PREDICTION OF STUDENTS DROPOUT USING DATA MINING

BY

NAZIFA TABASSUM

ID:161-15-7253

AND

REZOYANA ISLAM BONNA

ID: 161-15-6830

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

DR. SHEAK RASHED HAIDER NOORI

Associate Professor and Associate Head

Department of CSE

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

DECEMBER 2019

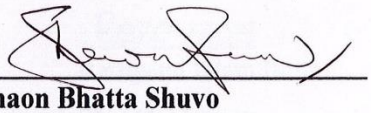
APPROVAL

This Project/internship titled **"Prediction of students Drop out using Data mining"**, submitted by Nazifa Tabassum, ID No: 161-15-7253 and Rezoyana Islam Bonna ID No: 161-15-6830 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 05/12/2019.

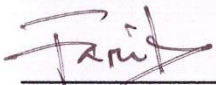
BOARD OF EXAMINERS

Dr. Syed Akhter Hossain — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Saiful Islam — **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Shaon Bhatta Shuvo — **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

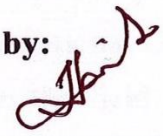


Dr. Dewan Md. Farid — **External Examiner**
Associate Professor
Department of Computer Science and Engineering
United International University

DECLARATION

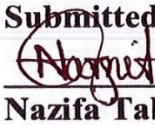
We hereby declare that, this project has been done by us under the supervision of **Dr. Sheak Rashed Haider Noori**, Associate professor and Associate Head, Department of CSE Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Dr. Sheak Rashed Haider Noori
Associate professor and Associate Head
Department of CSE
Daffodil International University

Submitted by:



Nazifa Tabassum
ID: 161-15-7253
Department of CSE
Daffodil International University

Rezoyana Islam Bonna
ID: 161-15-6830
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Dr. Sheak Rashed Haider Noori, Associate Professor and Associate Head,** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Data mining*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Syed Akhter Hossain Professor and Head,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

As a developing country, dropout is the dominant obstacle for our educational sectors. Therefore, it is important to develop the logical method for prediction of the students at the risk point of dropping out, allowing a proactive process to reduce this problem. This research work develops a prototype which can automatically recognize either the students will continue his/her study or dropout, using classification rules. Data were performed at one of the famous and prestigious university named Daffodil international university, with the main goal to reveal the high prospective of data mining applications. Data were collected from the students mainly focused on their personal and Family problems and university-performance. The responsible factors for dropping out were found through the Association technique using Apriori algorithm. Pre-processed factors were applied on the running students who were already completed one years or 3rd semester of their study. Classification method can be highly supportive in predicting student's dropout reasons. Selected 10 best attributes using CFS which were directly affected on the analysis. Finally, decision was making based onC4.5, Naïve Bayes algorithms that a running student would continue their study or would drop out. C4.5 algorithms was found the best classifier with 86.014% accuracy whereas Naïve Bayes was 76.22% of accuracy.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	I.
Declaration	II.
Acknowledgements	III.
Abstract	IV.
CHAPTER	
CHAPTER 1: INTRODUCTION	**1-3**
1.1.Introduction	**1**
1.2.Motivation	**1**
1.3.Rationale of the study	**1**
1.4.Objectives	**2**
1.5.Research Questions	**2**
1.6. Research Layout	**3**
CHAPTER 2: BACKGROUND	**4 - 7**
2.1.Introduction	**4**
2.2. Related Work	**4**
2.3. Bangladesh Perspective	**5**
2.4. Scope of The Study	**7**
2.5.Government Goals and Regulation	**7**
CHAPTER 3: METHODOLOGY	**8-14**
3.1. Introduction	**8**
3.2.Research Subject and Instrumentation	**8**
3.3.Data Collection Procedure	**8**
3.4. Work Methodology	**8**

3.5.Find the dropout students problems **9**

3.6. Data Collection **10**

3.7. Data Selection/ Data pre-procedure **12**

3.8. Techniques of data analysis **13**

CHAPTER 4: RESULT AND DISCUSSION **15-31**

4.1. Introduction **15**

4.2. Experimental Analysis **15**

4.3. Statistical Analysis **15**

4.4. Data Mining Analysis **21**

4.5. Application of Association Rule **29**

CHAPTER 5: CONCLUSION AND FUTURE WORK **32-33**

APPENDIX **34-35**

Appendix-1 **34**

Appendix-2 **35**

REFERENCES **36**

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.4.1: Work methodology	9
Figure 4.3.1: Types of Family problems	19
Figure 4.3.2: Types of personal problems	20
Figure 4.3.3: Types of Academic problem	20
Figure 4.4.1: Attribute Selected Using Correlation Feature Selection Background of C4.5	21
Figure 4.4.2: Decision Tree Using c4.5 Algorithm	23
Figure 4.4.3: Find the accuracy using Naive bayes algorithms	29
Figure 4.5.1: Select best rule using apriori algorithms	31

LIST OF TABLES

TABLES	PAGE NO
Table 2.3.1: Students' Enrollment Rate in a Few Countries	6
Table 3.6.1: Dropout students related variables	10
Table 3.6.2: Presented the data collection format	10
Table 4.3.1: Response the student's performance	16
Table 4.3.2: Response of dropout student and their reasons	18
Table 4.4.1: Confusion matrix of c4.5 algorithm	22
Table 4.4.2: Results for the C4.5 Decision Tree Algorithm Using training set (Accuracy by Class)	23
Table 4.4.3: Confusion matrix of c4.5 algorithm	24
Table 4.4.4: Classifier Rules	25
Table 4.4.5: Attributes were calculated with respect to Probability	26
Table 4.4.6: Results for the Naïve bayes Algorithm Using training set (Accuracy by Class)	27
Table 4.4.7: Confusion matrix of Naïve bayes algorithm	27
Table 4.5.1: Dataset with Description of Attribute and Possible Values	30

CHAPTER 1

Introduction

1.1. Introduction

One of the biggest challenging issue in our educational area is dropout. Increasing level of growing university is high in Bangladesh. The development of an individual and the forward movement of a nation depend on education. It is the main instrument in awakening the student to cultural values and thus is the strongest power in the development and growth of a student in preparing him/her to be a responsible, intelligent and capable citizen. Education is also equally important for both male and female. Nowadays the rate of female dropout is higher than male students. Dropout occurs for many reasons poverty, early marriage, financial problem, missed classes and many others. Data mining is the main approach for finding out the appropriate reasons of a student dropout. First data mining is collected vast area of dataset of a student.

1.2. Motivation

Nowadays universities environment is abundant combative to operate. The chief focused point is to highly analyzing students' performance, to noticed their uniqueness and to make a strategy for developing and taking action against drop out. Afterward detach the reasons and factors, generate an authorization to abatement the dropout rate. Yearly can attain more proficient graduate for make an educated nation. Helpful for assuage unemployment also.

1.3. Rationale of the study

It's not possible to ensure proper development in education sector without the proper maintenance of standard education. University education is like a crown for an educated person. It's a matter of sorrow that most of the students are not getting this facility due to some major problems. But we don't have enough researcher to find out those problems perfectly. In 2013 government nationalize the Registered Non-Government primary schools to govt. primary school. Most of

the non-govt. high schools are financed partially by the govt. directly. That's why students don't have to worry about being dropout due to educational expense. In the colleges, there are some lack of proper funding though. On the other hand, private universities get nothing from the government. Some of the students have to support themselves financially which hampers their study too. So in this sector financial problem is a great problem to continue the study. In primary, high-school and college the number of drug addicted students is comparatively low. Most of the students are under their guardian's supervision. But in the university they are free and become addicted very easily which is the big reason of dropout. Still our country doesn't have quality drug rehab. One another important reason is low quality high school and college education. This is a big reason, which stops a student to go with the flow in the university. They can't match with the university education due to lack of prior knowledge. All this thing has to find out if we want to solve the dropout problem. Still govt. or private institutions didn't take any necessary steps to do a research on this important topic. That's why it's necessary to go through a study and find out the possible solutions for this great problem.

