CHAPTER ONE
INTRODUCTION

1.1 Introduction:

The main objective of higher education institutes is to provide quality education to its student and to improve the quality of managerial decisions. One way to achieve highest level of quality in higher education system is by discovering knowledge from educational data to study the main attributes that may affect the student's performance. The discovered knowledge can be used to offer a helpful and constructive recommendations to the academic planners in higher education institutes to enhance their decision making process, to improve student's academic performance and trim down failure rate, to better understand student's behavior, to assist instructors, to improve teaching and many other benefits. There exists some evidence of students' background information that contributes immensely to the early prediction of student success. Though none of the studies directly shows how family background factors relate to student performance, is necessary to construct a model to capture students' success at the first-year level. Moreover, most researchers using prediction systems are doing this only for a particular course however, giving a good prediction for student academic performances for a particular course is no guarantee that it will give a good prediction for another course. Therefore, it is necessary to design a system that can give a global prediction at the end of each student semester. This prediction system not only helps to predict the grades of students; it also helps to prediction the appropriate socialization to the students by considering the result predict. It is therefore possible to employ scientific exploration of knowledge and mining of data in the evaluation process, which is a necessary process as an integral part of the process of development of universities and colleges affiliated to it and can be identified through the achievement of its objectives in raising the scientific level and achieve scientific rigor and evaluate the activities and work of those universities to Know the extent to which theoretical plans are transformed into reality(Jehad Mergani, 2014).

1.2 Problem statement:

There are different specializations in Faculty of Economics and Business Studies to distribute the students for them from second year, but the is no rule for distribution it's difficult to determine which specialization is better for the student.

1.3 Research Objectives:

The main aim objective of this research is to predict student's academic performance using data mining techniques to Design classification model to Predict the best student's specialization and using association rules to discovered knowledge and describe a close correlation between courses and specialization.

Other Objectives**:**

-collecting academic data of the students.

-processing and analyzing the data to get new results.

1.4 Importance of research:

The study, present application study in the field of Data Mining, to support decision maker, through extract some patterns which can be participate in processing educational development via technical applications Data Mining to improve the academic performance in University

1.5 Research Scope:

The research was conducted in University of Kordofan, Faculty of Economics and Business Studies, in North Kordofan State. Also applied to the academic data without financial and personal data between 2009 and 2016 it contains 1778 records.

1.6 Structure of the research

This research has five chapters, followed as references

Chapter 1: gives introduction, defining the problem, objectives, Importance of research and scope.

Chapter 2: discusses the literature review and related work, which contain represents general background about data mining and educational data mining, the related studies and techniques that used in educational data mining.

Chapter 3: explains the research methodology and the implementation of the techniques.

Chapter 4: presents the results and their discussion.

Finally, the conclusion of the research and the future work in Chapter 5.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction:

This chapter discusses the state of the studies in analyzing the students' performance using different techniques. The first part presents data mining techniques and especially that was implemented in the educational section. The second part presents the related work which analysis of academic performance approaches in the educational section**.**

The objective of data mining in each application area is different. For example, in business the main objective is to increase profit, which is tangible and can be measured in term of amounts of money, number of customers and customer loyalty. But educational data mining has both applied research objectives, such as improving the learning process and guiding students' learning; as well as pure research objectives, such as achieving a deeper understanding of educational phenomena. These goals are sometimes difficult to quantify and require their own special set of measurement techniques. In educational environments there are many different types of data available for mining. These data are specific to the educational area and which have intrinsic semantic information, relationships with other data and multiple levels of meaningful hierarchy(Jehad Mergani, 2014).

2.2 Data Mining(DM):

Automated data collection tools and mature database technology help Individuals organizations to collect and produce huge volume of data every day. These vast volumes of data lead to the Data explosion problem. The major challenge facing organizations Discovering useful knowledge from the database and transforming information into actionable results.DM (knowledge discovery from data) is an automatic or semi-automatic process of extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data (Kamber, 2006)

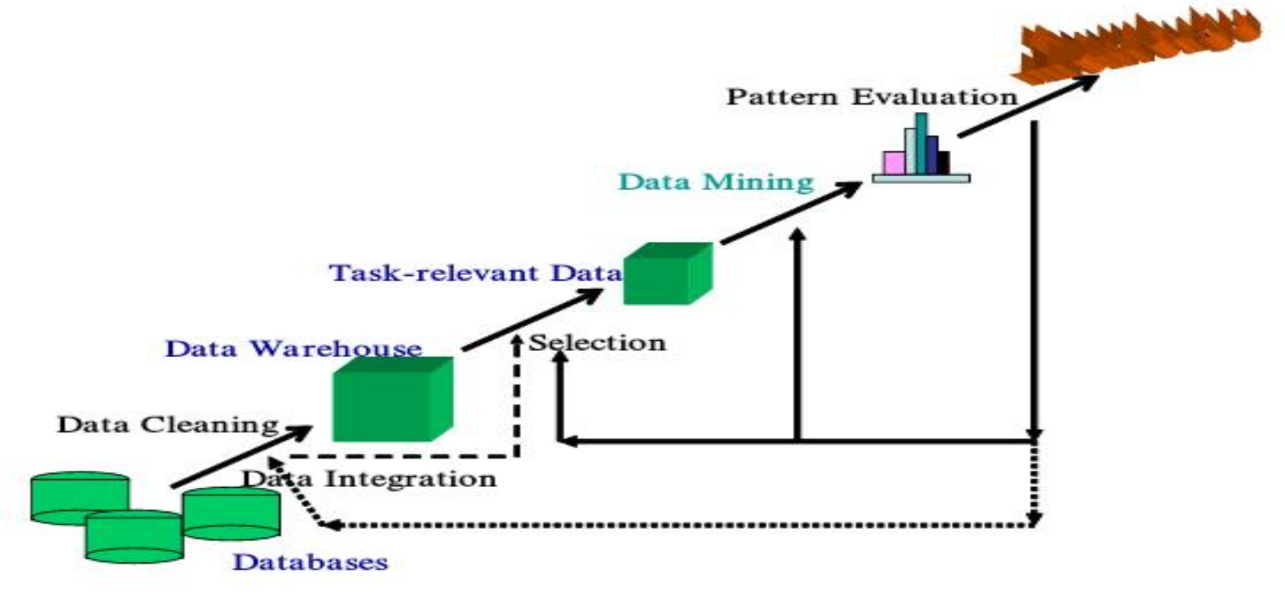


Figure (2.1): Steps of Data Mining Process

2.2.1 Data Mining Tasks:

Data mining tasks can be classified into two categories:

- Predictive mining tasks: perform inference on the current data in order to make predictions. Classification, regression, deviation detection are examples of predictive tasks(Jehad Mergani, 2014)
- Descriptive mining tasks: characterize the general properties of the data in the database. Clustering, association rule discovery, and sequential pattern discovery are examples of descriptive tasks (Jehad Mergani, 2014).

2.3 DM techniques:

2.3.1 Classification:

Classification is the process of finding a model (or function) that describes and distinguishes data classes or concepts. The model is derived based on the analysis of a set of training data. The model is used to predict the class label of objects for which the class label is unknown (S.Venkata Krishna Kumar, 2015). various classification methods were used in their studies to predict students' academic performance such as naive bayes, j48 and Random Forest these classifiers selected because they are most famous classifiers used in other researches.

2.3.2 Decision Tree:

A decision tree can be a flow chart resembling a tree structure, where every internal node is denoted by rectangles and the leaf nodes are denoted by ovals. This is often used algorithm because of easy implementation and easier to understand compared to different classification algorithms. Decision tree starts with a root node that helps the users to take required actions. From this node, users split every node recursively according to decision tree learning algorithm. The ultimate result is a decision tree in which each branch represents an outcome (Pant, 2017).

2.3.3 C4.5 (J48):

This algorithm can be a successor to ID3 developed by Quinlan Ross. It is additionally supported the Hunt's algorithm.C4.5 handles each categorical and continuous attributes to create a decision tree, so as to handle continuous attributes. C4.5 splits the attribute values into 2 partitions based on the chosen threshold. It additionally handles missing attribute values. C4.5 has the concept of Gain Ratio as an attribute selection measure to create a decision tree. It prunes the biasness of information gain once there are many outcome values of an attribute. At first, calculate the gain ratio of every attribute. The root nodes are the attribute whose gain ratio is a maximum. C4.5 uses pessimistic pruning to get rid of unessential branches with in the decision tree to enhance the accuracy of classification (Sujith Jayaprakash, 2018).

2.3.3.1 Random Forest:

Random Forests is a bagging tool that leverages the ability of multiple varied analyses, organization strategies, and ensemble learning to supply correct models, perceptive variable importance ranking, and laser-sharp coverage on a record-by-record basis for deep data understanding. Its strengths are recognizing outliers and anomalies in knowledgeable data, displaying proximity clusters, predicting future outcomes, characteristic necessary predictions, discovering data patterns, exchange missing values with imputations, and providing perceptive graphics (Sujith Jayaprakash, 2018).

2.3.3.2 Naive Bayes classifier:

The Bayesian Classification represents a supervised learning method as well as a statistical method for classification. Assumes an underlying probabilistic model and it allows us to capture uncertainty about the model in a principled way by determining probabilities of the outcomes. It can solve predictive problems. This Classification is named after Thomas Bayes, who proposed the Bayes Theorem. Bayesian classification provides practical learning algorithms and prior knowledge and observed data can be combined. Bayesian Classification provides a useful perspective for understanding and evaluating many learning algorithms. It calculates explicit probabilities for hypothesis and it is robust to noise in input data**.** Naive Bayes is a classification technique based on Bayes Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as 'Naïve' (1Y Divyabharathi, 2018).

Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods. Bayes theorem provides a way of calculating posterior probability P (c|x) from P(c), P(x) and P (x|c). Look at the equation below:

$$\underbrace{P(c \mid x)}_{\text{Posterior Probability}} = \frac{\overbrace{P(x \mid c)}^{\text{Likelihood}}\,\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Where:

- P (c|x) is the posterior probability of class (c, target) given predictor (x, attributes).
- P(c) is the prior probability of class.
- P (x|c) is the likelihood which is the probability of predictor given class.
- P(x) is the prior probability of predictor.

2.3.3.3 Neural Network:

Multilayer Perception (MLP) algorithm is one of the most widely used and common neural networks. Multilayer Perception (MLP) is a feed forward artificial neural network model that maps sets of input data onto a collection of acceptable output. An MLP consists of multiple layers of nodes in an exceedingly directed graph, with every layer totally connected to the consequent one. Their current output depends solely on the present input instance. It trains victimization back propagation (Sujith Jayaprakash, 2018).

2.3.4 Clustering:

Clustering is a collection of similar data object. Dissimilar object is another cluster. It is way finding similarities between data according to their characteristic. This technique based on the unsupervised learning. For example, image processing, pattern recognition and city are planning (Patil, 2015). Clustering and classification both, is a basic function in data mining. Classification is used mainly as a way of learning under the supervision block to learn uncensored. The objective of the meeting is descriptive; this classification is predictive. Because the goal of the meeting is to find a new set of new categories and groups are important in themselves, and their evaluation is essential. As we mentioned before, classification can be taken as supervised learning process, clustering is another mining technique similar to classification. However, clustering is an unsupervised learning process. Clustering is the process of grouping a set of physical or abstract objects into classes of similar objects, so that objects within the same cluster must be similar to some extent, also they should be dissimilar to those objects in other clusters. In classification which record belongs which class is predefined, while in clustering there is no predefined classes. In clustering, objects are grouped together based on their similarities. Similarity between objects is defined by similarity functions; usually similarities are quantitatively specified as distance or other measures by corresponding domain experts. Most clustering applications are used in market segmentation. By clustering their customers are divided into different groups, business organizations can provide different personalized services to different group of markets. For example, based on the expense, deposit and draw patterns of the customers, a bank can be clustering the market into different groups of people. For different groups of market, the bank can provide different kinds of loans for houses or cars with different budget plans (HAN, 2011)

2.3.5 Association Rule Mining:

Association and correlation is usually to find frequent item set findings among large data sets. This type of finding helps businesses to make certain decisions, such as catalogue design, cross marketing and customer shopping behavior analysis. Association Rule algorithms need to be able to generate rules with confidence values less than one. However, the number of possible Association Rules for a given dataset is generally very large and a high proportion of the rules are usually of little (if any) value(B.SANTHOSH KUMAR, 2010).

