The first year rules

Rules of cluster number (1) in first year



P O I S=D, ESS MATH=B, Islamic C1=D, ARABIC L2=C+ → GPA=good

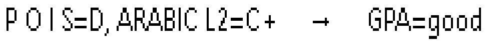
intro CS =C+, P O I S=D, ARABIC L2=C+, E L2=D → GPA=good

P O I S=D, E L1=C, ARABIC L2=C+, E L2=D → GPA=good

intro CS =C+, P O I S=D, E L1=C, ARABIC L2=C+ → GPA=good

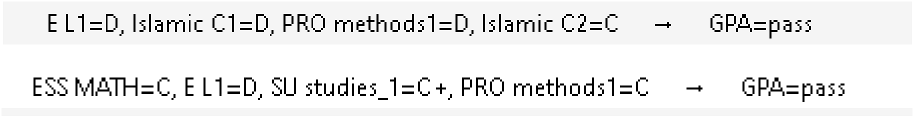
intro CS =C+, P O I S=D, E L1=C, ARABIC L2=C+, E L2=D → GPA=good

intro CS =C, P O I S=D, ESS MATH=B, ARABIC L2=C+ → GPA=good



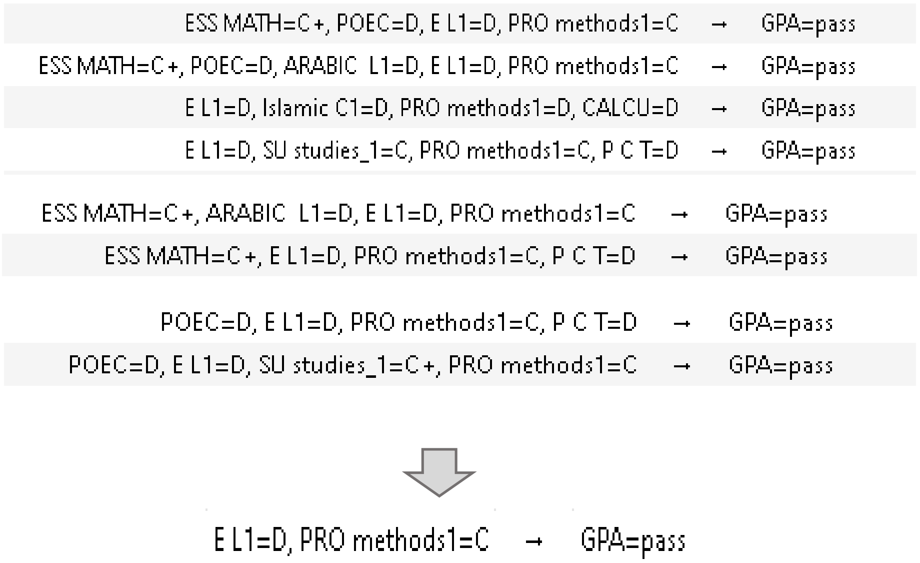
P O I S=D, ARABIC L2=C+ → GPA=good

Rules of cluster number (2) in first year



E L1=D, Islamic C1=D, PRO methods1=D, Islamic C2=C → GPA=pass

ESS MATH=C, E L1=D, SU studies_1=C+, PRO methods1=C → GPA=pass



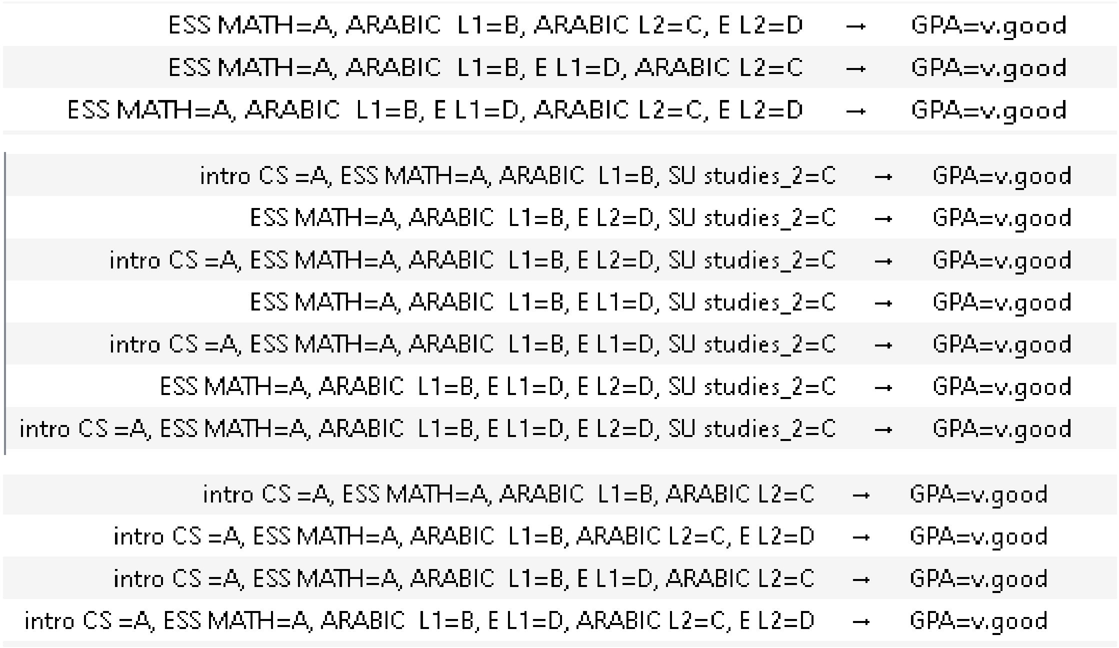
ESS MATH=C+, POEC=D, E L1=D, PRO methods1=C	→	GPA=pass
ESS MATH=C+, POEC=D, ARABIC L1=D, E L1=D, PRO methods1=C	→	GPA=pass
E L1=D, Islamic C1=D, PRO methods1=D, CALCU=D	→	GPA=pass
E L1=D, SU studies_1=C, PRO methods1=C, P C T=D	→	GPA=pass

ESS MATH=C+, ARABIC L1=D, E L1=D, PRO methods1=C	→	GPA=pass
ESS MATH=C+, E L1=D, PRO methods1=C, P C T=D	→	GPA=pass

POEC=D, E L1=D, PRO methods1=C, P C T=D	→	GPA=pass
POEC=D, E L1=D, SU studies_1=C+, PRO methods1=C	→	GPA=pass