1.4. Objectives

1.To find out the main reasons of student's dropout.

2. To find out the effect university and education system for dropping out.

3.To Find out the future or career effect of student

4.To find out the way to reduce the student's dropout

1.5. Research Questions

1.What are the reasons of student dropouts?

2.What are the effects of student dropouts on University and Education System?

3. What are the effect of student's future career?

4.What are the ways to reduce student dropouts?

1.6. Research Layout

Chapter 1: we will discuss about introduction, motivation, Problem Definition, Research Question, Research Methodology and the expected outcome of our project.

Chapter 2: we will discuss about background of this research and the related work and current status based on Bangladesh.

Chapter 3: it is focus to analysis of Methodology of Dropout.

Chapter 4: it is focus to the result and discussion of Dropout.

Chapter 5: Here concluded all the analysis and its future work.

CHAPTER 2

Background

2.1.Introduction

In the fundamental law of Bangladesh, Education is the basic needs for a nation and the obstacle is drop out. Well number of cases about dropout in the graduation level. Respect the consequence, Bangladeshi government took different enter pristineness of the growth of education system. University education is exigently essential to every citizen as primary and secondary education. For the socio-economic alteration and progress of a country, education is the basic need. It is the central component of human resource development.

2.2.Related Works

a. Khawar Shakeel and Naveed Anwer Butt model aim is to be identified the weak and risk student for dropout by measuring the students' performance through data mining Decision tree and Bayes algorithms. The data was collected from BS students of their university enrolled of year 2011 who were already completed two semesters. Data is concerning to student's previous academics, core subjects studied, student aptitude in specific assessment type, favorite subjects of students. Comparing the performance of Decision Tree and Bayes algorithms, it was concluded that Naive Bayes achieved highest accuracy of 91.9355% using our educational dataset.

b. Latif, Choudhary AI and HammayunAA The main purpose of this study is to explore the causes of students drop outs and their impact on economy. Financial problems, parents' unwillingness, distance and lack of basic facilities, bad quality of the education, inadequate school environment and building, overloaded class rooms, improper languages of teaching, carelessness of teachers and security problem in girl's school are found as major causes of student dropouts in different countries.

c. Sunil kumar and Ashok kumar based on a survey work which proposes to apply association rule mining measures like support confidence and other

interesting measures on these major factors of school failures to understand the problem in a better way and to have a proper planning for the academicians.

d. Dr. Saurabh Pal aimed to predictive their first year dropout students of engineering. Classification method and Bayesian classification method are used to evaluate previous year's student dropout data. Collected variable was marks in High School, marks in Senior Secondary, student's family position from the student's management system, to predict list of students who need special attention. Used different attribute table for predict drop out accuracy rate using Naïve Bayesian classification Algorithm.

e. Gerben W. Dekker, Mykola Pechenizkiy and Jan M. Vleeshouwers discuss the educational data mining case study aimed at predicting the Electrical Engineering (EE) students drop out after the first semester. They selected a target dataset of 648 students who were in their first year phase at the department and came either from VWO (which is pre-university secondary education) or from polytechnic education. In their experimental study they used different type of method and classification like CART,Bayes classifier, J48,JRip,Random firestone. The accuracy rate of the result was between 75%-80%.

f. Miguel Gil, Norma Reyes, María Juarez, Emmanuel Espitia, Julio Mosqueda and Myriam Soria focused on student's dropout of university using Artificial Neural Network. they analyzed and compared to know how the factors affect to the model behavior and the predicted result.

2.3.Bangladesh Perspective

The education quality of Bangladesh is developing day by day but the quantity of graduate students is still not at a satisfactory level. The enrolment of students in universities is only 6.6%. (Nation Master,2015). By the involvement of some NGOs primary education is now almost able to touch the expected margin. But after completing primary education a huge number of village students drop out of

primary school. Moreover, the same thing is happening for high school, college and university students in Bangladesh. From the perspective of Bangladesh if we make a shortlist of the reasons for dropping out from universities, then the reason comes first is drug addiction. The second reason is poverty or economic problem. And the next thing is the imbalance of versions of education among schools, colleges and universities. After going through a survey, it is very clear that drug addiction is one of the common reasons for university dropouts. Drug addiction hampers the attention of the students and at the end of the day it results in dropout. Compared to primary and high school education, university education is expensive. Some dropouts occur due to poverty or other economic reason. Imbalance of versions of education is responsible for dropout in some cases. Almost 97% of schools and colleges are Bengali medium in Bangladesh. But university education is English medium. It's very hard for a Bengali medium student to feel comfortable with the new version of education. As a result, some students give up study and dropout from universities. So the dropout percentage must be regulated to make the education sector secure.

Table 2.3.1: Students' Enrollment Rate in a Few Countries

Name of the country	Enrollment Rate in tertiary Education	Academic Year
United States	72.6%	2000
Australia	63.3%	2000
India	10.5%	2002
Bangladesh	6.6%	2000
Nepal	4.6%	2000
Malawi	0.3%	2002

Source: Nation Master, 2015

This statistical report prepared by UNESCO, here shows that the enrollment rate in tertiary education in Bangladesh 6.6%. Create huge stare of establishment need in tertiary system.

2.4.Scope of the study

Dropout research should have done for all the institutions in Bangladesh but due to some limitations, this research based on only Daffodil International University. Research data is collected from that university student only. However, it did not focus on the school or college level dropout problems in detail. Being a micro-level study that analyses the problems of university dropout students only. This study focuses on only the financial, educational, environmental and drug addiction problem of students. However, there are many reasons left but this short study will help the system to find out the core reasons of dropout. Which have a largest impact on overall education system. It can be used as a tool of improving the education quality of universities in future.

2.5.Government Goals and Regulation

The government finances public universities of Bangladesh but the Board of Trustees of the respective universities finances the private universities. Because of personal funding students have to pay a lot for getting education from private universities. In some cases, students are unable to continue their studies in private universities due to lack of funding, which is one of the big reason for dropout. To resolve this problem government of Bangladesh offers some scholarship for undergraduate and postgraduate students. UGC observes the education quality and overall activities of universities regularly to maintain the quality education. However, the government of Bangladesh is trying to develop a healthy environment of study and engage more students in higher studies.

CHAPTER 3

Methodology

3.1.Introduction

Success of a university mainly depend on the percentage rate of dropout student. For this study, information was collected through a questionnaire from 143 students of undergraduate courses of CSE department.

Predicting the status of a student is dropout whether he/she continue to their study or not. For the effective prediction, needs a huge amount of variable such as personal, academic, family, social, background of student, performance and so many.

This information will be helpful for the University to reduce the dropout student because it is related to classify the many quantitative factors to find the main causes of dropout which is to be part of the process of KDD and data mining.

3.2.Research subject and instrumentation

Research subject is prediction of dropout students using data mining.