2.3.5.1 Apriori Algorithm:

In computer science and data mining, Apriori is a classic algorithm for learning association rules. Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers, or details of a website frequentation). Apriori uses breadth-first search and a tree structure to count candidate item sets efficiently. It generates candidate item sets of length k from item sets of length k - 1. Then it prunes the candidates which have an infrequent sub pattern. According to the downward closure lemma, the candidate set contains all frequent k-length item sets. After that, it scans the transaction database to determine frequent item sets among the candidates (B.SANTHOSH KUMAR, 2010).

The key concepts in this algorithm are:

- Frequent Item sets: The sets of item which has minimum support (denoted by Liforith-Item set).
- Apriori Property: Any subset of frequent item set must be frequent.
- Join Operation: To find Lk, a set of candidate k-item sets is generated by joining Lk-1with itself.

The advantages of using Apriori algorithm are:

- Uses large item set property.
- Easily parallelized.
- Easy to implement.

The Apriori algorithm is an efficient algorithm for finding all frequent item sets. It implements level-wise search using frequent item property and can be additionally optimized. The Apriori algorithm used is given below

- Lk: Set of frequent item sets of size k (with min support).
- Ck: Set of candidate item set of size k (potentially frequent item sets).

L1= {frequent items};

For (k= 1; Lk! =∅; k++) **do**

Ck+1= candidates generated from Lk;

For each transaction tin database **do**

Increment the count of all candidate's in

Ck+1that are contained in t

Lk+1= candidates in Ck+1with min_support

Return ∪kLk;

2.3.6 Regression:

Regression is used to map a data item to a real valued prediction variable. In other words, regression can be adapted for prediction. In the regression techniques target value are known. For example, you can predict the child behavior based on family history (HAN, 2011).

2.4 Educational Data Mining (EDM):

Educational Data Mining (EDM) is concerned with developing methods and analyzing educational content to enable better understanding of students' performance (HAN, 2011). It is also important to enhance teaching and learning process. The data can be collected form historical and operational data reside in the databases of educational institutes. The student's data can be personal or academic. Also it can be collected from e learning systems which have a large amount of information used by most institutes. Prediction models that include all personal, social, psychological and other environmental variables are necessitated for the effective prediction of the performance of the students (HAN, 2011). The prediction of student performance with high accuracy is beneficial for identify the students with low academic achievements initially. It is required that the identified students can be assisted more by the teacher so that their performance is improved in future. So the Educational Data Mining (EDM) have different ways that is defined , most notably:(defined educational data mining as “Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings and using those methods to better understand students and the settings which they learn in EDM is emerging discipline with a suite of computational and psychological methods and research approaches to understanding how students learn and setting which they learn in it (Bassil, 2012). Educational data mining (EDM) is an emerging discipline which focuses applying data mining tools and techniques to educationally related data (Norhidayah Mohamad Yatima, 2015). An educational data mining is a broader term that focuses on nearly any type of data in educational institutional, while academic analytics is specific to data related institutional effectiveness and student retention issues. The scope of educational data mining includes areas that directly impact students. Other areas within EDM include analysis of educational processes including admissions, alumni relations and course selections (Norhidayah Mohamad Yatima, 2015). Also defined as academic analytics as the use of statistical techniques and data mining in ways that will help faculty and advisors become more proactive in identifying the position of student (Joshua D. Fostera, 2007), as well as the Educational data mining (EDM) deals with the developing and applying the computerized methods to detect patterns in large amount of educational data, that would be impossible to analyze. The main objective of any higher educational system is to improve the quality of education. Educational data mining has some advantages over the higher educational system such as decreasing student’s drop-out rate, increasing student’s promotion rate, increasing student’s retention rate, increasing student’s transition rate, increasing educational improvement ratio, increasing student’s learning outcome, maximizing educational system efficiency and reducing the cost of system processes. To achieve these goals, the data mining system will be helpful to put insights for decision makers in the higher educational system(Norhidayah Mohamad Yatima, 2015).

2.5 Related Works:

Data mining can be used in the educational field to enhance the understanding of the learning process to focus on identifying, extracting and evaluating variables related to the learning process of students. Some researchers used data mining in educational data to improve the learning process and enhance the academic performance of the students as illustrated the flowing**:**

(Al-Razgan, Mashael A. Al-Barrak and Muna, 2016)used j48 algorithm for classification. Transcripts data for female students who graduated from Computer Sciences College at King Saud University in the year 2012, the total number of students was236 students. The researcher focus on the mandatory courses offered by the program because of their domination in the study plan, which in turn have a greater effect on the final graduation grade. The result of the experimentation showed Software Engineering-1is the most important course and it is closely related to students' final GPA that if a student received an A+ in Software Engineering-1, she would graduate with an Excellent GPA.

(Kabakchieva, Dorina, 2012) predict Student Performance based on decision tree algorithm C4.5 (J48), a neural network (Multilayer Perceptron), and a Nearest Neighbor algorithm (IBK). These classification algorithms are selected because they are very often used for research purposes and have potential to yield good results. Moreover, they use different approaches for generating the classification models, which increases the chances for finding a prediction model with high classification accuracy. They have used one of famous and prestigious Bulgarian universities dataset; there are 10330 students that have been enrolled as university students during the period between 2007 and 2009, after preprocessing the final dataset, on which the selected classification data mining algorithms are applied, contains 10067 instances and 14 attributes, each time an algorithm is run, 2/3 of the data is used for training of the classification model and 1/3 of the data is used for testing and evaluation of the model. The result show that the classification models, generated by applying the selected four data mining algorithms – OneR Rule Learner, Decision Tree, Neural Network and K-Nearest Neighbour, on the available and carefully preprocessed student data, reveal classification accuracy between 67.46% and 73.59%. The highest accuracy is achieved for the Neural Network model (73.59%), followed by the Decision Tree model (72.74%) and the k-NN model (70.49%). The Neural Network model predicts with higher accuracy the “Strong” class, while the other three models perform better for the “Weak” class.

(V.Ramesh, 2013)it was collected data from higher secondary students. Nine schools were randomly selected from Kanchipuram district. The primary data was collected using a questionnaire which includes questions related to several personal, socioeconomic, psychological and school related variables that were expected to affect student performance. The questionnaire was reviewed by the professionals and tested on a small set of 45 students in order to get a feedback. The final version contained 50 questions and it was answered by more than 900 students. Latter a sample of 500 were selected from the whole. All 500 questionnaires were filled with the response rate of 100% out of which 316 were females and 184 were males. The researchers used five different classification algorithms: Naive Bayes, Multi-Layer Perception, SMO, J48, REPTree. The correctly classified instances with different algorithms were compared. The overall accuracy of classifiers' performance on our dataset is shown in the Table (2.1)

	Naïve Bayes	**MLP**	**SOM**	**J48**	**REPTREE**
Accuracy	49.5%	72.38%	57.25	64.88%	60.13%

Table (2.1): comparisons of algorithms

The evaluation results of this study, it was proven that Multi-Layer Perception (MLP) classifier is most appropriate for predicting student performance. MLP gives 72.38% prediction which is relatively higher than other algorithms.

(Rohaila Abdul Razak1, 2018)used linear regression and j48, Linear regression is a statistical method that can be used to summarize and identify the relationship between two (2) or more variables, the main purpose of linear regression analysis is to find out the associations between dependent variable and independent variables when the dependent and independent variables relationships are almost linear. In this study, there were 257 data set taken from the student of semester 6 which involved with 4 programs – Diploma in Computer System and Networking, Diploma in Information Technology, Diploma in Business Management and Diploma in Accountancy from Campus Management System (CMS), KPTM AlorSetar, there are 11 (eleven) attributes chose such as Gender, Program, State, Sponsorship, SPM, CGPA, GPA Sem 1, Sem 2, Sem 3, Sem 4, and Sem 5.The prediction accuracy in linear regression is 96.2% which indicates that the predictors has a strong correlation with the dependent variable, the J48 algorithm with 10-fold cross validation approached and has shown that the prediction accuracy is 82.5%. It indicates that the data set are classified as positive value. There is a strong significant relation between the GPASem1 and CGPA. The result shows that Linear Regression has a higher prediction accuracy compared to J48 Decision Tree. The analysis also shows factors contribute most to the performance of the students.

(Kalpesh Adhatrao, 2013)applied the ID3 and C4.5 algorithms on the training data to obtain decision trees of both the algorithms, information about students currently enrolled in the first year of engineering, was applied to the decision trees, the dataset contain 182 instance. The result of experimentation shows that the ID3 and C4.5 achieved both accuracy (75.145%).

(Mohammed M. Abu Tair, 2012)the dataset collected from the college of Science and Technology – Khanyounis for a period of fifteen years in period from 1993 to 2007. The graduate student's data set consists of 3360 records and 18 attribute. The researcher used four data mining tasks; Association, classification, clustering and outlier detection. The result shows that the association rules and sorted the rules using lift metric, and used two classification methods which are Rule Induction and Naïve Bayesian classifier to predict the Grade of the graduate student. Also the researcher used clustered the students into groups using K-Means clustering algorithm. Finally, used outlier detection to detect all outliers in the data, two outlier methods are used which are Distance-based Approach and Density-Based Approach. Each one of these tasks can be used to improve the performance of graduate student.

(Humera Shaziya, 2015)used naïve bayes classifier to predict student's performance; data was data chosen for the prediction of students‟ results are MCA students‟ data set consisting of the eight attributes. The results of experimentation show that the prediction of students‟ performance in their semester exams can be done by using their previous semester marks and their overall performance in various activities of the current semester. This helps all the stakeholders of the educational system to take necessary actions to improve the results.

(Suchita Borkar, 2013)predicted Students Academic Performance applied on data of Master of Computer Application (MCA) degree from Pune University based on association rule, the Variables used for judging the students' performance in university results are Graduation%, Attendance%, Assignment%, Unit Test% and University Result%. The analysis revealed after used apriori algorithm that student's university performance is dependent on Unit test, Assignment, Attendance and graduation percentage. The results reveal that the student's performance level can be improved in university result by identifying students who are poor unit Test, Attendance, Assignment and graduation and giving them additional guidance to improve the university result.

(Angeline, 2013)used Apriori Algorithm to analyze Student Performance based on collected department of Computer Science Dr. G.U.Pope College of Engineering dataset's in 2011 to 2012.

The initial data contains the details gathered from a number of 21 students with 15 attributes. The accuracy of rules was attained according to the value of confidence value. The number of rules generated was 127 with confidence 100% and support 38.095%. Since confidence gets a value of 100 % the rule is an exact rule. The running time for the application using apriori algorithm is 15ms. From the generated rules the students were categorized into good, average and poor. The extracted rules help to predict the performance of the students and it identify the average, below average and good students. The performance report of the student also helps to improve the result of the student.

(Sundar, 2013)compared between four classifiers: Naive Bayes Updateable, Hidden Naive Bayes, WAODE and AODEsr to predict student’s academic performance used First Year students of MCA Hindusthan college of Arts & Science- Coimbatore in the period of 2012-2013Intially student dataset contains 48 records and 10 Attribute. The result of experimentation show that the AODEsr algorithm has provides high overall accuracy rate than other algorithms (more than 64%).This study helped the students improve their performance and also it helps teacher to identify those students which needs a special attention to reduce failing ration and taking appropriate action at right time.

(Edin Osmanbegović, 2012)used three supervised data mining algorithms (J48, naïve Bayes and neural network) to predicted student performance. Data was collected from the surveys conducted during the summer semester at the University of Tuzla, the Faculty of Economics, academic year 2010-2011, among first year students and the data taken during the enrollment 257 students, the success was evaluated with the passing grade at the exam. The results indicate that the Naïve Bayes classifier outperforms in prediction decision tree and neural network methods.