E L1=D, PRO methods1=C	→	GPA=pass

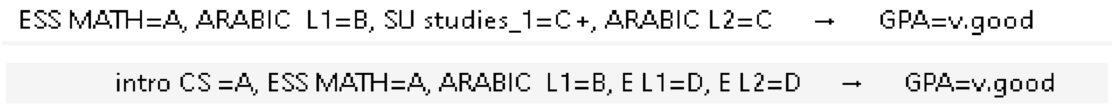
Rules of cluster number (3) in first year



ESS MATH=A, ARABIC L1=B, ARABIC L2=C, E L2=D	→	GPA=v.good
ESS MATH=A, ARABIC L1=B, E L1=D, ARABIC L2=C	→	GPA=v.good
ESS MATH=A, ARABIC L1=B, E L1=D, ARABIC L2=C, E L2=D	→	GPA=v.good

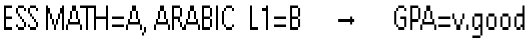
intro CS =A, ESS MATH=A, ARABIC L1=B, SU studies_2=C	→	GPA=v.good
ESS MATH=A, ARABIC L1=B, E L2=D, SU studies_2=C	→	GPA=v.good
intro CS =A, ESS MATH=A, ARABIC L1=B, E L2=D, SU studies_2=C	→	GPA=v.good
ESS MATH=A, ARABIC L1=B, E L1=D, SU studies_2=C	→	GPA=v.good
intro CS =A, ESS MATH=A, ARABIC L1=B, E L1=D, SU studies_2=C	→	GPA=v.good
ESS MATH=A, ARABIC L1=B, E L1=D, E L2=D, SU studies_2=C	→	GPA=v.good
intro CS =A, ESS MATH=A, ARABIC L1=B, E L1=D, E L2=D, SU studies_2=C	→	GPA=v.good

intro CS =A, ESS MATH=A, ARABIC L1=B, ARABIC L2=C	→	GPA=v.good
intro CS =A, ESS MATH=A, ARABIC L1=B, ARABIC L2=C, E L2=D	→	GPA=v.good
intro CS =A, ESS MATH=A, ARABIC L1=B, E L1=D, ARABIC L2=C	→	GPA=v.good
intro CS =A, ESS MATH=A, ARABIC L1=B, E L1=D, ARABIC L2=C, E L2=D	→	GPA=v.good



ESS MATH=A, ARABIC L1=B, SU studies_1=C+, ARABIC L2=C → GPA=v.good

intro CS =A, ESS MATH=A, ARABIC L1=B, E L1=D, E L2=D → GPA=v.good



ESS MATH=A, ARABIC L1=B → GPA=v.good

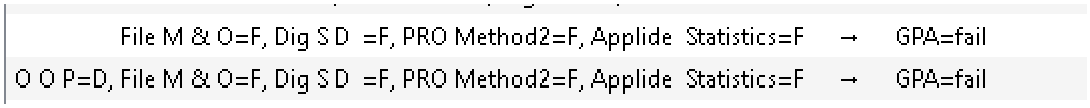
-The above outcomes clearly indicate that Probability of the student obtaining the grade (v. good) if the student obtains the grade(A) in Essential Mathematics , grade(B) in Arabic Language I and should at least pass all other courses.

-Probability of the student obtaining the grade (good) if the student obtains the grade(D) in Principles of Information System and grade(c+) in Arabic Language II.

-Probability of the student obtaining the grade (pass) if the student obtains the grade(D) in English language I and grade(c) in Programming Methods I.

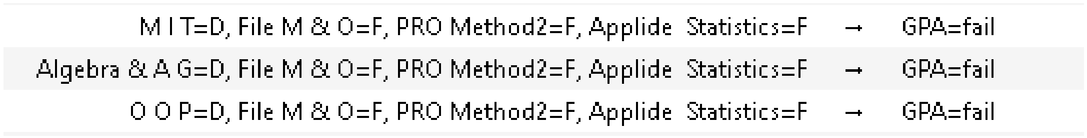
4.2.2 The second year rules

Rules of cluster number (1) in second year



File M & O=F, Dig S D =F, PRO Method2=F, Applide Statistics=F → GPA=fail

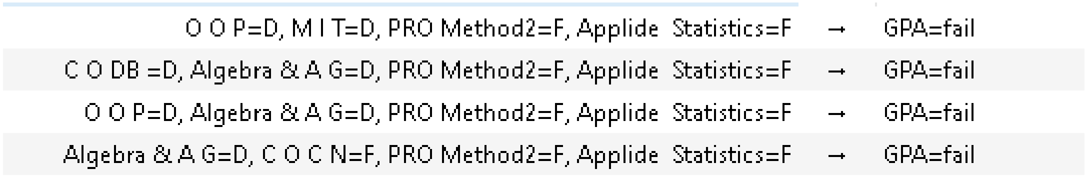
O O P=D, File M & O=F, Dig S D =F, PRO Method2=F, Applide Statistics=F → GPA=fail



M I T=D, File M & O=F, PRO Method2=F, Applide Statistics=F → GPA=fail

Algebra & A G=D, File M & O=F, PRO Method2=F, Applide Statistics=F → GPA=fail

O O P=D, File M & O=F, PRO Method2=F, Applide Statistics=F → GPA=fail

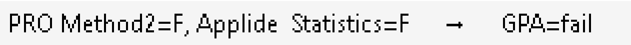


O O P=D, M I T=D, PRO Method2=F, Applide Statistics=F → GPA=fail

C O DB =D, Algebra & A G=D, PRO Method2=F, Applide Statistics=F → GPA=fail

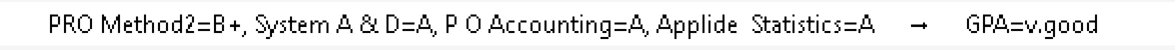
O O P=D, Algebra & A G=D, PRO Method2=F, Applide Statistics=F → GPA=fail

Algebra & A G=D, C O C N=F, PRO Method2=F, Applide Statistics=F → GPA=fail

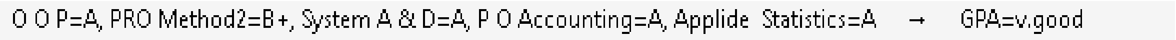


PRO Method2=F, Applide Statistics=F → GPA=fail

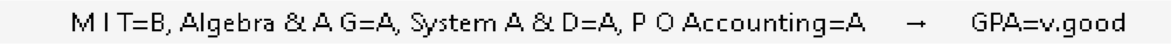
Rules of cluster number (2) in second year