Software: Weka Tools, Excel Sheet, Notepad++, Google form

3.3.Data Collection procedure

Problems are found through

1. Survey questionnaires,
2. Personal interview and
3. Academic.

3.4.Work Methodology

Achieving the main object, the steps were followed (Fig 3.4.1):

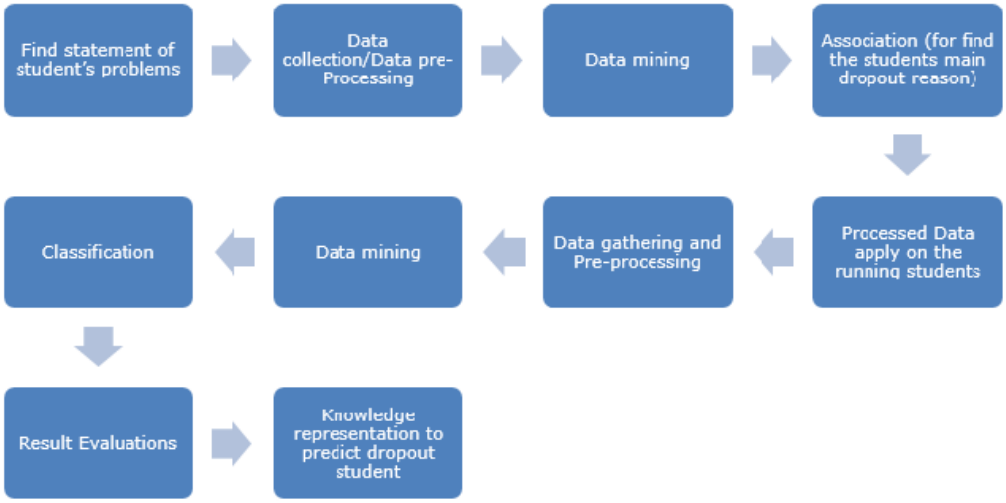


Figure 3.4.1: Work methodology

3.5.Find the dropout students problems

Problems are found through survey questionnaires, personal interview and academic only from the students who are dropped out. The questionnaire has been constructed based on the personal, family and academic problems such as personal (Illness, homesickness, facing depression, Change the future goal, no interest to study, Married/relationship, Financial problem, Learning problem, addicted with bad habits, engaged with other activities, Lower result/ CGPA, move to other university, living problem, did not adjust with environment), Family (Less support from mother and father, less financial support from family, family did not want to continue study, Lower family status, Far from family, Suddenly with any accident (death any family member)), Academic (Dislike University/Campus, Missed too many classes, Lengthy course, Too many rules, Course fee is too high, Dislike classes ,Did not understand teachers lecture, Bad understanding with teachers, Low placement rule, Did not understand about course, Move to other department, Credit transfer ,Distance from residence, Transportation problem ,Unwell performance in the class).

Through this problem statements, will find the main reason of dropping out which will be helpful data pre- processing.

3.6. Data Collection

Data are collected from the students who are already faced drop out. Number of the collected sample are 100 based on their age, semester, gender, types of dropout, causes of dropout (personal, family, academic). Table presented the data format.

Table 3.6.1: Dropout students related variables

Variables	Description	Values
Gender	gender	{Male, Female}
Age	age	{18-20,21-23,24-26}
DSemester	Dropout semester	{3-6,7-9,9-12}
TOD	Types of dropout	{SemesterDropout, CourseDropout, permamentDropout}
COD	Causes of dropout	{Personal,Family,Academic}
TOPP	Types of personal problems	{Illness,homesickness,Facingdepression,Changethefuturegoal,Nointeresttostudy ,Married/relationship,Financialproblem,Learningproblem,Addictedwithbadhabits ,Engagedwithotheractivities, Lowerresult/CGPA,Movtootheruniversity ,Livingproblem,Didnotadjustwithenvironment}
TOFP	Types of family problems	{LesssupportfromMother,LesssupportfroFather,Lesssupportfromboth, Lessfinancialsupportfromfamily,familydidnotwanttocontinue study, Lowerfamilystatus,Farfromfamily,anyaccident}
TOAP	Types of Academic problems	{DislikeUniversity/Campus,Missedtoomanyclasses, Lengthycourse, Toomanyrules, Coursefeeistoohigh, Dislikeclasses , Didnot understandteacherslecture, Badunderstandingwithteachers, Lowplacementrule, Didnotunderstandaboutcourse, Movetootherdepartment,Credittransfer, Distancefromresidence, Transportationproblem , Unwellperformanceintheclass}

Table 3.6.2: the data collection format.

Factors/variables	Description	Values
1. Age	Age of student	{18-20,21-23,24=<}
2. Semester	Current semester	{1-3,4-10,11-12}
3. Credit	Total no.of credit they are completed	{1-40,41-124,125-148}
4. Gender	Gender	{Male,Female}
5. EcoCondition	Economic condition	{Poor,Medium,High}
6. MaritalStatus	Marital status	{Single,Relationship,Married}
7. Member	Total family members	{1-4,5-8}
8. backMedium	Background medium	{Bangla,English}
9. Income	Family income	{20-50,51-80,>80}
10. FOccupation	Fathers occupation	{GovtJob,PrivateJobholder, Business,Retired/Nothing, Agriculture}
11. MOccupation	Mother occupations	{Housewife, Jobholder, Business}
12. Genvironment	Growing up environment	{UrbanArea,RuralArea}
13. SOccupation	Self-occupation	{OnlyStudent,Studentwithjob, StudentWithTution}
14. SubInterested	Interested in subject	{Yes,No,Average}
15. Lecunderstand	Understand the lecture	{Yes,No,Average}
16. adjustFriend	Adjust withclassfriends	{Yes,No,Average}
17. completeClasswork	Complete class work	{Yes,No}
18. missedClass	Missed the no. of classes	{0-5,6-10,10-15,16-20}
19. self-confidenceINstudy	Self confidence	{Yes,No}

	in study	
20. supportFroMteachers	Support from teachers	{Yes,No}
21. Goalconfident	Goal confident	{Yes,No}
22. likeClassesORUniversity	Like classes or university	{Yes,No}
23. engageActivity	Engaged with other activities	{Yes,No}
24. SpendSocialMediabad	Spend time in social media	{0-2,3-5,6-10}
25. BadAddiction	Addicted with bad habits	{Yes,No}
26. URcontribution	Contribution in family	{Yes,No}
27. Healthproblems	Faces health problem/illness	{Yes,No}
28. campusEnvironment	Like campus environment	{Yes,No}

Secondly, data are collected from the running students to predict student who will drop out or not in future based on survey questionnaires, personal interview and online survey. Students are performed 28 factors based on first data sets (main factors collect from drop out student) and add with some addition factors such as Gender, age, semester, credit, marital Status, family member, background medium, Income, father occupations, mother occupations, subject interest, lecture understand, adjust with friends, like university campus, bad addiction, spend time in social media, goal confident, support from teachers, missed classes, complete class work etc. This data is essential to find the student for dropping out.

3.7.Data Selection/Data pre-processing

After collection of all data, the datasets were ready to apply the techniques of data mining. Data was applied to measure the quality and suitability of data before prescribing the model of data mining. In this steps only those attributes were

selected which were more essential for data mining and removed the irrelevant attributes, missing values, noisy data, parameters.

The irrelevant data from first datasets were change the future goal, move to other university, engaged with other activities, family did not to continue education, less family support from mother, less family support from father, less family support from both, course fee is too high, too many rules, low placement of rule, transportation problem. From Second datasets removed all unnecessary data which was not more effective for the prediction.

3.8.Techniques of data analysis

In this study, qualitative and quantitative data analysis techniques have to be used for statistical and data mining methods.

Statistical methods:

Excel sheet are used for statistical analysis.

Data mining methods:

Pre-processing data were applied to measure appropriate factors using data mining methods. Classification and decision tree were applied to predict student drop out causes and rates in early stage of their study either before or after completing of their 3rd semester. For predict appropriate factors, WEKA tools was used for having many tools like C4.5 which is extension of ID3 and Naive bayes. Important attributes were ranked by using information gain. Correlation based Feature Selection using BFS(Best First Search) algorithm to select useful attributes. On the other hand Association rules was used to analysis the main factor which occurred the dropout. Apriori algorithms was for the Association rules.