(Kabakchieva, 2013)used different classifiers: decision tree algorithm C4.5 (J48), two Bayesian classifiers (Naive Bayes and Bayes Net), a Nearest Neighbour algorithm (IBk) and two rule learners (OneR and JRip).The final dataset used for the project implementation contained 10330 instances (539 in the “excellent” category, 4336 in the “very good” category, 4543 in the “good” category, 347 in the “average” category, and 564 in the “bad” category), each described with 14 attributes (1 output and 13 input variables), nominal and numeric (for the time period between 2007 and 2009).The achieved results: The J48 classifier classifies correctly about 2/3 of the instances (65.94 % for the 10-fold cross-validation testing and 66.59 % for the percentage split testing), produces a classification tree with a size of 1173 nodes and 1080 leaves. The attribute

Number of Failures appears at the first level of the tree, the Admission Score and Current Semester attributes appear at the second and third levels of the tree, the attributes University Specialty/Direction and Gender – at the third level of the tree, which means that these attributes influence most the classification of the instances into the five classes. The Bayesian classifies is about (but below) 60 % which is not very high, and it is worse compared to the performance of the decision tree classifier (66-67 %).The k-NN classifier accuracy is about 60 % and varies in accordance with the selected value for k.TheJRip rule learner performs better than the OneR rule learner, the overall accuracy of the JRip classifier is about 63 %, and for the OneR classifier it is about 54-55 %.The achieved results reveal that the decision tree classifier (J48) performs best (with the highest overall accuracy), followed by the rule learner (JRip) and the KNN classifier. The Bayes classifiers are less accurate than the others.

(Wan Fairos Wan Yaacob, 2019)used several predictive modelling techniques of K-Nearest Neighbor, Naïve Bayes, Decision Tree and Logistic Regression Model models were used to predict student's performance, the data of undergraduate students of Bachelor of Science in Statistics programmer has been selected from Faculty of Computer and Mathematical Sciences at Universities Teknologi MARA Cawangan Kelantan and Universities Teknologi MARA Cawangan Negeri Sembilan, collected 631 transcripts from year 2013 to 2016 for students who have completed their academic degrees. The result of experimentation shows that the Naïve Bayes classifier perform best (with the highest overall accuracy) followed by LR, K-NN, DT Information Gain and DT GINI. All classifier tested are performing with overall accuracy above 80% which means the error rate is low and predictions are reliable.

(C. Anuradha, 2015)compared between J48 decision tree, Naive Bayes Classifiers, k-Nearest Neighbours algorithm, OneR and JRip algorithm27, 28. A student's dataset was created based on the demographic and pre-collegiate characteristics of the students along with their performance in the class and university examinations. The author used for training and 1 fold is used for testing) and percentage split (2/3 of the dataset used for training and 1/3 – for testing). The result of experimentation show that the Bayesian classifiers outperforms the other (more than 70%).

(Bichkar, 2011)used Decision Trees to Predicted. The author used Data set of 346 students of the institute is collected who appeared for the first year of engineering in the year 2009-10, 2010-11. The data was collected through the enrolment form filled by the student at the time of admission.

The result show that the accuracy of this model is 69.94 %, that is out of 346 instances 242 are correctly classified.

(Surjeet Kumar Yadav, 2012)used to classification ID3, C4.5 and CART. Under the "Test options", the 10-fold cross validation is selected to evaluation approach. The researcher used the data set from VBS Purvanchal University, Jaunpur (Uttar Pradesh) on the sampling method for Institute of Engineering and Technology for session 2010. Initially size of the data is 90. The result of classifiers accuracy it was cleared that the true positive rate of the model for the FAIL class is 0.786 for ID3 and C4.5 decision trees that means model is successfully identifying the students who are likely to fail. These students can be considered for proper counseling so as to improve their result.

(Brijesh Kumar Baradwaj, 2011)used decision tree to Analyze Students‟ Performance. The data set used obtained from VBS Purvanchal University, Jaunpur (Uttar Pradesh) on the sampling method of computer Applications department of course MCA (Master of Computer Applications) from session 2007 to 2010. Initially size of the data is 50. The result shows that the classification rules can be generated for each path from each terminal node to root node. Pruning technique was executed by removing nodes with less than desired number of objects.

(Hafez Mousa1, 2017)used students' performance prediction model based on DM classification algorithms (Naïve Bayes, Decision Tree and K-NN). The researcher collected the dataset from a preparatory male schooling Gaza strip, includes (1036 records) and (9 features), (7th, 8th, 9th grade) in scholastic year (2014-2015). The result experimentation shows that when author used all of data features (academic and social features) the Decision Tree (DT) outperforms high accuracy (92.96%).

(Abeer Badr El Din Ahmed1, 2014)used ID3 Decision Tree. The data set used in this study was obtained from a student's database used in one of the educational institutions, on the sampling method of Information system department from session 2005 to 2010. Initially size of the data is 1547 records. The result was helped x to improve the student's performance, to identify those students which needed special attention to reduce failing ration and taking appropriate action at right time.

(Annisa Uswatun Khasanah, 2018)compared between Decision Tree and Bayesian Network for Predicted Student’s Performance. The dataset collected from student data base that can be accessed from Universities Islam Indonesia’s information system (UNISYS). The initial data consisted of 178 data of student in academic year 2007 and after cleaned the data only 104 data set available with 12 attributes. After three experiments the result shows that, the Bayesian Network achieved high accuracy more than Decision Tree (98%08).

Table (2.2): Summary of Related Work

Author	**Techniques**	**Details of the data**	**Results**
(Mashael A. Al-Barrak and Muna Al-Razgan, 2016)	Decision Trees (j48)	Size of data 236 records	if a student received an A+ in Software Engineering-1, she would graduate with an Excellent GPA
(DorinaKabakchieva, 2012)	decision tree (J48), a neural network (Multilayer Perceptron), and a Nearest Neighbor algorithm (IBk)	Size of data 10067 records	The Neural Network model predicts with higher accuracy the “Strong” class, while the other three models perform better for the “Weak” class
(K.Ramar,P.Parkavi ,V.Ramesh, 2013)	Naive Bayes, Multi-Layer Perception, SMO, J48, REPTree	Size of data 500 records	Multi-Layer Perception (MLP) gives 72.38% prediction, MLP classifier is most appropriate for predicting student performance
(Rohaila Abdul Razak,MazniOmar, Mazida Ahmad, 2018).	linear regression,j48	Size of data 257 records	The Linear Regression has a higher prediction accuracy compared to J48 Decision Tree the prediction accuracy is 96.2% , There is a strong

			significant relation between the GPASem1 and CGPA
(KalpeshAdhatrao, AdityaGaykarAdityaGaykar,RohitJha ,Vipul Honrao,2013)	ID3, decision trees (C4.5)	Size of data 182 records	The ID3 and C4.5 achieved both accuracy (75.145%)
(Mohammed M. Abu Tair,Alaa M. El-Halees, 2012)	Association, classification, clustering and outlier detection	Size of data 3360 records	Each one of these tasks can be used to improve the performance of graduate student
(HumeraShaziya,RaniahZaheer,G.Kavitha, 2015)	Naïve bayes Classifier	-	prediction of students‟ performance in their semester exams can be done by using their previous semester marks and their overall performance in various activities of the current semester
(SuchitaBorkar,K. Rajeswari, 2013)	association rule(apriori algorithm)	-	the student's performance level can be improved in university result by identifying students who are poor unit(Test, Attendance, Assignment and graduation) and giving them additional guidance to improve the university result
(D. Magdalene	association	Size of	The extracted rules helps to

Delighta Angeline, 2013)	rule(apriori algorithm)	data 21 records	predict the performance of the students and it identify the average, below average and good students
(P.V.Praveen Sundar,2013)	Naive Bayes Updateable,Hidden Naive Bayes,WAODEandAODEsr	Size of data 48 records	AODEsr algorithm has provides high overall accuracy rate than other algorithms (more than 64%), the result helped students improved their performance and helped teacher to identify those students which needs a special attention to reduce failing ration and taking appropriate action at right time
(Mirza Suljić.2012)	J48, naïve Bayes and neural network	Size of data 257 records	the Naïve Bayes classifier outperforms in prediction decision tree and neural network methods
(DorinaKabakchieva, 2013).	decision tree algorithm C4.5 (J48), two Bayesian classifiers (NaiveBayes and BayesNet), a Nearest Neighbour algorithm (IBk)	Size of data 10330 records	the decision tree classifier (J48) performebest (with the highest overall accuracy)

	and two rule learners (OneR and JRip)		
(Wan Fairos Wan Yaacob,SyerinaAzlinMdNasir,WanFaizah Wan Yaacob,Norafefah Mohd Sobri,2019)	K-Nearest Neighbor, Naïve Bayes, Decision Tree and Logistic Regression	Size of data 631 records	The Naïve Bayes classifier perform best (with the highest overall accuracy) followed by LR, K-NN, DT Information Gain and DT GINI
(C. Anuradha,T. Velmurugan, 2015)	decision tree J48, Naive Bayes Classifiers, k-Nearest Neighbours algorithm, OneR and JRip algorithm27, 28	-	That the Bayesian classifier outperforms the other (more than 70%)
(R. R. Kabra,R. S. Bichkar, 2011).	Decision Trees	Size of data 346 records	the accuracy of this model is 69.94 %, that is out of 346 instances 242 are correctly classified
(Surjeet Kumar Yadav,Saurabh Pal, 2012)	ID3, C4.5 and CART	Size of data 90 records	model is successfully identifying the students who are likely to fail
(Brijesh Kumar Baradwaj,Saurabh Pal, 2011)	decision tree	Size of data 50 recordss	classification rules can be generated for each path from each terminal node to root node

(Hafez Mousa,AshrafMaghari, 2017)	Decision Tree	Size of data 1036 recordss	The Decision Tree (DT) outperforms high accuracy (92.96%)
(AbeerBadr El Din Ahmed,Ibrahim Sayed Elaraby, 2014)	ID3 Decision Tree	Size of data 1547 records	identified those students which needed special attention to reduce failing ration and taking appropriate action at right time
(AnnisaUswatunKhasanah,Harwati, 2018)	Decision Tree, Bayesian Network	Size of data 104 records	the Bayesian Network achieved high accuracy more than Decision Tree (98%08)

2.6 Summary:

This chapter reviewed the data mining techniques and especially that was implemented in the educational section existing surveys and journal papers about analyzing the student's performance using different techniques, educational data mining as well as Predict of students Performance based on data mining techniques. Therefore, depending on reviewed papers. This research adopts methodology according to Predict student's performance based on data mining techniques which is explaining broadly in next chapter.

CHAPTER THREE

METHODOLOGY

This chapter explains the methodology used in this research.

3.1 Methodology:

The methodology was carried on four phases in order to achieve the objective of this research. The first phase presents the Data Collection. The second phase explains the data preprocessing techniques applied to the dataset. The third phase reconsiders the algorithms implemented by the researcher based on Predict of student's performance as discussed in previous chapter last phase implements the association rule algorithm. all the methodology phases implementing using WEKA tools. Workbench contains a collection of visualization tools and algorithms for data analysisandpredictive modeling, together with graphical user interfaces for easy access to this functionality. Weka is a collection of machine learning algorithms for solving real world data mining problems. It is written in Java and runs on almost any platform. The algorithms can either be applied directly to a dataset or called from your own Java code. The original non-Java version of Weka was aTCL/TKfront-end to (mostly third-party) modeling algorithms implemented in other programming languages, plus data preprocessing utilities inC, and aMake file-based system for running machine learning experiments. This original version was primarily designed as a tool for analyzing data from agricultural domains, but the more recent fully Java-based version (Weka 3), for which development started in 1997, is now used in many different application areas, in particular for educational purposes and research (Dr. Sudhir B. Jagtap, 2013). explains the four steps to implement the mentioned methodology **Figure (3.1).**

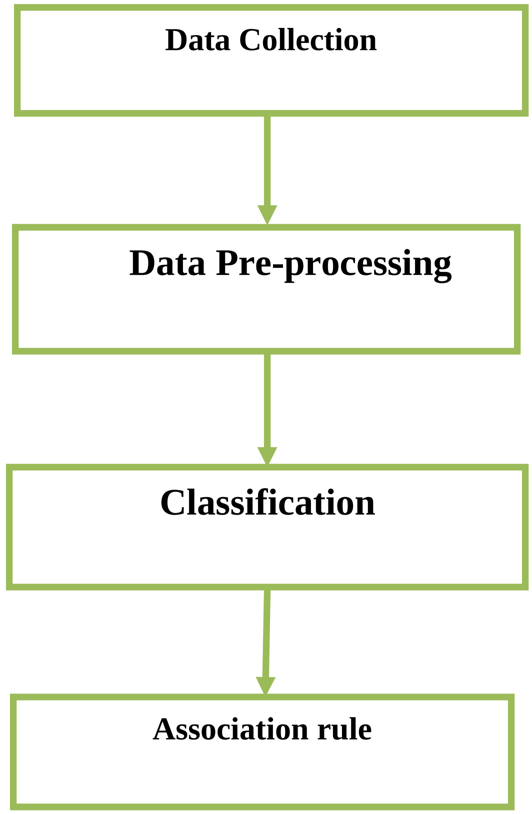


Figure (3.1): Methodology for proposed model

3.1.1 Data Collection phase:

Student data was obtained from student's files from the Registrar Office and the Office of Examinations at the Faculty of Economics and Business Studies, where it was available in paper files, which is data for regular baccalaureate students enrolled through the year's study 2010 to 2014. The structure of study in faculty has been distributed into four academic years equivalent to eight semesters. Each year study included two semesters, the number of courses in faculty more than sixty-two courses affording to all the semesters. The dataset contains the number of the index, the name of the student, University ID, courses, GPA and departments, all data were written manually. The data before collection as shown in Appendix A.