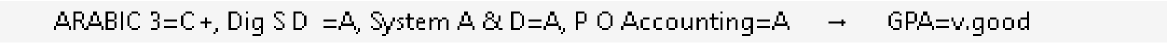
PRO Method2=B+, System A & D=A, P O Accounting=A, Applide Statistics=A → GPA=v.good



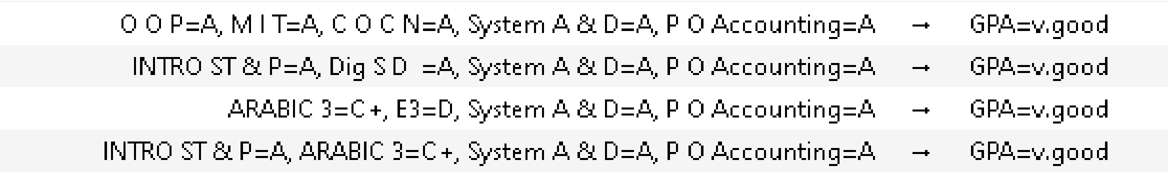
O O P=A, PRO Method2=B+, System A & D=A, P O Accounting=A, Applide Statistics=A → GPA=v.good



M I T=B, Algebra & A G=A, System A & D=A, P O Accounting=A → GPA=v.good



ARABIC 3=C+, Dig S D =A, System A & D=A, P O Accounting=A → GPA=v.good

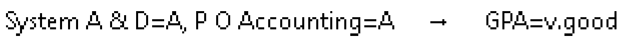


O O P=A, M I T=A, C O C N=A, System A & D=A, P O Accounting=A → GPA=v.good

INTRO ST & P=A, Dig S D =A, System A & D=A, P O Accounting=A → GPA=v.good

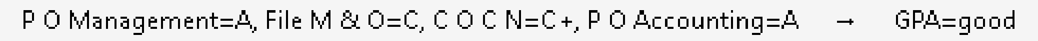
ARABIC 3=C+, E3=D, System A & D=A, P O Accounting=A → GPA=v.good

INTRO ST & P=A, ARABIC 3=C+, System A & D=A, P O Accounting=A → GPA=v.good

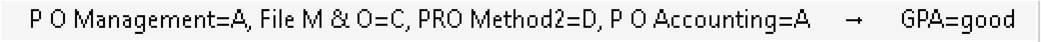


System A & D=A, P O Accounting=A → GPA=v.good

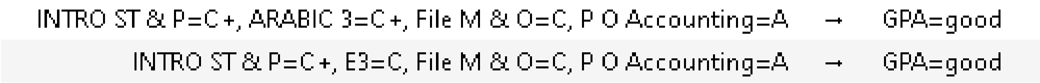
Rules of cluster number (3) in second year



P O Management=A, File M & O=C, C O C N=C+, P O Accounting=A → GPA=good

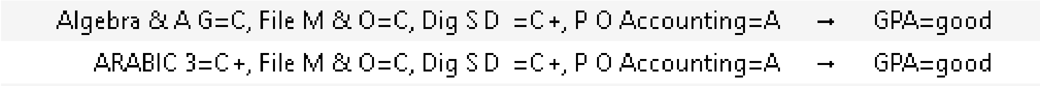


P O Management=A, File M & O=C, PRO Method2=D, P O Accounting=A → GPA=good



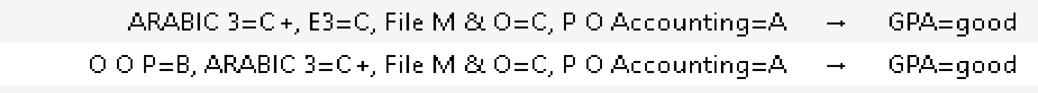
INTRO ST & P=C+, ARABIC 3=C+, File M & O=C, P O Accounting=A → GPA=good

INTRO ST & P=C+, E3=C, File M & O=C, P O Accounting=A → GPA=good



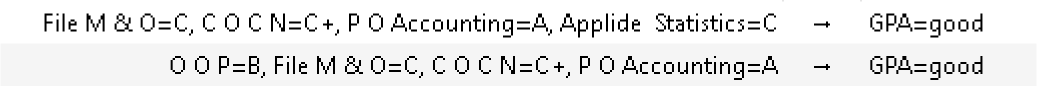
Algebra & A G=C, File M & O=C, Dig S D =C+, P O Accounting=A → GPA=good

ARABIC 3=C+, File M & O=C, Dig S D =C+, P O Accounting=A → GPA=good



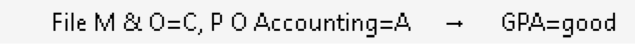
ARABIC 3=C+, E3=C, File M & O=C, P O Accounting=A → GPA=good

O O P=B, ARABIC 3=C+, File M & O=C, P O Accounting=A → GPA=good



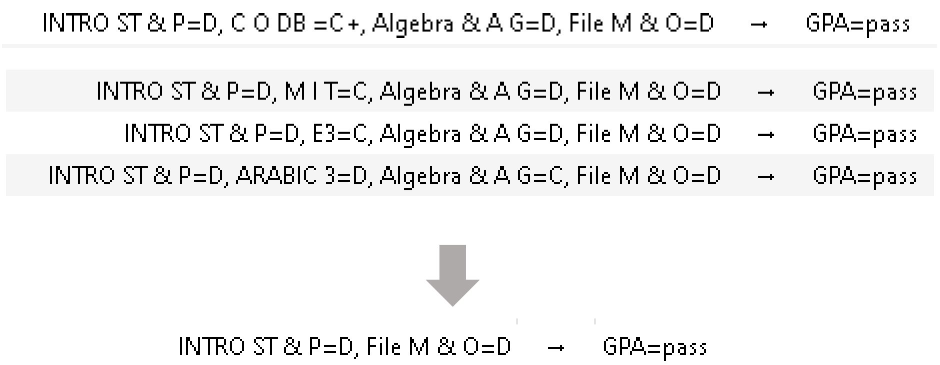
File M & O=C, C O C N=C+, P O Accounting=A, Applide Statistics=C → GPA=good

O O P=B, File M & O=C, C O C N=C+, P O Accounting=A → GPA=good



File M & O=C, P O Accounting=A → GPA=good

Rules of cluster number (4) in second year



INTRO ST & P=D, C O DB =C+, Algebra & A G=D, File M & O=D → GPA=pass

INTRO ST & P=D, M I T=C, Algebra & A G=D, File M & O=D → GPA=pass

INTRO ST & P=D, E3=C, Algebra & A G=D, File M & O=D → GPA=pass

INTRO ST & P=D, ARABIC 3=D, Algebra & A G=C, File M & O=D → GPA=pass

INTRO ST & P=D, File M & O=D → GPA=pass

-The above outcomes clearly indicate that Probability of the student obtaining the grade (v. good) if the student obtains the grade(A) in System Analysis and Design , grade(A) in Principles of Accounting and should at least pass all other courses..