Correlation-Based Feature Selection:

Feature selection is used to select a subset of input data most useful for analysis and future prediction by eliminating features, which are irrelevant of no predictive information. Features selection is use for increasing the predictive accuracy and reducing complexity of learner results. In present study Correlation-Based Feature Selection (CFS) was used to find the feature subsets that are highly correlated

with the class but minimal correlation between features combined with search strategy best-first search (BFS). Best First Search method starts with empty set of features and generates all possible single feature expansions.

C4.5:

C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan. C4.5 is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier.

Naive Bayes:

It 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.

Association rule mining:

Association rule mining is a procedure which is meant to find frequent patterns, correlations, associations, or causal structures from data sets found in various kinds of databases such as relational databases, transactional databases, and other forms of data repositories.

CHAPTER 4

Result and Discussion

4.1. Introduction

In this study, data was analyzed through the Classification and Association rules techniques of data mining. For the purpose of predicting students who would be dropped in future or not, 28 attributes were applied on 143 Instance where 42 attributes were applied on the 64 instance who are already dropped out to find out the main reasons of dropping out. Apriori algorithm of association rule was used for generate the rules of drop out. And C4.5 and decision tree were used for predict student drop out in future.

4.2. Experimental Analysis

Data were analyzed both statistically and data mining. Statistical analysis can visualize the data response graphically where data mining process is used for analyzed the data logically.

4.3. Statistical Analysis

In this study demographic issues were explored through Age, gender, semester, economic conditions, marital status, Family member, background medium, Family income, Father Occupation, Mother Occupation, Growing up environment, Self Occupation ,Subject Environment, Lecture Understanding level, Adjust Friend, complete class work, number of missed classes, self-confidence level, engage with other activities, Goal confident , spend time in social media, bad Addiction, Contribution in family, Dislike university on the running students. The maximum responses students ages were 21-23 years and 66.43%, 81.11% of male,81.11% of 4-10 semester, 71.32% of small family members, 35.66% and 86.72% of business and housewife parents, 67.78% of students, 81.11% of students are single,58.04% students growing up area was urban, 64.33% of students were interested in their subjects, 48.25% of average understand class lectures, 71.32% adjust with their friends,51.04% students did not complete class work, 58.74% of students missed the number of classes were 0-5,63.63% got support from their teacher,73.42% were confident with their

goal,58.04% students did not engage with other activities,50.34% students spent time in social media were 3-5 hours in a day, 86.72% had no health problem or illness, 53.14% students dislike university.Table4.3.1 show the percentage of these variables responses:

Table 4.3.1: Response the student's performance

Variable	Particulars	Frequency(No. of students)	Percentage
Age	18-20	17	11.89
	21-23	95	66.43
	24=<	31	21.68
Gender	Male	116	81.11
	Female	27	18.88
Semester	1-3	24	16.78
	4-10	116	81.11
	11-12	3	2.09
Economic condition	Poor	14	9.79
	Medium	126	88.11
	High	3	2.09
Family members	1-4	102	71.32
	5-8	41	28.67
Fathers Occupation	Govt.job	25	17.48
	Private job holder	25	17.48
	Business	51	35.66
	Retired/nothing	31	21.67
	Agriculture	11	7.69
Mother Occupation	Housewife	124	86.72
	job holder	17	11.88
	Business	2	1.4
Self-Occupation	Only student	110	67.92

	Student with job	0	0
	Student with Tuition	33	23.07
Marital status	Single	116	81.11
	Relationship	25	17.48
	Married	2	1.3
Growing up environment	Urban	83	58.04
	Rural	60	41.95
Subject in interested	Yes	92	64.33
	No	11	7.69
	Average	40	27.97
Lecture Understand	Yes	62	43.35
	No	12	8.39
	Average	69	48.25
Adjust with friends	Yes	102	71.32
	No	6	4.19
	Average	35	24.47
Complete classwork	Yes	70	48.95
	No	73	51.04
No. of missed classes	0-5	84	58.74
	6-10	1	0.69
	10-15	47	32.86
	16-20	11	7.69
Self confidence in study	Yes	110	76.92
	No	33	23.07
Support from teacher	Yes	91	63.63
	no	52	36.36
Goal confident	Yes	96	67.13
	No	47	32.86
Like classes or university	Yes	105	73.42

	No	38	26.57
Engage other activity	Yes	60	41.97
	no	83	58.04
Spend time in social media	0-2	40	27.97
	3-5	72	50.34
	6-10	31	21.67
Bad addiction	Yes	45	31.46
	No	98	68.53
Contribution in family	Yes	47	32.86
	No	96	67.13
Health problem/illness	Yes	19	13.28
	No	124	86.72
Dislike university	Yes	67	46.85
	No	76	53.14

The number of respondent dropout students were 64 and they responded on 42 variables. The main factors for dropout were related to personal, family and academic problems and their respondent percentage were 40.90%, 30.30% and 28.78% respectively. The maximum factors from the personal problems were facing depression, financial problem and lowest cgpa with 22.22% and 20.04% of illness and go to abroad (show in figure 4.3.2). From the family problem maximum response were from less financial support from family, accident and distance from home with 43.9%, 31.7% and 26.8% respectively (show in figure 4.3.1). The maximum causes from the academic problems were transportation problem with 40% response and missed many classes with 35% response (show in figure 4.3.3). Table 4.3.2 show the response from dropout students.

Table 4.3.2: Response of dropout student and their reasons

Variable	particular	Frequency	Percentage
Age	18-20	10	15.6

	21-23	40	62.5
	24-26	14	21.9
Gender	Male	37	57.8
	Female	27	42.2
Dropout Semester	1-3	2	3.125
	4-6	34	53.12
	7-9	16	25.00
	10-12	12	18.75
Types of dropout	Personal	54	40.90
	Family	40	30.30
	Academic	38	28.80