3.1.1.1 Dataset Description:

The number of Instances: 1778 instance (the number of rows in the dataset which rows refer to the instances). It contains 36 attributes: attribute (the number of column in the dataset which columns refer to the attributes).

3.1.1.2 Attribute Information:

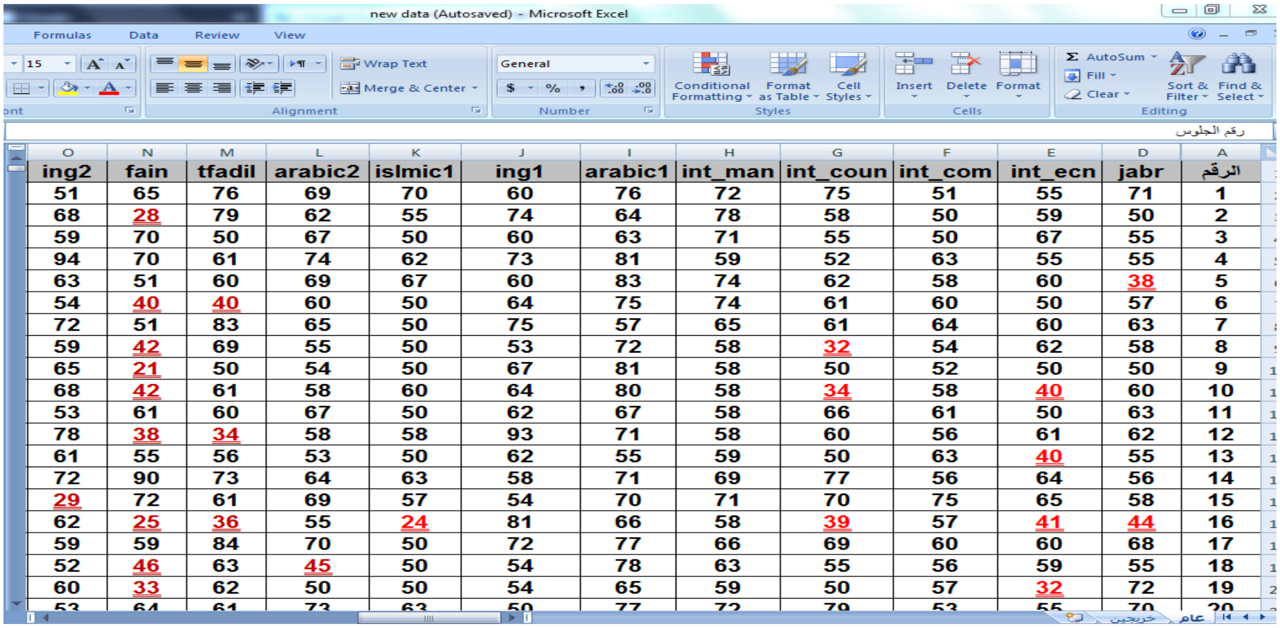
- Index: student index a digital sequence for preparing students, can do without.
- Jabr, int_ecn,com,int_coun,int_man,arabic1,ing1,islmic1: the first semester subjects include eight courses.
- arabic2,tfadil,fain,ing2,mang1,islmic2,ecno_j,mis:the second semester subjects include eight courses.
- arabic3,mid_coun1,mang2,econ_k,islmic3,int_sta,prins,int_low: the third semester subjects include eight courses.
- ecno_s,low_mar,mid_coun2,p_sta,ing3,p_bank1,org_be,com_ap:the fourth semester subjects include eight courses.
- GPA: GPA for the student who graduated with (3:4 excellent, 2.99:2.7 very good, 2.69:2.41good, 2.40:2.00pass and Less than 2 GPA failure).
- Dep: The Department of the student in the faculty (Accounting, Management Information Systems, Business Administration, Economic, Banking and Financial Studies).
- Class: selected as label class**.**

3.1.2 Data Preprocessing phase:

Applying data preprocessing techniques before mining will improve and reduce the required time for the actual mining. In this study some general tasks of the data preprocessing have to be performed on the dataset, such as data integration, data cleaning, data reduction, data transformation.

3.1.2.1 Data Integration:

The first task of the data preprocessing is the data integration. In this research The data was integrated in one excel sheet file, it contains the first year (two semesters) and second year (two semesters), these two years are called the general section before specialization, from the fifth semester to the eight semester (department of specialization) we taken specialization and GPA. The data contains 1778records, which turns to be 1402 records after preprocessing Figure (3.3) explains the sample of data for one semester; we have hidden the name and University ID due to privacy.



ing2	fain	tfadil	arabic2	islmic1	ing1	arabic1	int_man	int_coun	int_com	int_ecn	jabr	الرقم
51	65	76	69	70	60	76	72	75	51	55	71	1
68	28	79	62	55	74	64	78	58	50	59	50	2
59	70	50	67	50	60	63	71	55	50	67	55	3
94	70	61	74	62	73	81	59	52	63	55	55	4
63	51	60	69	67	60	83	74	62	58	60	38	5
54	40	40	60	50	64	75	74	61	60	50	57	6
72	51	83	65	50	75	57	65	61	64	60	63	7
59	42	69	55	50	53	72	58	32	54	62	58	8
65	21	50	54	50	67	81	58	50	52	50	50	9
68	42	61	58	60	64	80	58	34	58	40	60	10
53	61	60	67	50	62	67	58	66	61	50	63	11
78	38	34	58	58	93	71	58	60	56	61	62	12
61	55	56	53	50	62	55	59	50	63	40	55	13
72	90	73	64	63	58	71	69	77	56	64	56	14
29	72	61	69	57	54	70	71	70	75	65	58	15
62	25	36	55	24	81	66	58	39	57	41	44	16
59	59	84	70	50	72	77	66	69	60	60	68	17
52	46	63	45	50	54	78	63	55	56	59	55	18
60	33	62	50	50	54	65	59	50	57	32	72	19

Figure (3.3) sample dataset

3.1.2.2 Missing Values Manipulation:

The second task of the data preprocessing is handling the missing data; the concept of missing values is important to understand in order to successfully manage data. The problem of missing data generally arises due to the absence of data in a statement for any variable during the experiments, or when no information is provided or unavailable for the variables. If the missing values are not handled properly by the researcher, then he/she may end up drawing an inaccurate inference about the data. In this study, there are two cases of the missing data will be handling. First case; when the student is absent for one exam, which was handled by using the attribute mean for Replace missing values of an attribute with the mean value for that attribute in the same class as the given tuple. The second case; when the student is absent of the more than exam, which was handled by Ignoring the data row (remove row), Figure (3.4) explains the sample for missing data.

M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
68	62	61	64	54	75	50	60	51	50	50	52	58	56	51	61	50	72	73
50	50	79	75	50	55	50	50	62	41	60	67	55	55	50	60	54	76	63
64	71	82	74	54	85	61	60	78	62	58	87	64	69	59	71	66	67	74
50	69	67	78	50	65	50	70	69	50	54	57	65	60	50	61	57	64	70
96	92	91	77	61	85	81	65	88	64	67	83	79	77	68	79	59	73	83
79	92	63	50	57	71	72	50	84	57	59	77	53	65	55	60	52	75	58
59	10	57	70	40	75	50	60	79	50	52	71	75	64	50	64	57	64	73
58	63	63	69	57	70	66	79	91	50	54	70	53	67	75	79	58	74	74
71	73	83	74	60	75	54	87	97	51	66	85	76	74	63	80	71	78	89
67	87	91	78	57	50	64	50	55	59	54	52	50	54	50	50	50	50	70
60	66	50	50	67	75	35	50	64	31	56	69	51	54	50	46	47	70	64
50	50	64	50	50	85	54	67	72	50	50	77	69	66	70	59	72	65	86
50	80	79	65	50	70	50	50	67	50	52	62	55	57	60	63	59	73	86
50	50	50	62	60	85	57	58	93	68	62	69	67	70	80	66	50	73	80
86	78	68	88	50	80	50	60	88	50	50	63	70	64	71	69	50	75	77
71	50	79	69	50	75	50	84	71	50	71	64	57	65	50	51	69	72	83
65	81	77	75	60	50	34	a	a	a	a	a	a	42	50	50	50	80	83
50	58	67	55	56	65	50	78	79	50	50	80	60	64	55	71	50	81	86
62	61	73	69	50	75	57	59	88	54	50	79	58	65	71	58	50	58	82
72	50	64	69	60	65	90	50	100	81	53	82	50	71	50	59	61	72	68
90	73	81	74	55	70	50	77	89	55	61	75	71	69	58	72	77	77	84
50	50	54	40	42	60	52	70	73	50	50	81	52	61	74	64	52	80	87
50	54	62	66	71	75	67	78	95	59	65	77	74	74	50	55	50	60	75
61	67	76	88	64	70	50	62	80	50	50	77	53	62	79	67	73	78	88
72	79	81	73	54	65	50	50	64	39	50	66	59	55	55	50	30	69	60
50	63	85	59	66	65	50	53	62	39	50	50	55	53	50	59	60	77	81
50	62	67	68	53	2	3	3	3	3	2	2	3		42	50	60	77	75

Figure (3.4) sample for missing data

3.1.2.3 Data Transformation:

The last task in the data preprocessing is the data transformation. Data transformation is the process of changing the format, structure, or values of data into forms appropriate for mining. Data transformations involve many techniques; we transform the dataset format to a file in a format ARFF (Attribute Relation File Format) to be ready to work in the WEKA tool. Finally, we will take from student grade system four year for applying prediction techniques to know the distribute of student of different specialization in the in college. And we will take the first two years and applying the prediction techniques, and we will take two instances (final GPA and department) form last two years, and we have the “class” attribute was generated and it held the predicted result, which can be either “yes” or “no” or “failure”, Where Yes means that the student has chosen the appropriate major and graduated from it at a rate above 2.7, and no, mean that the student has chosen an inappropriate major and graduated from it at a rate lower than 2.7, failure means that the student has failure in GPA.

3.1.3 Classification phase:

Classification is the process of finding a model (or function) that describes and distinguishes data classes, in order to be able using this model to predict the class of objects whose class label is unknown. This phase the Naïve Bayes, j48 and Random Forest algorithms were applied on the dataset. There have three experiments, in the first experiment the dataset was divided up to 60% for training and 40% for testing, in second experiment the dataset was divided up to 70% for training and 30% for testing and last experiment the dataset was divided up to 80% for training and 20% for testing.

3.1.4 Association Rule phase:

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. An $X \Rightarrow Y$ type association rule expresses a close correlation between items (attribute-value) in a database. There are many association rule mining algorithms such as Apriori algorithmApriori is designed to operate on databases containing transactions (for example, collections of items bought by customers, or details of a website frequentation). The whole point of the algorithm (and data mining, in general) is to extract useful information from large amounts of data. The algorithm aims to find the rules which satisfy both a minimum support threshold and a minimum confidence threshold (Strong Rules) (codeproject, n.d.). Apriori is the first and foremost algorithm in association rule. Most association rule mining algorithms require the user to set at least two thresholds, one of the minimum support and the other of minimum confidence. The support S of a rule 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. we used Association rules for build model that helped to identify if the students' academic qualify to study the Specialization and predict the result that students can obtain based on available data. After many experiments were conducted on the Weka program, by changing values (Confidence) and (Support), and change the attributes selected in the configuration phase of the data(Preprocessing), A large amount of rules has been obtained. After loading those rules, it is found that, some of them are classified as trivial and some of them have a logical degree of value (Confidence). As a result, we notice that the greater the number of (item set) and the smaller the number of items produced, the more the relations between them are more logical, the change in both the values of "confidence" and " (increase or decrease) causes different rules with varying degrees of logic to appear. There was also a large amount of intuitive results and illogical results when determining all the attributes to find the relations between them, which made me, identify certain qualities and conduct operations on them to discover more relationships logical between results (Angeline, 2013).