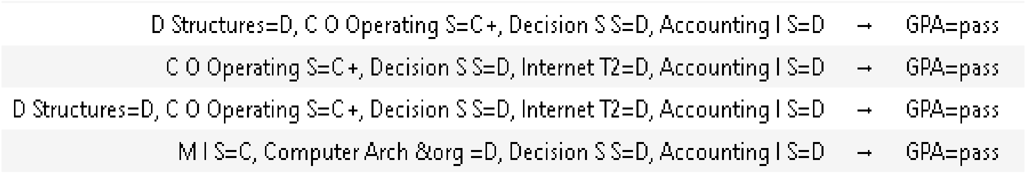
-Probability of the student obtaining the grade (good) if the student obtains the grade (A) in Principles of Accounting and grade (c) in File Management and Organization.

-Probability of the student obtaining the grade (pass) if the student obtains the grade (D) in Introduction to Statistics and grade (D) in File Management and Organization.

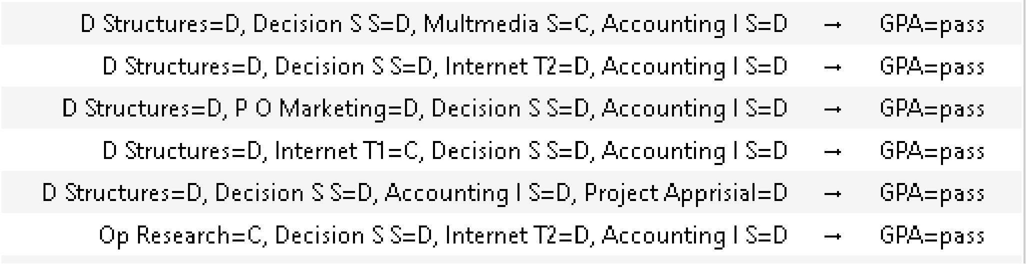
-Probability of the student obtaining the grade (Fail) if the student obtains the grade (F) in Programming Method II and grade (F) in applied statistics.

4.3.3 The third year rules

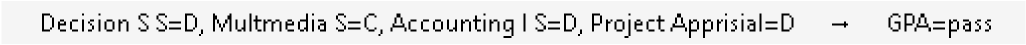
Rules of cluster number (1) in third year



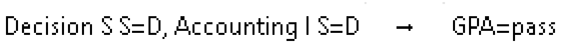
D Structures=D, C O Operating S=C+, Decision S S=D, Accounting I S=D	→	GPA=pass
C O Operating S=C+, Decision S S=D, Internet T2=D, Accounting I S=D	→	GPA=pass
D Structures=D, C O Operating S=C+, Decision S S=D, Internet T2=D, Accounting I S=D	→	GPA=pass
M I S=C, Computer Arch &org =D, Decision S S=D, Accounting I S=D	→	GPA=pass



D Structures=D, Decision S S=D, Multmedia S=C, Accounting I S=D	→	GPA=pass
D Structures=D, Decision S S=D, Internet T2=D, Accounting I S=D	→	GPA=pass
D Structures=D, P O Marketing=D, Decision S S=D, Accounting I S=D	→	GPA=pass
D Structures=D, Internet T1=C, Decision S S=D, Accounting I S=D	→	GPA=pass
D Structures=D, Decision S S=D, Accounting I S=D, Project Apprisial=D	→	GPA=pass
Op Research=C, Decision S S=D, Internet T2=D, Accounting I S=D	→	GPA=pass

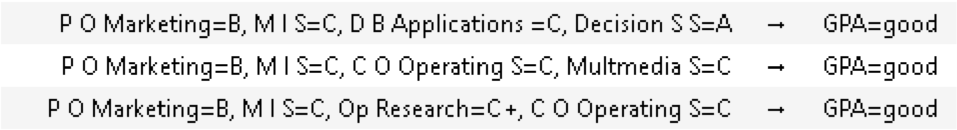


Decision S S=D, Multmedia S=C, Accounting I S=D, Project Apprisial=D	→	GPA=pass

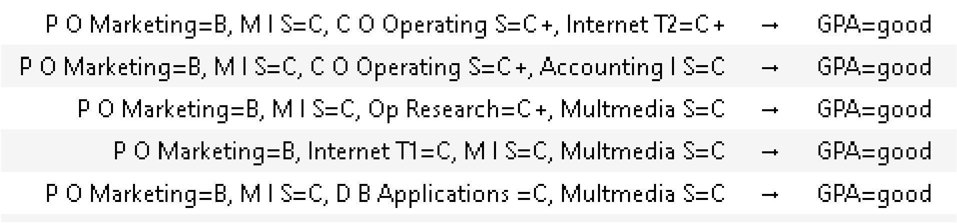


Decision S S=D, Accounting I S=D	→	GPA=pass

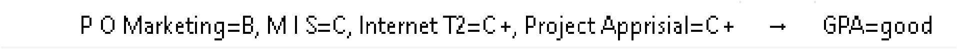
Rules of cluster number (2) in third year



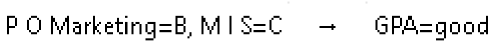
P O Marketing=B, M I S=C, D B Applications =C, Decision S S=A	→	GPA=good
P O Marketing=B, M I S=C, C O Operating S=C, Multmedia S=C	→	GPA=good
P O Marketing=B, M I S=C, Op Research=C+, C O Operating S=C	→	GPA=good



P O Marketing=B, M I S=C, C O Operating S=C+, Internet T2=C+	→	GPA=good
P O Marketing=B, M I S=C, C O Operating S=C+, Accounting I S=C	→	GPA=good
P O Marketing=B, M I S=C, Op Research=C+, Multmedia S=C	→	GPA=good
P O Marketing=B, Internet T1=C, M I S=C, Multmedia S=C	→	GPA=good
P O Marketing=B, M I S=C, D B Applications =C, Multmedia S=C	→	GPA=good