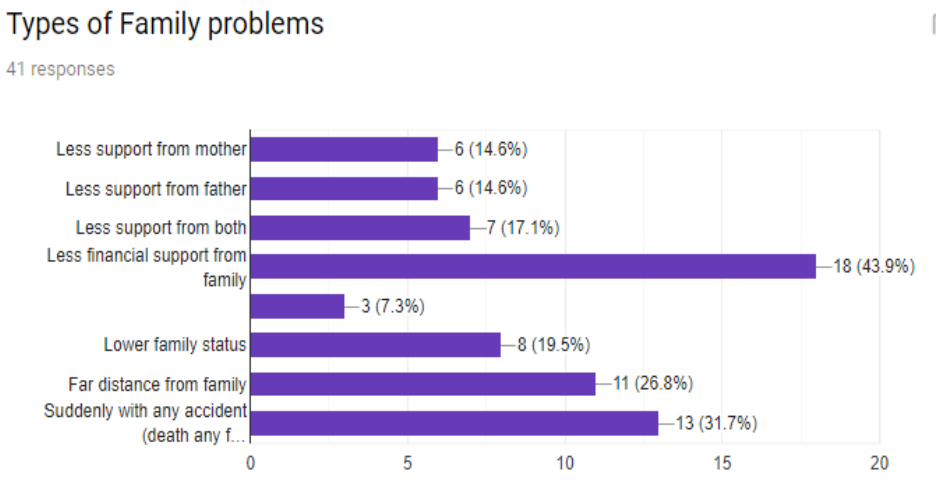


Figure 4.3.1: Types of Family problems

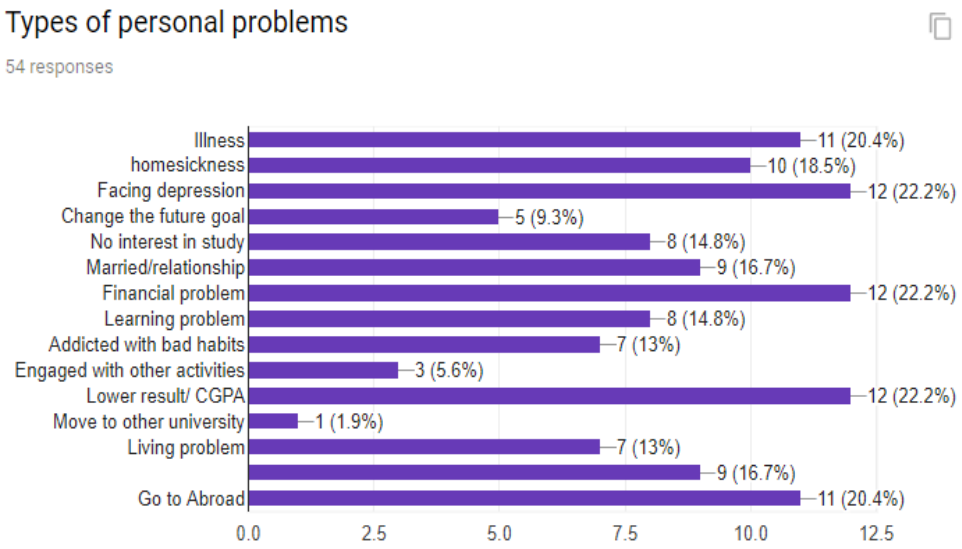


Figure 4.3.2: Types of personal problems

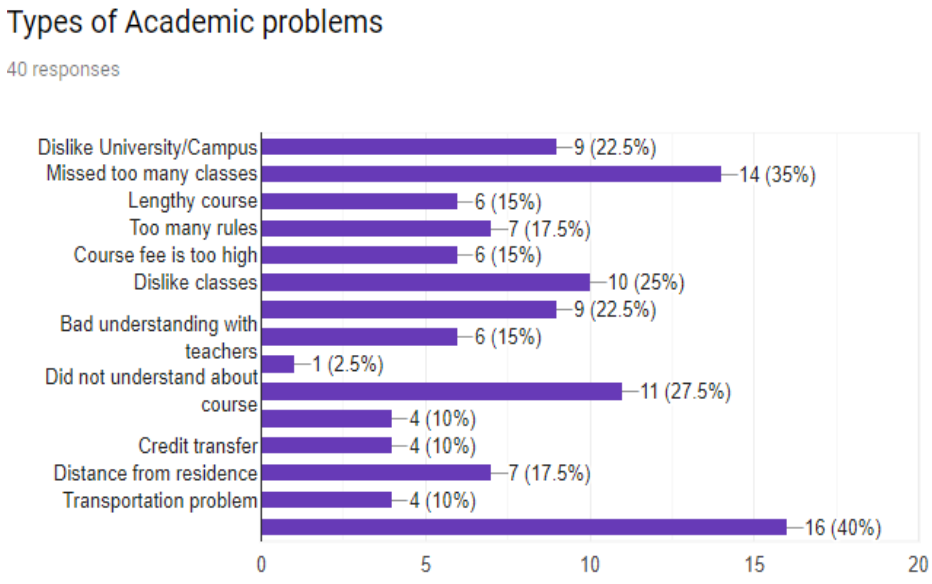


Figure 4.3.3: Types of academic problems

Overall observation, 8 variables were selected out of 28 variables for predictions such as income, self-Occupation, subject in interested, lecture understand, missed class, like classes or university, engage with activity, dislike university.

4.4. Data mining Analysis

First, selected highly correlated feature which are associated with dropout student to design classification model for future prediction of student's dropout or not. At this stage, generate the KDD 8 most representative attributes were selected for classification

The aim of this study is to select a highly correlated feature which are associated with dropout student and design a classification model for future prediction of whether student will drop the course or continue and study the cause of dropout students. At this stage, in order to generate knowledge, the 8 most representative attribute were selected from database based on correlation feature selection method .

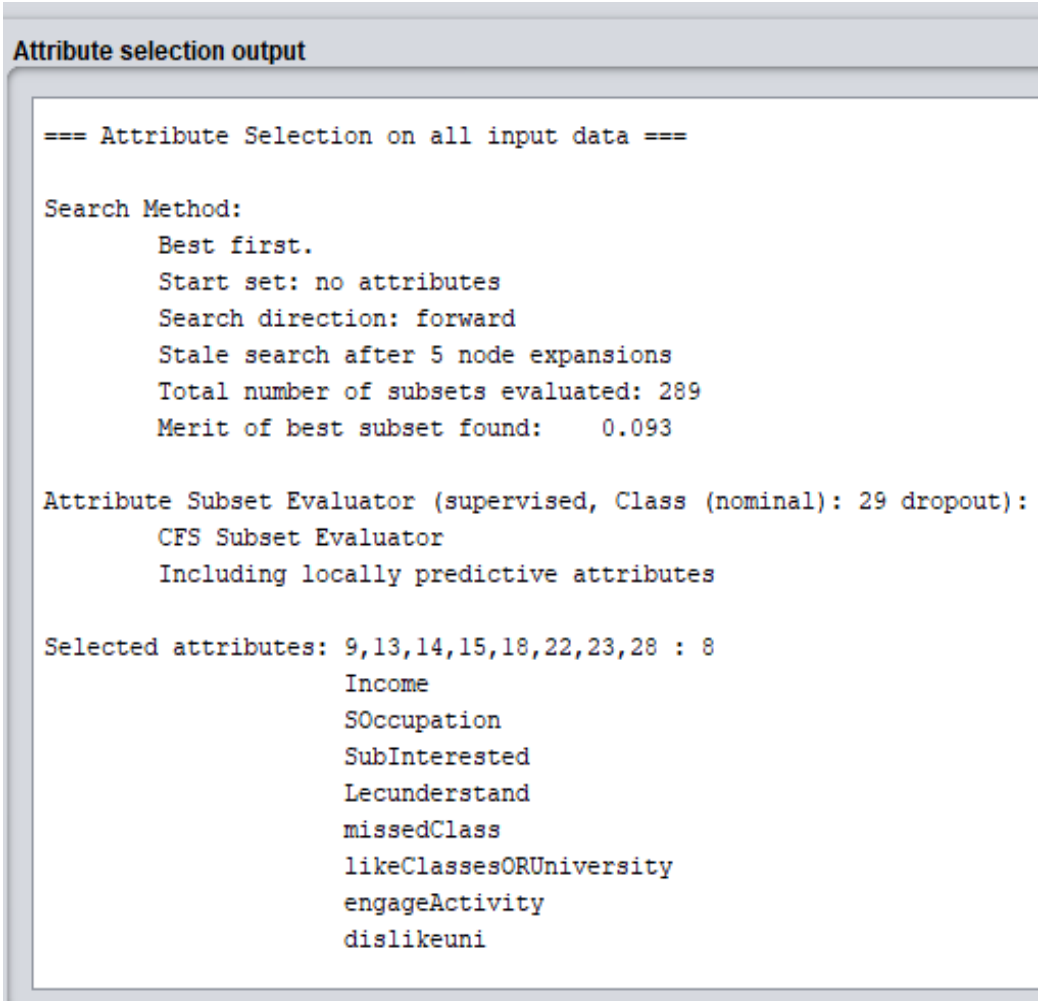


Figure 4.4.1: Attribute Selected Using Correlation Feature Selection Background of C4.5

Through csf technique with bfs algorithm 8 attribute were selected which are income, self-Occupation, subject in interested, lecture understand, missed class, like classes or university, engage with activity, dislike university. Then C4.5 algorithms was applied and decision was made using information gained ranked. The highest information gain attribute used as root node and this process was used until all attributes were perfectly classified. Table 4.4.1 show the ranked of information gain of 8 attributes. Then decision table was made using c4.5 algorithms show in figure 4.4.2