3.2 Evaluation Measures:

The evaluation criteria that used in this research are: Accuracy, Perception, Recall and F-Measure

3.2.1 Accuracy:

Accuracy rate (AC): the percentage of correct predictions. According to the confusion matrix(Ünal, 2019), it can be calculated as

$$AC = \frac{TN + TP}{TP + FP + FN + TN},$$

where TN is the true negative, TP is the true positive, FP is the false positive, and FN: false negative.

3.2.2 Precision:

Precision (P): the fraction of correctly predicted positive observations among the total predicted positive observations(Ünal, 2019).

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}},$$

where TP is the true positive and FP: false positive**.**

3.2.3 Recall:

Recall (R): the fraction of correctly predicted positive observations among all the observations in the class(Ünal, 2019).

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

where TP is the true positive and FN: false negative**.**

3.2.4 F-measure:

The Precision and Recall criteria can be interpreted together rather than individually. To accomplish this, we consider the F-Measure values generated by the harmonic mean of the Precision and Recall columns, as the harmonic mean provides the average of two separate factors produced per unit. Therefore, F provides both the level of accuracy of the classification and how robust (less data loss) it is(Ünal, 2019)

$$F - \text{measure} = \frac{2 \times P \times R}{P + R},$$

where P is the precision and R is the recall.

3.2.5 Basic quantitative quality indicators:

Predicted real	**Class1 Predicted**	**Class2 Predicted**
Class1 Actual	TP	FN
Class2 Actual	FP	TN

where:

Class 1: Positive(P)

Class 2: Negative (N)

TP: True Positive the number of observations correctly assigned to the positive class (instances correctly classified as a given class).

TN: True Negative the number of observations correctly assigned to the negative class (instances falsely classified as a given class).

FP: False Positive the number of observations assigned by the model to the positive class, which in reality belong to the negative class.

FN: False Negative the number of observations assigned by the model to the negative class, which in reality belong to the positive class.

3.3 Summary:

The purpose of this chapter is to present and discuss the approach and methods of the research. Hence, it covers the methodological aspects that have guided the present work. It starts with an introduction, which gives an overview of the methodology of work then data collection and data description and preprocessing steps. Several learning algorithms used in this research are described, which are mainly used to determine the performance of the proposed method.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Introduction:

This chapter discussed the results of the experiment which have been implemented in chapter three. The results were carried on two phases first; explain experiments of classification applied in chapter three, results analysis. Second, intersection of the strong rules was generated in chapter three. After training the models and tested, the results are:

4.1.1 Classification Experiments:

In the first experiment split the data to 60% for training and 40% for testing, the experiment result as shown in Appendix and the result summary appear below have got:

Time is taken to build model: 0. 18 seconds.

Correctly classified instances 555 65.8%

Incorrectly classified instances 289 34.2%

Total number of instances is 844.

Table (4.1) show the Precision, recall, and F-Measure for Naive Bayes Classifier.

Class	Precision	Recall	F-Measure
Yes	0. 724	0. 725	0. 724
No	0. 546	0. 544	0. 545
Failure	1	1	1

In the second experiment split the data to 70% for training and 30% for testing, and have got:

Time is taken to build model: 0.05 seconds.

Correctly classified instances 658 66.9%

Incorrectly classified instances 326 33.1%

Total number of instances is 984.

Table (4.2) show the Precision, recall, and F-Measure for Naive Bayes Classifier

Class	Precision	Recall	F-Measure
Yes	0. 605	0. 628	0. 616
No	0. 718	0. 697	0. 707
Failure	1	1	1

In the third experiment split the data to 80% for training and 20% for testing, and have got:

Time is taken to build model: 0.03 seconds.

Correctly classified instances 740 65.8%

Incorrectly classified instances 384 34.2%

Total number of instances is 1124.

Table (4.3) show the Precision, recall, and F-Measure for Naive Bayes Classifier

Class	Precision	Recall	F-Measure
Yes	0. 61	0. 63	0. 62
No	0. 698	0. 68	0. 689
Failure	1	1	1

In the first experiment split the data to 60% for training and 40% for testing, and have got:

Time is taken to build model: 0. 16 seconds.

Correctly classified instances 801 94.9%

Incorrectly classified instances 43 5.1%

Total number of instances is 844.

Table (4.4) show the Precision, recall, and F-Measure for Random Forest.

Class	Precision	Recall	F-Measure
Yes	0. 948	0. 971	0. 959
No	0. 951	0. 912	0. 931
Failure	1	1	1

In the second experiment split the data to 70% for training and 30% for testing, and have got:

Time is taken to build model: 0. 13 seconds.

Correctly classified instances 942 95.7%

Incorrectly classified instances 42 4.3%

Total number of instances is 984.

Table (4.5) show the Precision, recall, and F-Measure for Random Forest.

Class	Precision	Recall	F-Measure
Yes	0. 958	0. 94	0. 949
No	0. 956	0. 97	0. 963
Failure	1	1	1

In the third experiment split the data to 80% for training and 20% for testing, and have got:

Time is taken to build model: 0. 18 seconds.

Correctly classified instances 1085 96.5%

Incorrectly classified instances 39 3.4%

Total number of instances is 1124.

Table (4.6) show the Precision, recall, and F-Measure for Random Forest**.**

Class	Precision	Recall	F-Measure
Yes	0. 967	0. 954	0. 96
No	0. 964	0. 974	0. 969
Failure	1	1	1

In the first experiment split the data to 60% for training and 40% for testing, and have got:

Time is taken to build model: 0. 28 seconds.

Correctly classified instances 822 97.4%

Incorrectly classified instances 22 2.4%

Total number of instances is 844.

Table (4.7) show the Precision, recall, and F-Measure for J48**.**

Class	Precision	Recall	F-Measure
Yes	0. 981	0. 977	0. 979
No	0. 963	0. 969	0. 966
Failure	1	1	1

In the second experiment split the data to 70% for training and 30% for testing, and have got:

Time is taken to build model: 0. 33 seconds.

Correctly classified instances 955 97.0%

Incorrectly classified instances 29 2.9%

Total number of instances is 984.

Table (4.8) show the Precision, recall, and F-Measure for J48.

Class	Precision	Recall	F-Measure
Yes	0. 971	0. 959	0. 965
No	0. 97	0. 979	0. 974
Failure	1	1	1

In the third experiment split the data to 80% for training and 20% for testing, and have got:

Time is taken to build model: 0. 55 seconds.

Correctly classified instances 1084 96.4%

Incorrectly classified instances 40 3.6%

Total number of instances is 1124.

Table (4.9) show the Precision, recall, and F-Measure for J48**.**

Class	Precision	Recall	F-Measure
Yes	0. 953	0. 972	0. 962
No	0. 974	0. 962	0. 968
Failure	0	0	0.998

After conducted the experiments using three classifiers we can summarize that compare between the previous models depending on their accuracy for classifying our dataset to measure the accuracy of these algorithms on those dataset, all classification algorithms used in this thesis achieve a high accuracy in classifying. The best result obtained using j48 algorithm, which get higher accuracy than the other algorithm in all experiments.

Table (4.10) show the comparison between algorithms accuracy

Algorithm	60%	70%	80%
Naive Bayes	66%	67%	66%
Random Forest	95%	96%	96%
J48	97%	97%	96%

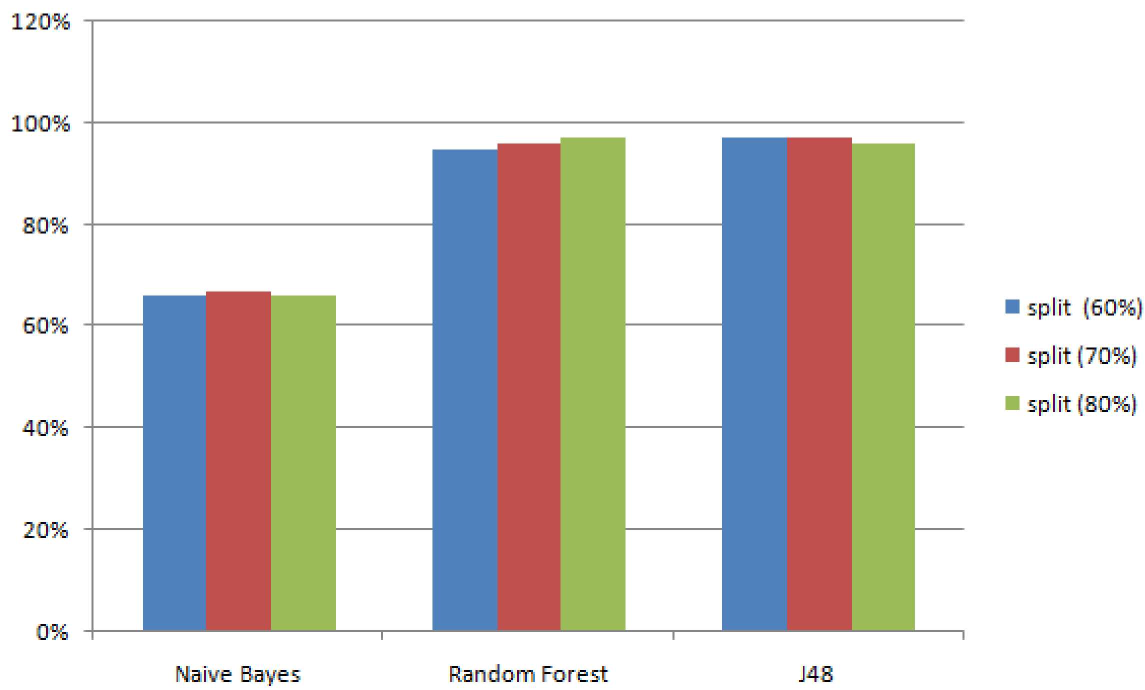


Figure (4.1): shows the accuracy of all models.

4.1.2 Association rule Experiments:

In order to the applying apriori algorithm on the classification results to be covert the numerical values to the nominal values. Many experiments are deployed on the apriori algorithm that considered the results of classifier were generated algorithm. In the first experiment: on the data was generated strong rule with (Support 11% and Confidence 100%) (see Appendix A1). The second experiment as shown in (Appendix A2) also generated strong rules it is observed with (Support 11% and Confidence 90%). The third experiment was generated strong rules as appear in (Appendix A3) with (Support 1% and Confidence 100%).The last experimented: applied on a dataset without using classifier that is appear weak rules as shown in (Appendix A4)

Best rules found:

4 Dep=Accounting degree=Good 155 ==> class=No 155 conf:(1).

5 arabic2=A 319 ==> dep=Accounting 217 conf:(0.68).

6 islmic1=D 381 ==> islmic2=D 248 conf:(0.65).

7 ing2=D 441 ==> islmic2=D 286 conf:(0.65).

8 islmic1=D 381 ==> ing2=D 246 conf:(0.65).

9 int_coun=D 415 ==> dep=Accounting 238 conf:(0.57).

10 islmic3=D 407 ==>int_low=D 232 conf:(0.57).

11 int_low=D degree=Good 145 ==> class=No 145 conf:(1).

4.1.3 Discussion the results:

4 Students who scored (C) in English language scored (A) in linear algebra graduated from the Accounting Department with a v. good grade.

5 Students who scored (C) in the English language subject scored A in the Introduction to Accounting course and graduated from the Accounting Department with a good grade.

6 Students who graduated with a good grade from the Department of Economics had obtained a (C) in the subjects of Microeconomics and Econometrics.

7 99% of the students who graduated with a pass grade from the Department of Management Information Systems had obtained a (D) in computer introduction, Commercial Law and linear algebra.

8 Students who graduated with excellent grade from Business Administration had obtained (A) degree in Accounting, Linear Algebra and Management Information Systems.

9 Students who graduated with a good grade from the Banking Studies Department had obtained a (B) in the subject of Commercial Law and (C) in the English2 and (D) in the Arabic2.

10 Students who graduated with a v. good grade from the Department of Economics had obtained a (B) in the subject of Commercial Law and (A) in the English2 and (C) in the Arabic2

4.2 Summary:

In this chapter, the data of the available students were examined. The data were categorized using several predictive modeling techniques; they were able to predict the students' rate according to the model they designed. We compared the algorithms used, but the J48 algorithm achieved best results were. Associated rules have been created and the best rules that have achieved better results have been selected and compared with the real reality.