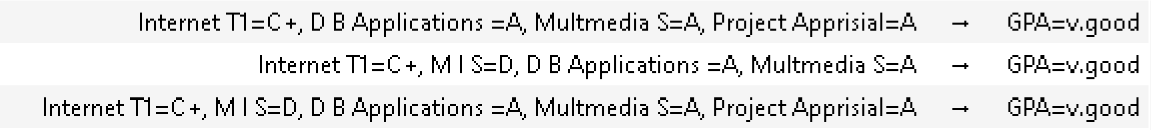


P O Marketing=B, M I S=C, Internet T2=C+, Project Apprisial=C+	→	GPA=good

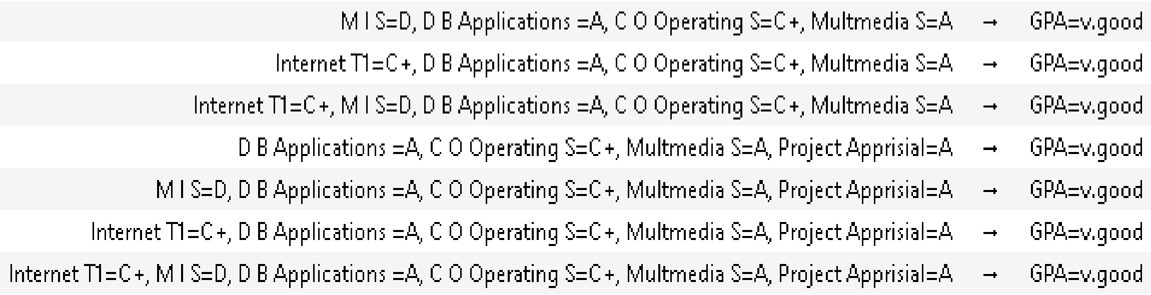


P O Marketing=B, M I S=C	→	GPA=good

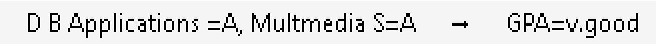
Rules of cluster number (3) in third year



Internet T1=C+, D B Applications =A, Multmedia S=A, Project Apprisial=A	→	GPA=v.good
Internet T1=C+, M I S=D, D B Applications =A, Multmedia S=A	→	GPA=v.good
Internet T1=C+, M I S=D, D B Applications =A, Multmedia S=A, Project Apprisial=A	→	GPA=v.good



M I S=D, D B Applications =A, C O Operating S=C+, Multmedia S=A	→	GPA=v.good
Internet T1=C+, D B Applications =A, C O Operating S=C+, Multmedia S=A	→	GPA=v.good
Internet T1=C+, M I S=D, D B Applications =A, C O Operating S=C+, Multmedia S=A	→	GPA=v.good
D B Applications =A, C O Operating S=C+, Multmedia S=A, Project Apprisial=A	→	GPA=v.good
M I S=D, D B Applications =A, C O Operating S=C+, Multmedia S=A, Project Apprisial=A	→	GPA=v.good
Internet T1=C+, D B Applications =A, C O Operating S=C+, Multmedia S=A, Project Apprisial=A	→	GPA=v.good
Internet T1=C+, M I S=D, D B Applications =A, C O Operating S=C+, Multmedia S=A, Project Apprisial=A	→	GPA=v.good



D B Applications =A, Multmedia S=A	→	GPA=v.good

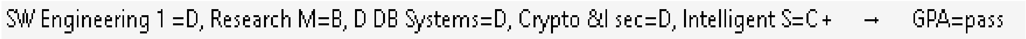
- The above outcomes clearly indicate that Probability of the student obtaining the grade (v. good) if the student obtains the grade(A) in Database Applications , grade(A) in Multimedia Systems and should at least pass all other courses..

-Probability of the student obtaining the grade (good) if the student obtains the grade(B) in Principles of Marketing and grade (c) in Management Information Systems.

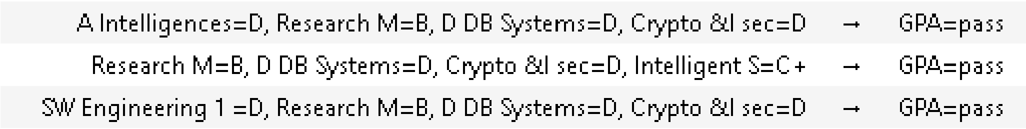
-Probability of the student obtaining the grade (pass) if the student obtains the grade(D) in Decision Support Systems and grade(D) in Accounting Information Systems.

4.3.4 The fourth year rules

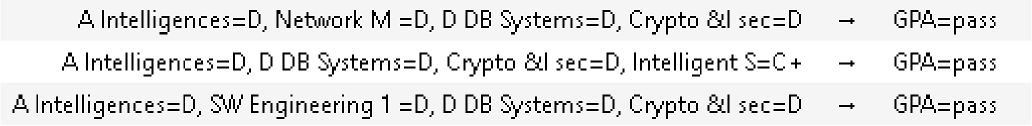
Rules of cluster number (1) in fourth year



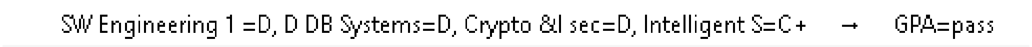
SW Engineering 1 =D, Research M=B, D DB Systems=D, Crypto &I sec=D, Intelligent S=C+	→	GPA=pass



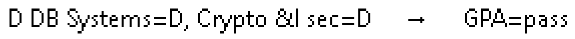
A Intelligences=D, Research M=B, D DB Systems=D, Crypto &I sec=D	→	GPA=pass
Research M=B, D DB Systems=D, Crypto &I sec=D, Intelligent S=C+	→	GPA=pass
SW Engineering 1 =D, Research M=B, D DB Systems=D, Crypto &I sec=D	→	GPA=pass



A Intelligences=D, Network M =D, D DB Systems=D, Crypto &I sec=D	→	GPA=pass
A Intelligences=D, D DB Systems=D, Crypto &I sec=D, Intelligent S=C+	→	GPA=pass
A Intelligences=D, SW Engineering 1 =D, D DB Systems=D, Crypto &I sec=D	→	GPA=pass

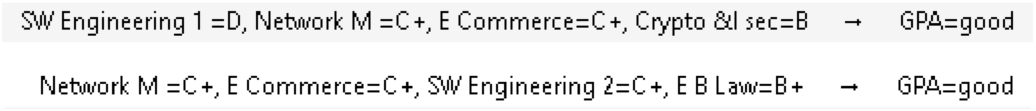


SW Engineering 1 =D, D DB Systems=D, Crypto &I sec=D, Intelligent S=C+	→	GPA=pass