Table 4.4.1: Attributes ranked with respect to information gain

Information Gain	**Attributes**
0.131	LecUnderstand
0.112	engageActivity
0.098	SubInterested
0.096	LikeClassorUni
0.093	SOccupation
0.091	dislikeUni
0.087	missedClasses
0.081	income

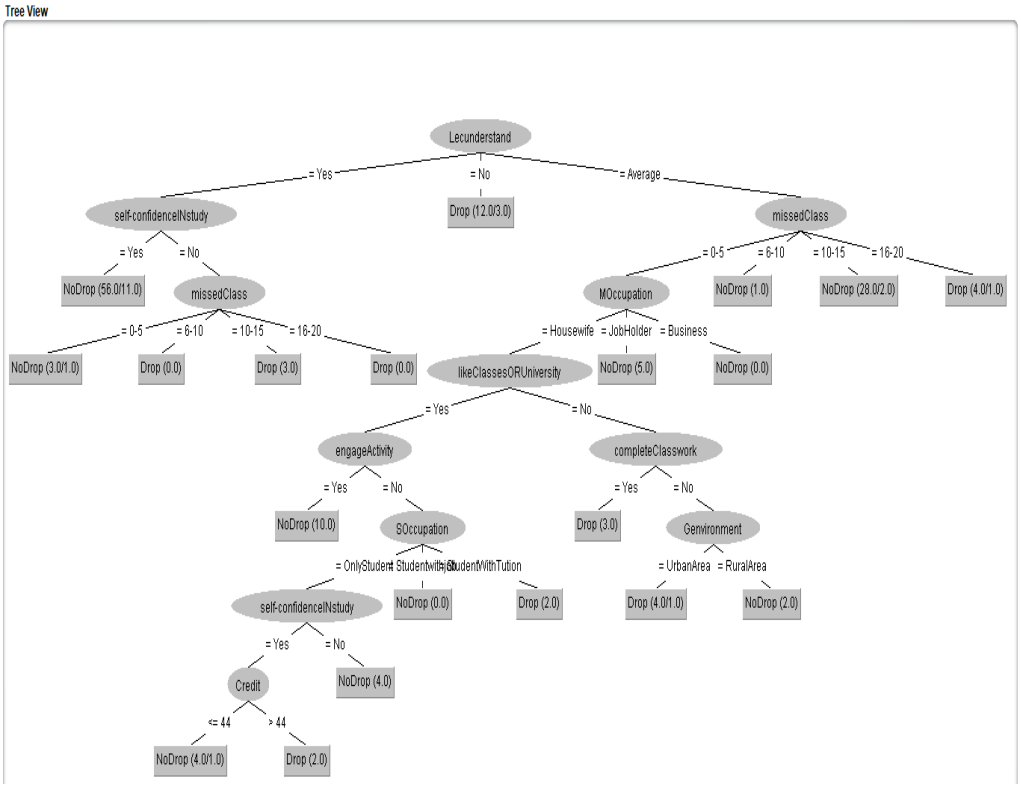


Figure 4.4.2: Decision Tree Using c4.5 Algorithm

After tree construction and confusion matrix, evaluation parameters such as Recall, Measure, Precision and Accuracy are calculated shown in Table 4.4.3.

Table 4.4.2: Results for the C4.5 Decision Tree Algorithm Using training set (Accuracy by Class)

Accuracy	Class	
	Drop	No drop
TP Rate	0.625	0.951
FP Rate	0.049	0.375
Precision	0.833	0.867
Recall	0.625	0.951
F-Measure	0.714	0.907

MCC	0.635	0.635
ROC Area	0.868	0.868
PRC Area	0.729	0.919

Table 4.4.3: Confusion matrix of c4.5 algorithm

		Predicted classes		
		No Drop	drop	Total
Actual class	No Drop	25 (TN)	15 (FP)	40
	Drop	5 (FN)	98 (TP)	103
	Total	30	113	143

$$Correctly\ classified\ instance\ (accuracy) = \frac{TP + TN}{TP + TN + FP + FN} =$$

$$\frac{98 + 25}{25 + 98 + 5 + 15} * 100 = 86.0139\%$$

$$Incorrectly\ classified\ instances = \frac{FP + FN}{N} = \frac{15 + 5}{143} * 100 = 13.98\%$$

$$Misclassification\ rate(mean\ absolute\ error) = \frac{FN + FP}{N} = \frac{5 + 15}{143}$$
$$= 0.1398$$

$$Sensitivity(Recall) = \frac{TP}{TP + FN} = \frac{98}{98 + 5} = 0.951$$

$$Specifity = \frac{TN}{FP + TN} = \frac{25}{15 + 25} = 0.625$$

$$Percision = \frac{TP}{TP + FP} = \frac{98}{98 + 15} = .8672$$

$$F - measure = \frac{2 * Recall * Precision}{Recall + Precision} = \frac{2 * 0.951 * 0.8672}{0.951 + 0.8672} = 0.907$$

Table 4.4.4: Classifier Rules

If LectureUnderstand=No THEN Dropout=**drop**
IF LectureUderstand= Yes AND Self-confident in study=Yes THEN Dropout=Nodrop
IF LectureUderstand= Yes AND Self-confident in study=No AND missedclasses=0-5 THENDropout=NoDrop
IF LectureUderstand= Yes AND Self-confident in study=No AND missedclasses=6-10 THEN Dropout=**Drop**
IF LectureUderstand= Yes AND Self-confident in study=No AND missedclasses=10-15 THEN Dropout=**Drop**
IF LectureUderstand= Yes AND Self-confident in study=No AND missedclasses=16-20 THEN Dropout=**Drop**
IF LectureUderstand= Average AND missedclasses=0-5 AND MotherOccupation=Housewife AND LikeclassesOrUniversity=Yes AND EngageotherActivity=Yes THEN Dropout=Nodrop
IF LectureUderstand= Average AND missedclasses=0-5 AND MotherOccupation=Housewife AND LikeclassesOrUniversity=Yes AND EngageotherActivity=No AND SelfOccupation=OnlyStudent AND Self-ConfidenceInStudy=Yes AND Age=22>= THEN Dropout=Nodrop
IF LectureUderstand= Average AND missedclasses=0-5 AND MotherOccupation=Housewife AND LikeclassesOrUniversity=Yes AND EngageotherActivity=No AND SelfOccupation=OnlyStudent AND Self-ConfidenceInStudy=Yes AND Age=22<= THEN Dropout=**Drop**
IF LectureUderstand= Average AND missedclasses=0-5 AND MotherOccupation=Housewife AND LikeclassesOrUniversity=Yes AND EngageotherActivity=No AND SelfOccupation=OnlyStudent AND Self-ConfidenceInStudy=No THEN Dropout=Nodrop
IF LectureUderstand= Average AND missedclasses=0-5 AND MotherOccupation=Housewife AND LikeclassesOrUniversity=Yes AND EngageotherActivity=No AND SelfOccupation=StudentWithJob THEN Dropout=Nodrop
IF LectureUderstand= Average AND missedclasses=0-5 AND MotherOccupation=Housewife AND LikeclassesOrUniversity=Yes AND EngageotherActivity=No AND SelfOccupation=StudentWithTution THEN Dropout=**Drop**
IF LectureUderstand= Average AND missedclasses=0-5 AND MotherOccupation=Housewife AND LikeclassesOrUniversity=No AND

CompleteClassWork=Yes THEN Dropout=**Drop**
IF LectureUderstand= Average AND missedclasses=0-5 AND MotherOccupation=Housewife AND LikeclassesOrUniversity=No AND CompleteClassWork=No AND GrowingUpEnvirnment=Urban THEN Dropout=**Drop**
IF LectureUderstand= Average AND missedclasses=0-5 AND MotherOccupation=Housewife AND LikeclassesOrUniversity=No AND CompleteClassWork=No AND GrowingUpEnvirnment=Rural THEN Dropout=**Drop**
IF LectureUderstand=Average AND missedclasses=6-10 THEN Dropout=Nodrop
IF LectureUderstand= Average AND missedclasses=10-15 THEN Dropout=Nodrop
IF LectureUderstand= Average AND missedclasses=16-20 THEN Dropout=**Drop**

By using classification rules, can easily improve the students monitoring planning and reduce the situation who have a high probability of dropping out from the university. This rules can easily predict the students that they will continue their study or not.