CHAPTER FIVE

CONCLUSION AND FUTURE WORK

5.1 Conclusion:

In this study, various data mining techniques were discussed to support education system via generating strategic information. Since the application of data mining brings a lot of advantages in higher learning institution, the study was applied on real data collected form faculty of Economics and Business Studies, University of KORDOFAN. Some preprocessing phases are applied on data such as handle missing data, data transformation and attribute selected. Then applied three data mining algorithms which are naive Bayes, J48, Random Forest and Apriori algorithm to achieve the objective of study. The results show that J48 classifier was achieved 97.6% of accuracy and the second technique is Random Forest with accuracy %96 and the worst technique is Naïve Bayes with accuracy %67Forest. In addition to this, Apriori algorithms used to generate strong rules to identify if the student's academic results are qualified to study the Specialization

5.2 Future work:

There are aspects that have not been addressed in this research due to the lack of complete data or that it is outside the limits of the research such as Subjects related to the major, data related to family income, type of acceptance and the percentage of the Sudanese certificate on it we recommend taking it into account in future studies**.**

5.3 Recommendations:

To obtain the best results, the university database as a whole must be improved in the future by Establish data center at the University to facilitate gets the students, staffs and courses information which helps in data collection.

The data should be stored well enough, to be accessible, thus helping to complete research in this area.

Correlation between specific subjects appears at lower levels for students. Therefore, we recommend that decision-makers stand up to it and try to explain it according to the nature of the subject or its location in the study plan or the performance of some professors.

To increase the efficiency of the results obtained, we recommend adding the results of bridging and online education.

.

REFERENCES:

1Y Divyabharathi, 2. S. (2018). A Framework for Student Academic Performance Using Naive Bayes Classification Technique. JOURNAL OF ADVANCEMENT IN ENGINEERING AND TECHNOLOGY, 4.

Abeer Badr El Din Ahmed1, I. S. (2014). Data Mining: A prediction for Student's Performance. World Journal of Computer Application and Technology 2(2): 43-47, 5.

Al-Razgan, Mashael A. Al-Barrak and Muna. (2016). Predicting Students Final GPA Using Decision Trees: A Case Study. International Journal of Information and Education Technology, 6.

Angeline, D. M. (2013). Association Rule Generation for Student Performance Analysis using Apriori Algorithm. The SIJ Transactions on Computer Science Engineering & its Applications (CSEA), 5.

Annisa Uswatun Khasanah, H. (2018). A Comparative Study to Predict Student’s Performance Using Educational Data Mining Techniques. IOP Conf. Series: Materials Science and Engineering, 8.

B.SANTHOSH KUMAR, K. (2010). MINING WEIGHTED ASSOCIATION RULE USING HITS.

Bassil, Y. (2012). A Data Warehouse Design forA Typical University Information System. Journal of Computer Science & Research (JCSCR), 6.

Bichkar, R. R. (2011). Performance Prediction of Engineering Students using. International Journal of Computer Applications (0975 – 8887), 5.

Brijesh Kumar Baradwaj, S. P. (2011). Mining Educational Data to Analyze Students‟Performance. International Journal of Advanced Computer Science and Applications, 7.

C. Anuradha, T. V. (2015). A Comparative Analysis on the Evaluation of Classifcation Algorithms in the Prediction of Students Performance. Indian Journal of Science and Technology, 12.

codeproject. (n.d.). Retrieved from codeproject.com: https://www.codeproject.com/Articles/70371/Apriori-Algorithm

Dr. Sudhir B. Jagtap, D. K. (2013). Census Data Mining andData Analysisusing WEKA. nternational Conference in “Emerging Trends in Science, Technology and Management(ICETSTM –2013) , 6.

Edin Osmanbegović, M. S. (2012). DATA MINING APPROACH FOR PREDICTING STUDENT PERFORMANCE. Economic Review – Journal of Economics and Business, 11.

Hafez Mousa1, A. M. (2017). School Students' Performance Predication Using Data Mining Classification. nternational Journal of Advanced Research in Computer and Communication Engineering, 6.

HAN, J. K. (2011). Data mining concepts and techniques third edition. Morgan Kaufmann.

Humera Shaziya, R. Z. (2015). Prediction of Students Performance in Semester Exams using a Naïve bayes Classifier. International Journal of Innovative Research in Science,Engineering and Technology, 8.

Jehad Mergani, E. M. (n.d.). Application of analyzing students’ academic performance using decision tree.

Joshua D. Fostera, W. K. (2007). Are there such things as ‘‘Narcissists’’ in social psychology? Ataxometric analysis of the Narcissistic Personality Inventory. Personality and Individual Differences, 12.

Kabakchieva, D. (2013). Predicting Student Performance by Using Data Mining. CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume 13, No 1, 12.

Kabakchieva, Dorina. (2012). Student Performance Prediction by Using Data Mining Classification Algorithms. International Journal of Computer Science and Management Research, 5.

Kalpesh Adhatrao, A. G. (2013). PREDICTING STUDENTS’ PERFORMANCE USING ID3 AND C4.5 CLASSIFICATION ALGORITHMS. International Journal of Data Mining & Knowledge Management Process (IJDKP), 14.

Kamber, H. a. (2006). Data Mining:Concepts and Techniques. Diane Cerra.

Mohammed M. Abu Tair, A. M.-H. (2012). Mining Educational Data to Improve Students’ Performance: A Case Study. International Journal of Information and Communication Technology Research, 7.

Norhidayah Mohamad Yatima, b. N. (2015). PARTICLE FILTER IN SIMULTANEOUS LOCALIZATION AND MAPPING (SLAM)USING DIFFERENTIAL DRIVE MOBILE ROBOT. JurnalTeknologi, 7.

Pant, A. (2017). Determining Factors to Predict Rate of Student Dropout. Department of Graduate StudiesNepal College of Information TechnologyBALKUMARI, LALITPUR, NEPAL, 43.

Paola Galdi, R. T. (2018). Data Mining: Accuracy and Error Measures for Classification and Prediction. 15.

Patil, D. P. (2015). A STUDY OF STUDENT’S ACADEMIC PERFORMANCE USING DATA MININGTECHNIQUES. INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS, 5.

Rohaila Abdul Razak1, M. O. (2018). A Student Performance Prediction Model Using Data Mining Technique. International Journal of Engineering & Technology, 3.

S.Venkata Krishna Kumar, P. (2015). An Overview of Classification Algorithm in Data mining. International Journal of Advanced Research in Computer and CommunicationEngineering, 3.

Suchita Borkar, K. R. (2013). Predicting Students Academic Performance Using Education Data Mining. International Journal of Computer Science and Mobile Computing, 7.

Sujith Jayaprakash, D. J. (2018). A Survey on Academic Progression of Students in TertiaryEducation using Classification Algorithms. International Journal of Engineering Technology Science and Research IJETSR, 7.

Sundar, P. (2013). A COMPARATIVE STUDY FOR PREDICTING STUDENT'S ACADEMIC PERFORMANCE USING BAYESIAN NETWORK CLASSIFIERS. IOSR Journal of Engineering (IOSRJEN), 6.

Surjeet Kumar Yadav, S. P. (2012). Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification. World of Computer Science and Information Technology Journal (WCSIT), 6.

Ünal, B. Ç. (2019). Comparison of Data Mining Classification Algorithms Determining the Default Risk.

V.Ramesh, P. (2013). Predicting Student Performance: A Statistical and Data Mining Approach. International Journal of Computer Applications, 6.

Wan Fairos Wan Yaacob, S. A. (2019). Supervised data mining approach for predicting student performance. Indonesian Journal of Electrical Engineering and Computer Science, 9.

Y Divyabharathi, P. S. (2018). A Framework for Student Academic Performance Using Naive Bayes Classification Technique. 4.