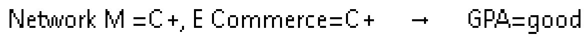
D DB Systems=D, Crypto &I sec=D → GPA=pass

Rules of cluster number (2) in fourth year



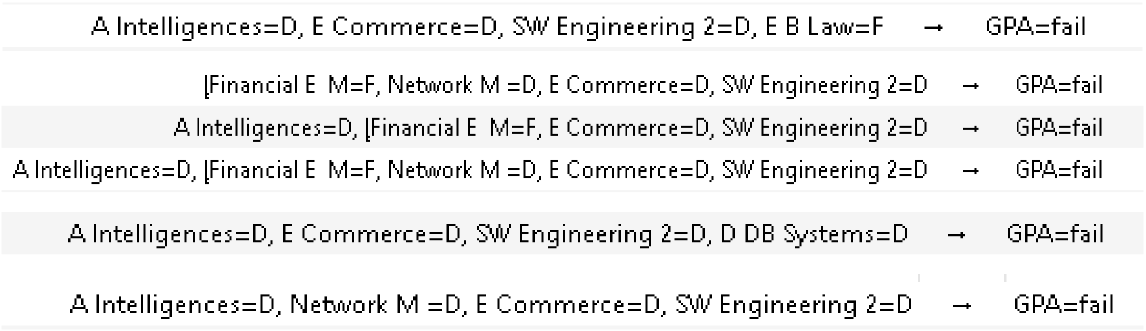
SW Engineering 1 =D, Network M =C+, E Commerce=C+, Crypto &I sec=B → GPA=good

Network M =C+, E Commerce=C+, SW Engineering 2=C+, E B Law=B+ → GPA=good



Network M =C+, E Commerce=C+ → GPA=good

Rules of cluster number (3) in fourth year



A Intelligences=D, E Commerce=D, SW Engineering 2=D, E B Law=F → GPA=fail

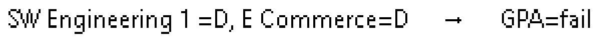
[Financial E M=F, Network M =D, E Commerce=D, SW Engineering 2=D → GPA=fail

A Intelligences=D, [Financial E M=F, E Commerce=D, SW Engineering 2=D → GPA=fail

A Intelligences=D, [Financial E M=F, Network M =D, E Commerce=D, SW Engineering 2=D → GPA=fail

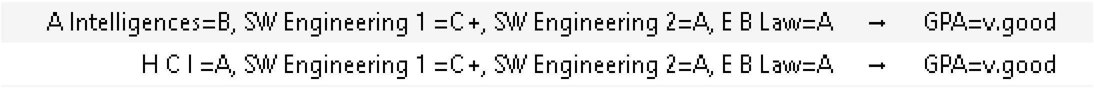
A Intelligences=D, E Commerce=D, SW Engineering 2=D, D DB Systems=D → GPA=fail

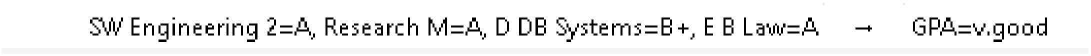
A Intelligences=D, Network M =D, E Commerce=D, SW Engineering 2=D → GPA=fail



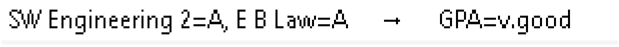
SW Engineering 1 =D, E Commerce=D → GPA=fail

Rules of cluster number (4) in fourth year


A Intelligences=B, SW Engineering 1 =C+, SW Engineering 2=A, E B Law=A → GPA=v.good
H C I =A, SW Engineering 1 =C+, SW Engineering 2=A, E B Law=A → GPA=v.good


SW Engineering 2=A, Research M=A, D DB Systems=B+, E B Law=A → GPA=v.good

↓


SW Engineering 2=A, E B Law=A → GPA=v.good

- The above outcomes clearly indicate that Probability of the student obtaining the grade (v. good) if the student obtains the grade(A) in Software Engineering II , grade(A) in Electronic Business Law and should at least pass all other courses..

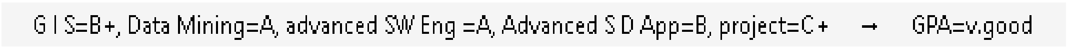
-Probability of the student obtaining the grade (good) if the student obtains the grade (c+) in Network Management and grade (c+) in Electronic commerce.

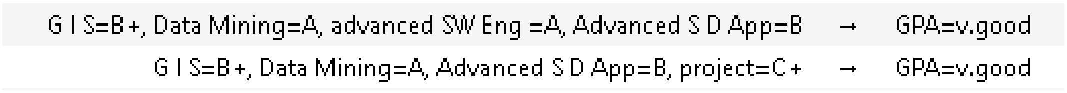
-Probability of the student obtaining the grade (pass) if the student obtains the grade (D) in Distributed Database Systems and grade(D) in cryptography and Information Security.

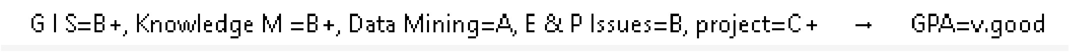
-Probability of the student obtaining the grade (Fail) if the student obtains the grade (D) in Software Engineering I and grade (D) in Electronic commerce.

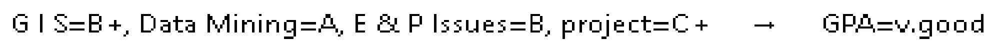
4.3.5 The fifth year rules

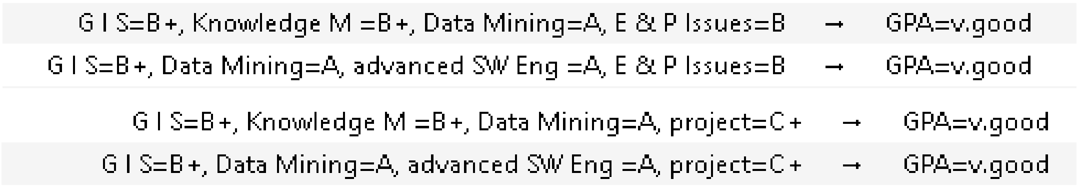
Rules of cluster number (1) in fifth year


G I S=B+, Data Mining=A, advanced SW Eng =A, Advanced S D App=B, project=C+ → GPA=v.good