Naive Bayes result summary:

Naive Bayes algorithm mainly works based on calculating the probability of the attributes.

Table 4.4.5: Attributes were calculated with respect to Probability

Attributes	**Values**	**Probability**
LecUnderstand	average	0.395
engageActivity	no	0.744
SubInterested	yes	0.581
LikeClassorUni	yes	0.571
SOccupation	only Student	0.860
dislikeUni	no	0.651
missedClasses	0-5	0.5

income	20-50	0.627

Table 4.4.6: Results for the Naïve bayes Algorithm Using training set (Accuracy by Class)

Accuracy	Class	
	Drop	No drop
TP Rate	0.500	0.864
FP Rate	0.136	0.500
Precision	0.588	0.817
Recall	0.500	0.864
F-Measure	0.541	0.840
MCC	0.384	0.384
ROC Area	0.794	0.794
PRC Area	0.620	0.916

Table 4.4.7: Confusion matrix of Naïve bayes algorithm

		Predicted classes		
		No Drop	drop	Total
Actual class	No Drop	20 (TN)	20 (FP)	40
	Drop	14 (FN)	89 (TP)	103
	Total	34	109	143

$$Correctly\ classified\ instance\ (accuracy) = \frac{TP + TN}{TP + TN + FP + FN}$$
$$= \frac{89 + 20}{89 + 20 + 20 + 14} * 100\ = 76.2238\%$$

$$Incorrectly\ classified\ instances = \frac{FP + FN}{N} = \frac{20 + 14}{143} * 100 = 23.7762\%$$

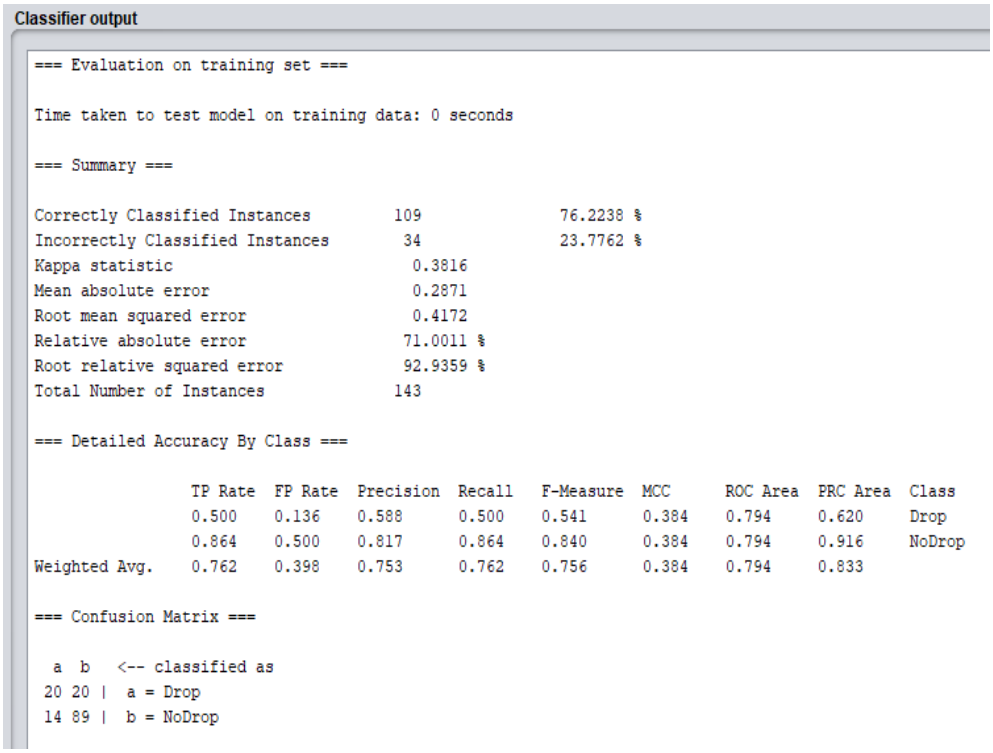
$$Misclassification\ rate(mean\ absolute\ error) = \frac{FN + FP}{N} = \frac{20 + 14}{143}$$
$$= 0.2377$$

$$Sensitivity(Recall) = \frac{TP}{TP + FN} = \frac{89}{89 + 14} = 0.864$$

$$Specifity = \frac{TN}{FP + TN} = \frac{20}{20 + 20} = 0.500$$

$$Percision = \frac{TP}{TP + FP} = \frac{89}{89 + 20} = .817$$

$$F - measure = \frac{2 * Recall * Precision}{Recall + Precision} = \frac{2 * 0.864 * 0.817}{0.864 + 0.817} = 0.840$$



Classifier output

```
=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances         109               76.2238 %
Incorrectly Classified Instances        34               23.7762 %
Kappa statistic                          0.3816
Mean absolute error                      0.2871
Root mean squared error                  0.4172
Relative absolute error                 71.0011 %
Root relative squared error             92.9359 %
Total Number of Instances              143

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.500    0.136    0.588      0.500    0.541      0.384    0.794     0.620     Drop
               0.864    0.500    0.817      0.864    0.840      0.384    0.794     0.916     NoDrop
Weighted Avg.  0.762    0.398    0.753      0.762    0.756      0.384    0.794     0.833

=== Confusion Matrix ===

  a  b   <-- classified as
 20 20 |  a = Drop
 14 89 |  b = NoDrop
```

Figure 4.4.3: Find the accuracy using Naive Bayes algorithms

Now, table 4.4.4 show the probability of student's dropout attributes such as self-occupation, engage activity, income, dislike university, subject interested, like classes, miss classes, lecture understand. Using the training data set Naive Bayes algorithms found 76.22% accuracy in weka software and also from the confusion matrix show in the table 4.4.6.

4.5. Application of Association Rule

In this study data was accumulated from the students who discontinue their studies. These data were analyzed using Association Rule Mining to find out the causes or factors behind their decision to discontinue their education at the university. In order to achieve the objective following steps were performed in sequential order:

Data Set

Under data set, the reason provided by the students for dropping out of the ICT courses at university level were divided into four factors such as family problem,

health related, personal problem and institutional problem listed in Table 4.5.1. The size of dataset was 41.