Appendix

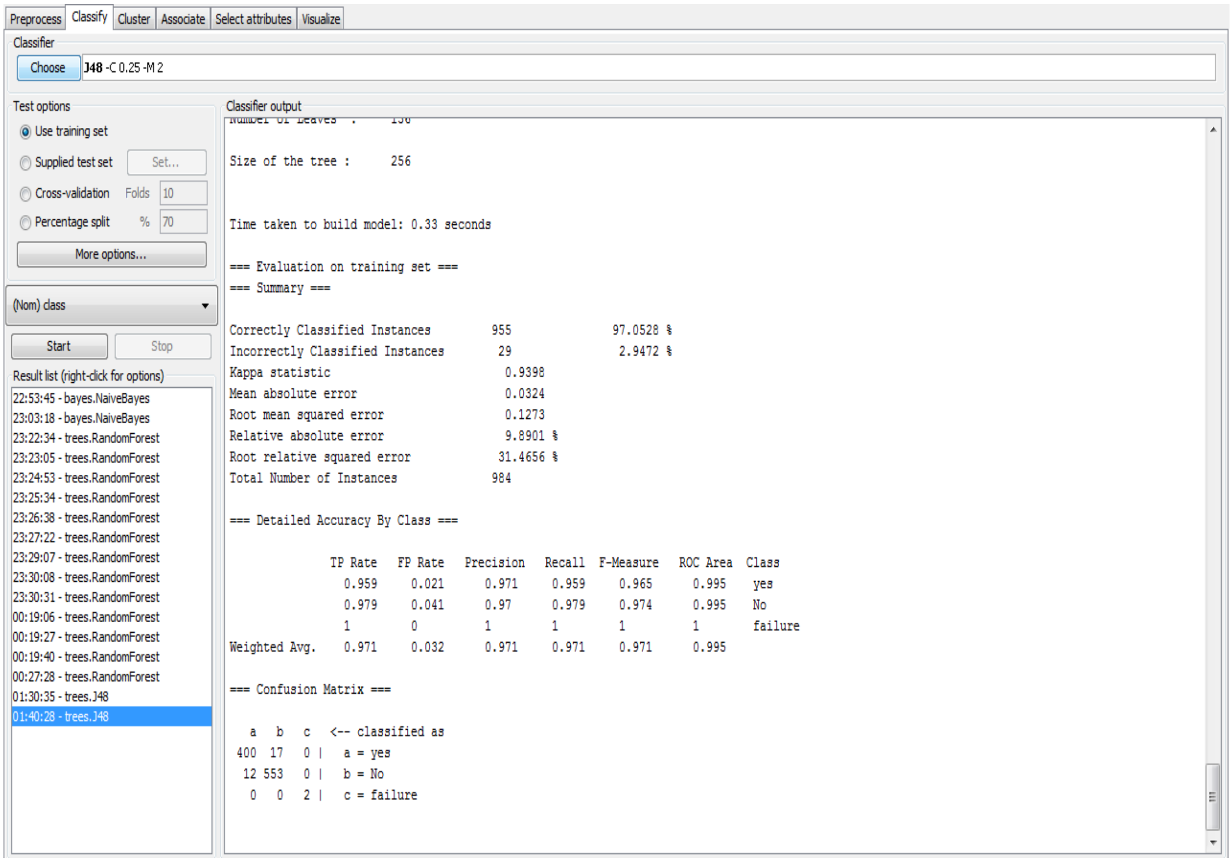


Figure. A.1:ApplyingJ48

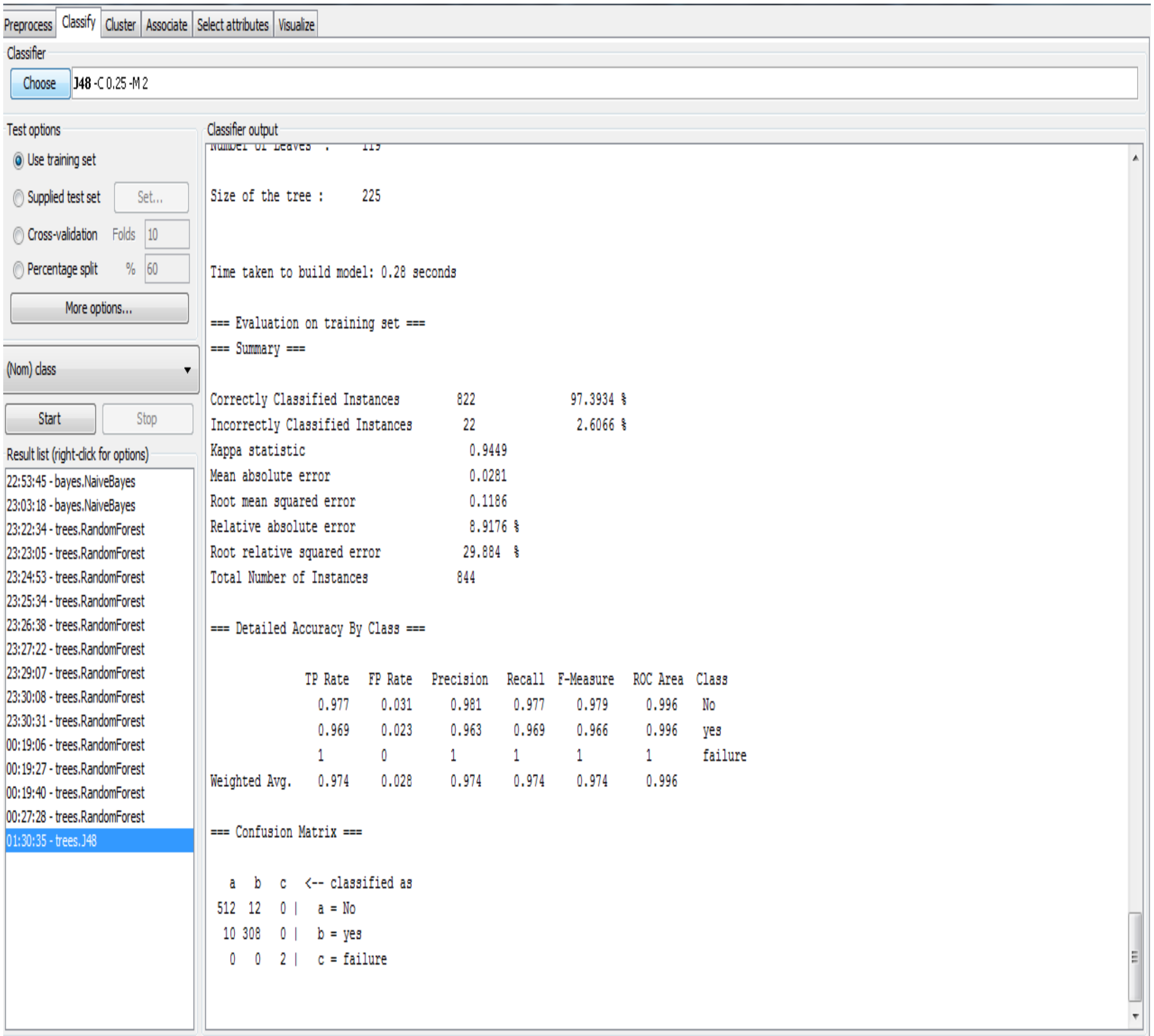


Figure. A.2: Applying Random Forest

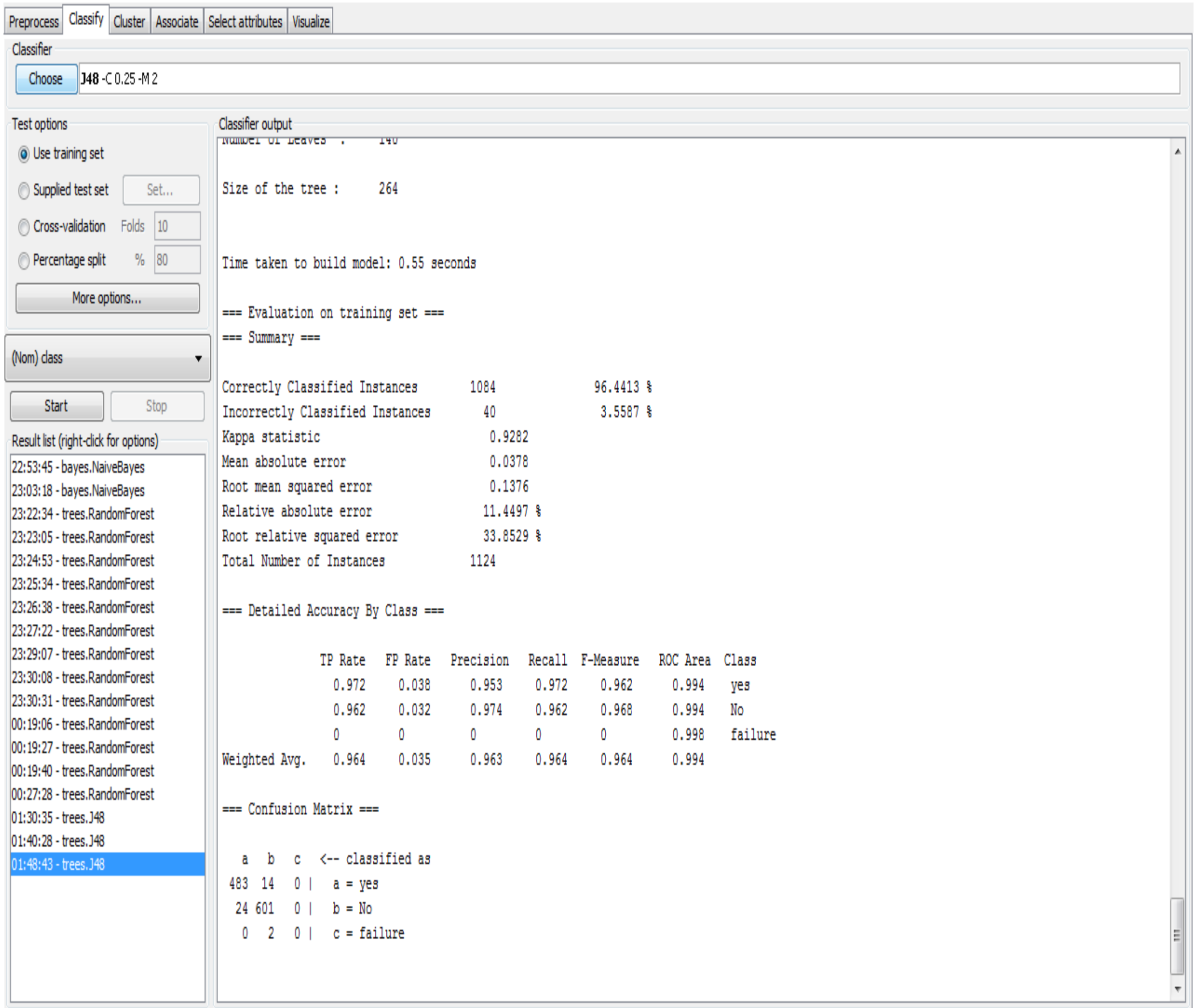


Figure. A.3: Applying Naïve bayes

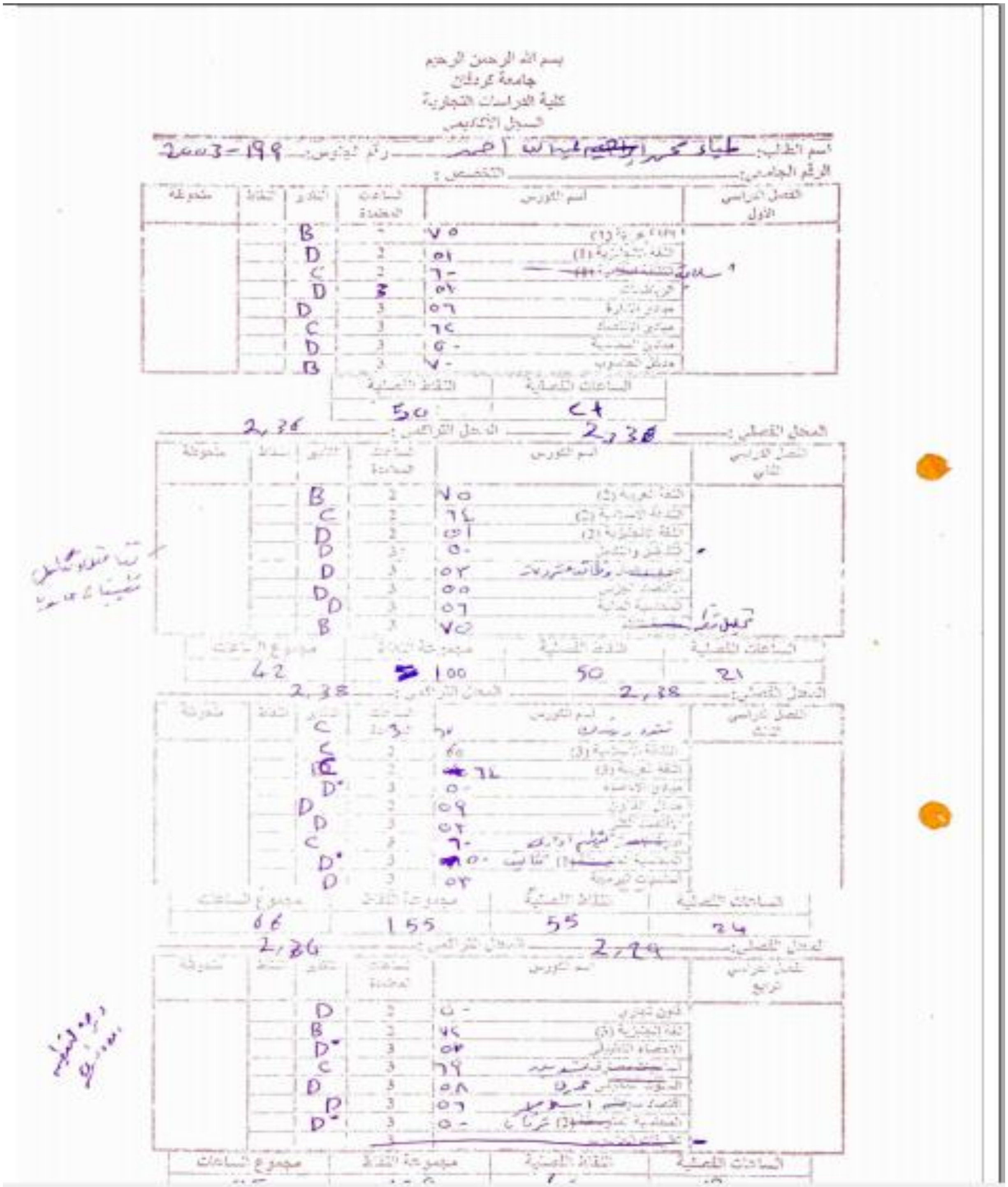

بسم الله الرحمن الرحيم
جامعة كردفان
كلية الدراسات التجارية
السجل الأكاديمي

Figure. A.4: data of students

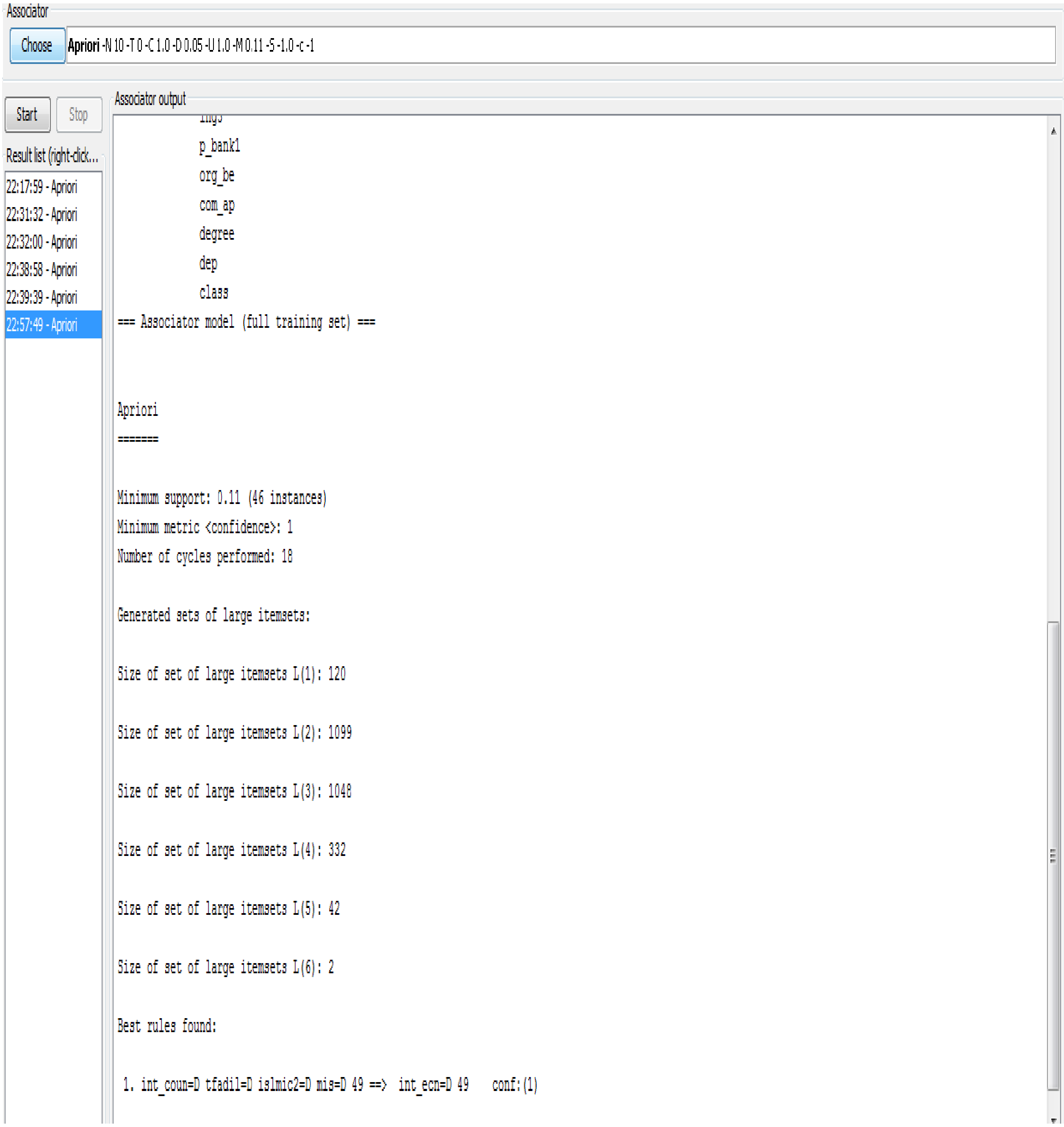


Figure.A.5: generated strong rules with Support 11% and Confidence 100%

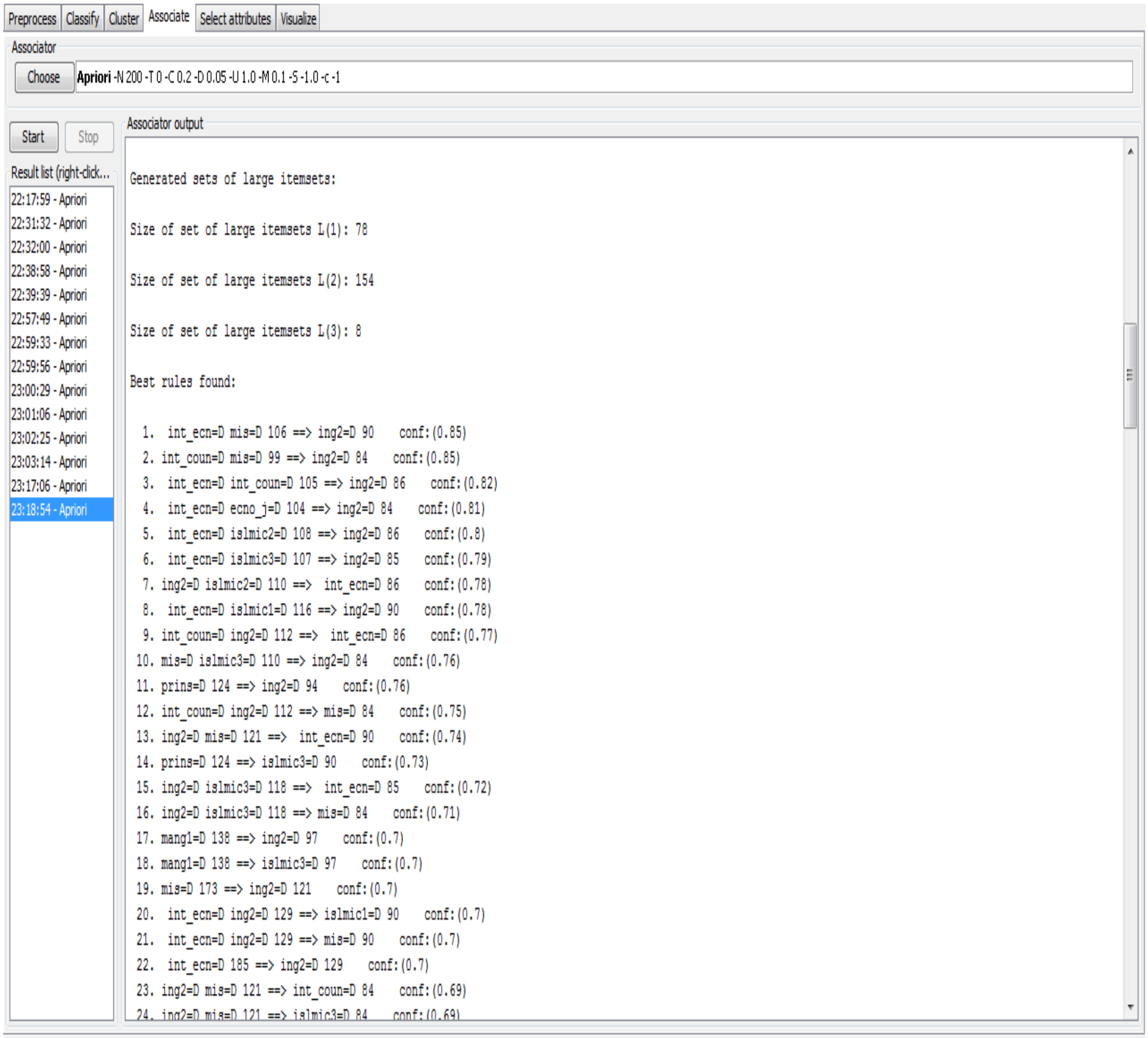


Figure. A.6: generated strong rules with Support 11% and Confidence 90%

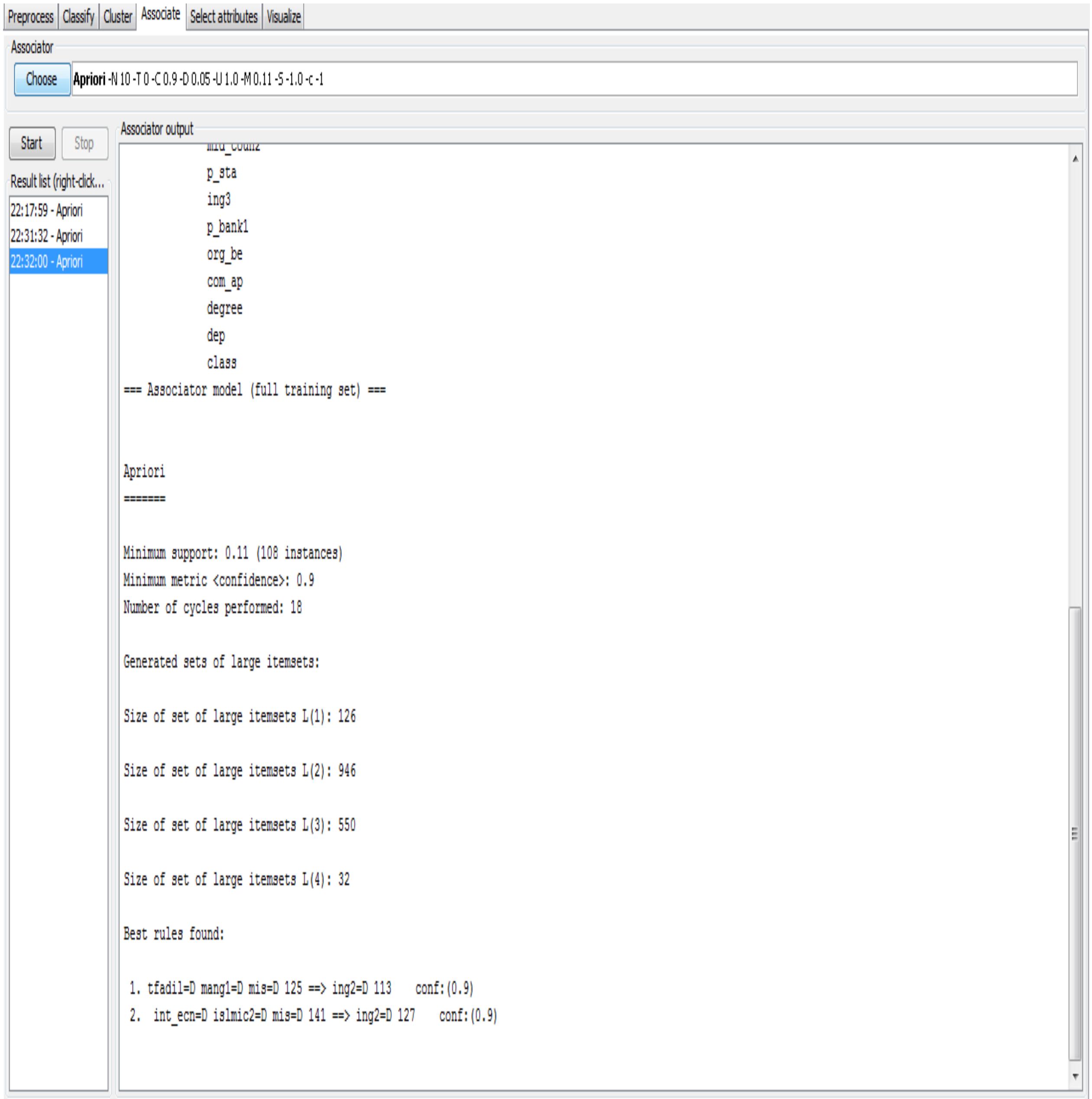


Figure.A.7: generated weak rules dataset without using classifier

```
19.  int_ecn=D 371 ==> islmic2=D 232    conf:(0.63)
20. int_low=D 422 ==> ing2=D 263    conf:(0.62)
21. islmic3=D 407 ==> islmic2=D 253    conf:(0.62)
22. islmic3=D 407 ==> ing2=D 250    conf:(0.61)
23. ecno_j=D 381 ==> ing2=D 234    conf:(0.61)
24.  int_ecn=D 371 ==> int_low=D 227    conf:(0.61)
25. arabic1=D 358 ==> ing2=D 218    conf:(0.61)
26.  int_ecn=D 371 ==> mid_coun1=D 225    conf:(0.61)
27. arabic3=D 369 ==> mid_coun1=D 223    conf:(0.6)