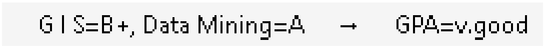

G I S=B+, Data Mining=A, advanced SW Eng =A, Advanced S D App=B → GPA=v.good
G I S=B+, Data Mining=A, Advanced S D App=B, project=C+ → GPA=v.good


G I S=B+, Knowledge M =B+, Data Mining=A, E & P Issues=B, project=C+ → GPA=v.good


G I S=B+, Data Mining=A, E & P Issues=B, project=C+ → GPA=v.good

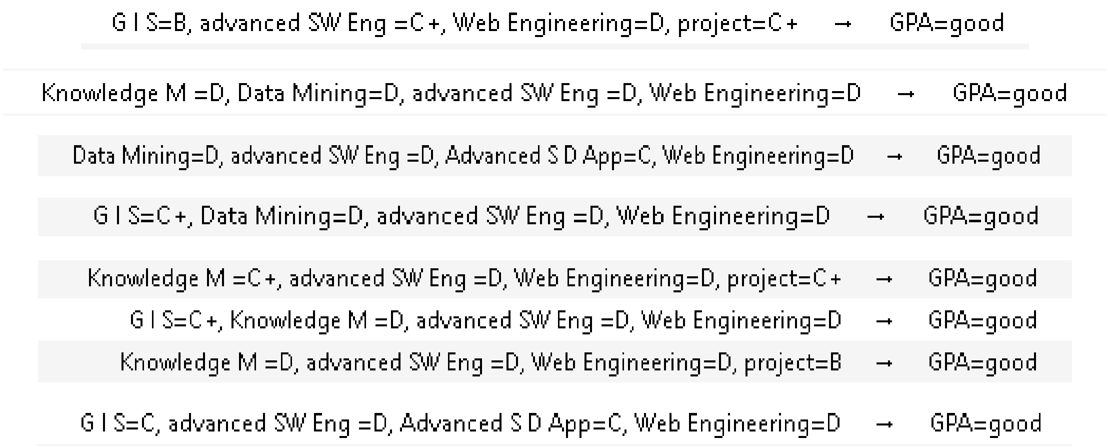


G I S=B+, Knowledge M =B+, Data Mining=A, E & P Issues=B → GPA=v.good
G I S=B+, Data Mining=A, advanced SW Eng =A, E & P Issues=B → GPA=v.good
G I S=B+, Knowledge M =B+, Data Mining=A, project=C+ → GPA=v.good
G I S=B+, Data Mining=A, advanced SW Eng =A, project=C+ → GPA=v.good

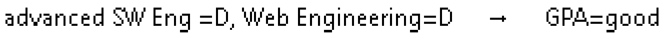


G I S=B+, Data Mining=A → GPA=v.good

Rules of cluster number (2) in fifth year

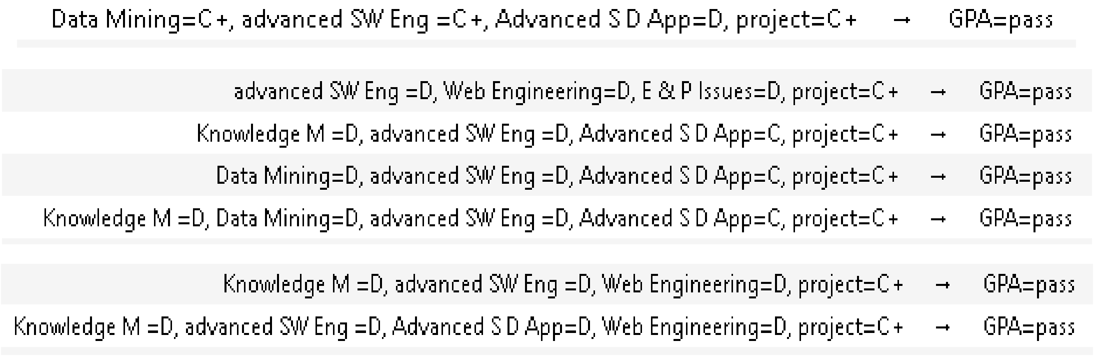


G I S=B, advanced SW Eng =C+, Web Engineering=D, project=C+ → GPA=good
Knowledge M =D, Data Mining=D, advanced SW Eng =D, Web Engineering=D → GPA=good
Data Mining=D, advanced SW Eng =D, Advanced S D App=C, Web Engineering=D → GPA=good
G I S=C+, Data Mining=D, advanced SW Eng =D, Web Engineering=D → GPA=good
Knowledge M =C+, advanced SW Eng =D, Web Engineering=D, project=C+ → GPA=good
G I S=C+, Knowledge M =D, advanced SW Eng =D, Web Engineering=D → GPA=good
Knowledge M =D, advanced SW Eng =D, Web Engineering=D, project=B → GPA=good
G I S=C, advanced SW Eng =D, Advanced S D App=C, Web Engineering=D → GPA=good

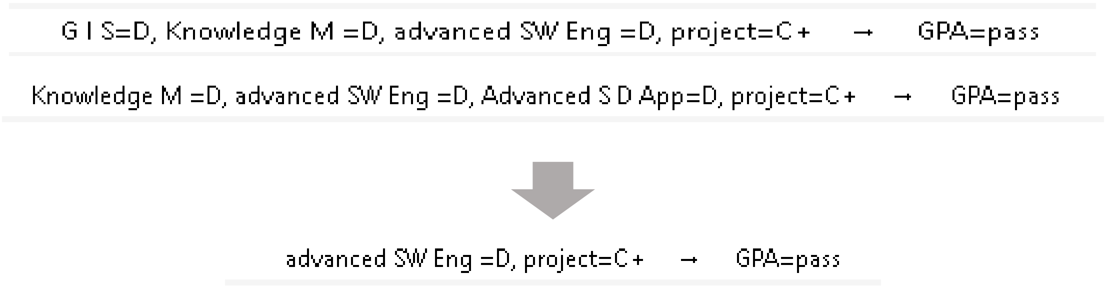


advanced SW Eng =D, Web Engineering=D → GPA=good

Rules of cluster number (3) in fifth year



Data Mining=C+, advanced SW Eng =C+, Advanced S D App=D, project=C+ → GPA=pass
advanced SW Eng =D, Web Engineering=D, E & P Issues=D, project=C+ → GPA=pass
Knowledge M =D, advanced SW Eng =D, Advanced S D App=C, project=C+ → GPA=pass
Data Mining=D, advanced SW Eng =D, Advanced S D App=C, project=C+ → GPA=pass
Knowledge M =D, Data Mining=D, advanced SW Eng =D, Advanced S D App=C, project=C+ → GPA=pass
Knowledge M =D, advanced SW Eng =D, Web Engineering=D, project=C+ → GPA=pass
Knowledge M =D, advanced SW Eng =D, Advanced S D App=D, Web Engineering=D, project=C+ → GPA=pass



G I S=D, Knowledge M =D, advanced SW Eng =D, project=C+ → GPA=pass

Knowledge M =D, advanced SW Eng =D, Advanced S D App=D, project=C+ → GPA=pass

advanced SW Eng =D, project=C+ → GPA=pass

- The above outcomes clearly indicate that Probability of the student obtaining the grade (v. good) if the student obtains the grade(A) Data mining , grade(B+) in Geographic Information System and should at least pass all other courses.