28. mid_coun1=D 439 ==> ing2=D 264    conf:(0.6)
29. int_low=D 422 ==> islmic2=D 253    conf:(0.6)
30. ing2=D 441 ==> mid_coun1=D 264    conf:(0.6)
31.  int_ecn=D 371 ==> int_coun=D 222    conf:(0.6)
32. ing2=D 441 ==> int_low=D 263    conf:(0.6)
33. ing2=D 441 ==>  int_ecn=D 261    conf:(0.59)
34. dep=Accounting 435 ==> class=No 256    conf:(0.59)
35. islmic2=D 437 ==> tfadil=D 257    conf:(0.59)
36. int_coun=D 415 ==> ing2=D 244    conf:(0.59)
37. ecno_j=D 381 ==> islmic2=D 223    conf:(0.59)
38. islmic3=D 407 ==> mid_coun1=D 238    conf:(0.58)
39. arabic3=D 369 ==> ing2=D 215    conf:(0.58)
40. islmic1=D 381 ==> int_low=D 221    conf:(0.58)
41. islmic2=D 437 ==> islmic3=D 253    conf:(0.58)
42. islmic2=D 437 ==> int_low=D 253    conf:(0.58)
43. int_sta=D 380 ==> class=No 220    conf:(0.58)
44. ing2=D 441 ==> tfadil=D 255    conf:(0.58)
45. ecno_j=D 381 ==> int_low=D 219    conf:(0.57)
46. class=No 475 ==> degree=Good 273    conf:(0.57)
47. int_coun=D 415 ==> dep=Accounting 238    conf:(0.57)
48. int_sta=D 380 ==> mid_coun1=D 217    conf:(0.57)
49. islmic3=D 407 ==> int_low=D 232    conf:(0.57)
50. tfadil=D 392 ==> int_coun=D 223    conf:(0.57)
```

Figure.A.8: sample of rules