-Probability of the student obtaining the grade (good) if the student obtains the grade (D) in Advance Software Engineering and grade (D) in Web Engineering.

-Probability of the student obtaining the grade (pass) if the student obtains the grade(D) in Advance Software Engineering and grade(c+) in Project.

4.4 Discussion guidance rules that help student to improve their performance

The students before they join each year must consider the courses which effect in their performance positively or negatively

- **In the first year:** the students when they enrolled in the first year study should care to the courses of Essential Mathematics and Arabic Language I which effect positively in their performance.
  English language I and Programming Methods I which effect negatively in their performance.

- **In the second year:** the students before they join in the second year study should care to the courses of System Analysis and Design Principles of Accounting which effect positively in their performance. Programming Method II and Applied statistics which effect negatively in their performance.

- **In the third year:** the students before they join in the third year study should care to the courses of Database Applications and Multimedia Systems which effect positively in their performance. Decision Support Systems and Accounting Information Systems which effect negatively in their performance.
- **In fourth year:** the students before they join in the fourth year study should care to the courses of Software Engineering II and Electronic Business Law which effect positively in their performance. Software Engineering I and Electronic commerce which effect negatively in their performance.
- **In fifth year:** the students before they join in the fourth year study should care to the courses of Data mining and Geographic Information System which effect positively in their performance. Advance Software Engineering and Project which effect negatively in their performance.

4.5 Conclusion

The study applied on real data from university of Kassala faculty of computer sciences and information technology. Preparing the data. Then grouping student into similar classes using k means cluster. Then applied Association rules algorithms on each cluster which generated strong rules in each year study. Then discus the result of strong rule which help student to improve their performance

4.6 future works

There are aspects that the research did not address due to the lack of data such as data related to family income, type of admission and data of teaching staff and their impact on student performance recommend that they be taken into account in future studies

References

Ahmad, F., Ismail, N. H. and Aziz, A. A. (2015) 'The prediction of students' academic performance using classification data mining techniques', *Applied Mathematical Sciences*, 9(129), pp. 6415–6426.

Ahmed, A. and Elaraby, I. S. (2014) 'Data mining: A prediction for student's performance using classification method', *World Journal of Computer Application and Technology*, 2(2), pp. 43–47.

Al-shargabi, A. A. and Nusari, A. N. (2010) 'Discovering vital patterns from UST students data by applying data mining techniques', in *2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)*. IEEE, pp. 547–551.

Asif, R. *et al.* (2017) 'Analyzing undergraduate students' performance using educational data mining', *Computers & Education*. Elsevier, 113, pp. 177–194.

Borkar, S. and Rajeswari, K. (2013) 'Predicting students academic performance using education data mining', *International Journal of Computer Science and Mobile Computing*, 2(7), pp. 273–279.

Chen, H. (2018) 'Predicting student performance using data from an Auto-grading system'. University of Waterloo.

Golding, P., Facey-Shaw, L. and Tennant, V. (2006) 'Effects of peer tutoring, attitude and personality on academic performance of first year introductory programming students', in *Proceedings. Frontiers in Education. 36th Annual Conference*. IEEE, pp. 7–12.

Gulati, P. and Sharma, A. (2012) 'Educational data mining for improving educational quality', *Int. J. Comput. Sci. Inf. Technol. Secur*, 2(3), pp. 648–650.

Guleria, P. and Sood, M. (2014) 'Data mining in education: a review on the knowledge discovery perspective', *International Journal of Data Mining & Knowledge Management Process*. Academy & Industry Research Collaboration Center (AIRCC), 4(5), p. 47.

Kapur, B., Ahluwalia, N. and Sathyaraj, R. (2017) 'Comparative study on marks prediction using data mining and classification algorithms', *International Journal of Advanced Research in Computer Science*. International Journal of Advanced Research in Computer Science, 8(3).

Mining, O. D. (2016) 'Orange Visual Programming Documentation'.

orange (2018) ‘Orange Data Mining - Data Mining’. Available at: https://orange.biolab.si/.

Patil, P. S. (2017) ‘Predicting instructor performance using na {\"\i} ve bayes classification algorithm in data mining technique: A survey’, *International Journal of Advanced Electronics and Communication Systems*, 6.

Saa, A. A. (2016) ‘Educational data mining & students’ performance prediction’, *International Journal of Advanced Computer Science and Applications*, 7(5), pp. 212–220.

Sembiring, S. *et al.* (2011) ‘Prediction of student academic performance by an application of data mining techniques’, in *International Conference on Management and Artificial Intelligence IPEDR*, pp. 110–114.

Shovon, M. H. I. and Haque, M. (2012) ‘Prediction of student academic performance by an application of k-means clustering algorithm’, *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(7).

Talley, C. P. and Scherer, S. (2013) ‘The enhanced flipped classroom: Increasing academic performance with student-recorded lectures and practice testing in a" flipped" STEM course’, *The Journal of Negro Education*. JSTOR, 82(3), pp. 339–347.

Tegegne, A. K. and Alemu, T. A. (2018) ‘Educational data mining for students’ academic performance analysis in selected Ethiopian universities’, *Information Impact: Journal of Information and Knowledge Management*, 9(2), pp. 1–15.

Yadav, S. K. and Pal, S. (2012) ‘Data mining: A prediction for performance improvement of engineering students using classification’, *arXiv preprint arXiv:1203.3832*.

Zeynu, S. and Patil, S. (2018) ‘Prediction of Chronic Kidney Disease using Data Mining Feature Selection and Ensemble Method’, *International Journal of Data Mining in Genomics & Proteomics*, 9(1), pp. 1–9.