Table 4.5.1: Dataset with Description of Attribute and Possible Values

Factors	Description	Possible Values
Gender	Gender	{Male, Female}
Age	Age	{18-20,21-23,24-26}
DSemester	Dropout semester	{3-6,7-9,9-12}
TOD	Types of dropout	{SemesterDropout, CourseDropout, permamentDropout}
COD	Causes of dropout	{Personal,Family,Academic}
TOPP	Types of personal problems	{Illness,homesickness,Facingdepression ,Changethefuturegoal,Nointeresttostudy ,Married/relationship,Financialproblem ,Learningproblem,Addictedwithbadhabits ,Engagedwithotheractivities, Lowerresult/CGPA,Movtootheruniversity ,Livingproblem,Didnotadjustwithenvironment}
TOFP	Types of family problems	{LesssupportfromMother, LesssupportfroFather,Lesssupportfromboth, Lessfinancialsupportfromfamily, familydidnotwanttocontinue study, Lowerfamilystatus,Farfromfamily,anyaccident}
TOAP	Types of Academic problems	{DislikeUniversity/Campus,Missedtoomanyclasses, Lengthycourse, Toomanyrules, Coursefeeistoohigh, Dislikeclasses , Didnot understandteacherslecture, Badunderstandingwithteachers, Lowplacementrule, Didnotunderstandaboutcourse, Movetootherdepartment,Credittransfer, Distancefromresidence, Transportationproblem , Unwellperformanceintheclass}

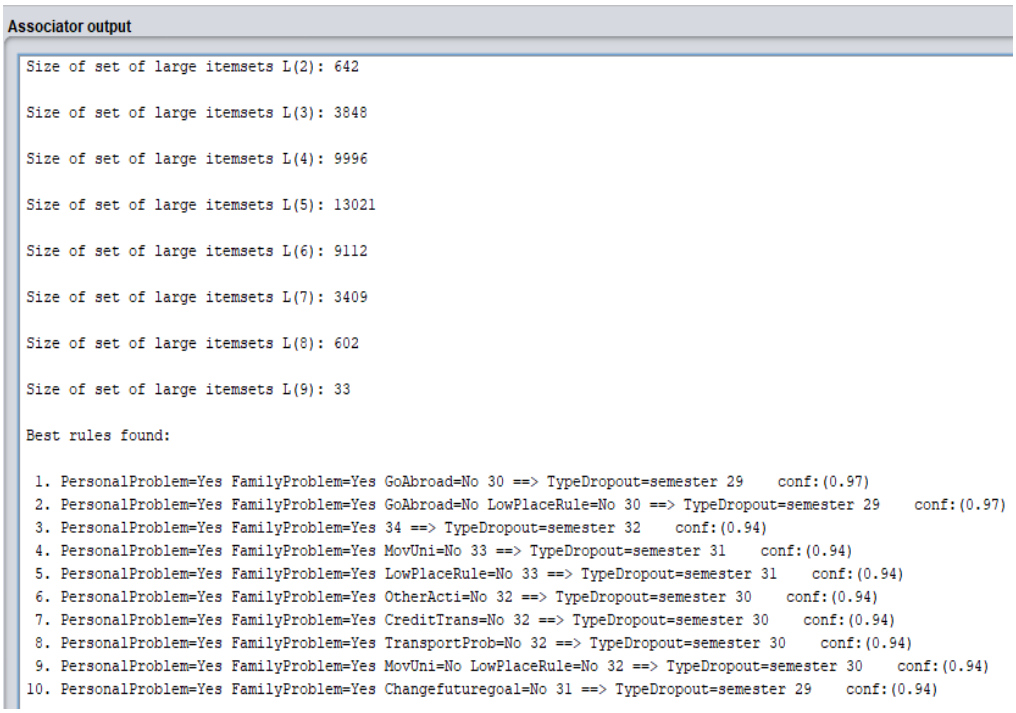


Figure 4.5.1: Select best rule using apriori algorithms

Above the figure which was implementation of Apriori algorithms of Association techniques was that maximum students were drop out for the personal and family problems with the conf: (0.97).

Over all, association rule mining was applied to find out the relationship of factors which affecting the students' dropout in the university. For prediction of future students drop out were calculated through C4.5 and Naive bayes algorithms. The calculation of C4.5 was based on information gain where Naïve Bayes was on probability. Using training data set C4.5 gave 86.014% whereas Naïve bayes 76.22% of accuracy.

CHAPTER 5

Conclusion and Future Work

Investigation of the major factor or reasons of students drop out in undergraduate students was purpose of the study. Through this reasons university authority could take proper action against this problem. This reasons can remove the student's dropout. Nowadays predicting students drop out is very important and difficult task for universities and educational sectors. Therefore, Data mining techniques were applied and this was very helpful for find out the main reasons of dropout.

The findings of student questionnaire were analyzed and interpreted. The computer software called Excel was used for the treatment of the collected data. Statistical techniques such as Frequency distribution for single variable, Discriminant analysis and Data mining techniques for application of decision tree (C4.5) and association rule have been used to study the causing factors for dropping out the students.

Based on the CFS (Correlation based feature selection) result, it was found that a student's dropout was affirmatively correlated with 8 response variables namely as income, self-Occupation, subject in interested, lecture understand, missed class, like classes or university, engage with activity, dislike university.

The association rule clearly specified that the students who have Personal problem were more chance to dropouts in comparison to Family problem and Academic problem. The analysis confidence represents that the maximum number of respondents having family problems suffered which was related to financial crisis and sudden accident. Besides that, few personal problems as facing depression, financial problem and lowest cgpa and illness and go to abroad were forced the students to discontinue their study.

Result indicates that C4.5 decision tree algorithm is best classifier with 86.014% accuracy. The generated information will be quite useful for management of university to develop policies and strategies for better planning and implementation of educational program and infrastructure under measurable condition to increase the enrolment rate in University and to take effective decision to reduce student dropout.

Future work is to study on large database of dropout student at the university using other data mining techniques such as Logistic Regression, Clustering and Neural Network in order to determine similarities and relationship between multiple factors.

Appendix

Appendix-1:

For Drop out students

Find out the student dropout reasons (only for dropout students):

Gender:

Age:

Semester (which semester your dropped):

Type of dropout: 1. semester dropout/temporary dropout

2.Course dropout

3.permanent dropout

Causes of dropout:

1.personal

2.family

3.educational

Types of personal problems:

Problems	Yes	No
Illness		
homesickness		
Facing depression		
Change the future goal		
No interest to study		
Married/relationship		
Financial problem		
Learning problem		
Addicted with bad habits		
Engaged with other activities		
Lower result/ CGPA		
Move to other university		
Living problem		

Types of Family problems:

Problems	Yes	No
Less support from mother		
Less support from father		
Less support from both		
Less financial support from family		
family did not want to continue study		
Lower family status		
Far distance from family		
Suddenly with any accident (death any family member)		

Types of University problems:

Problems	Yes	No
Dislike University/Campus		
Missed too many classes		
Lengthy course		
Too many rules		
Course fee is too high		
Dislike classes		
Did not understand teachers lecture		
Bad understanding with teachers		
Low placement rule		
Did not understand about course		
Move to other department		
Credit transfer		
Distance from residence		
Transportation problem		

Appendix-2 :

Running Students who can be drop out

Survey Question

Age :	Semester :

1. what is the percentage of credits have you completed?	25%	50%	75%	100%
2. Gender :	Male	Female		
3. Residence Type:	Mess / Hostel	Family	Relatives	
4. Economical Condition :	Poor	Medium	Rich	
5. Marital Status :	Married	Single	Relationship	
6. Number of Family Members :	1-4	5-8		
7. Education medium :	English	Bangla		
8. Average Monthly Income of the Family:	20k-50k	51k-80k	More than 80k	
9. Fathers' Occupation :	Govt job	Private Jobholder	Business	
	Nothing / Retired	Agriculture		
10. Mothers' Occupation :	Housewife	Job holder	Business	
11. Growing up environment:	Urban Area	Rural Area		
12. Your Occupation :	Only Student	Student with job		
	Student With Tution			
13. Are you interested in your subject ?	Yes	No	Average	
14. Can you understand your teachers' lecture?	Yes	No	Average	
15. Can you adjust with your friends or classmates?	Yes	No	Average	
16. Do you complete your classwork regularly?	Yes	No		
17. How many classes have you missed in last semester?	0-5	6-10	11-15	16-20

18. Have you enough self-confidence in your study?	Yes	No	
19. Are you getting enough support from your teachers?	Yes	No	
20. Still you confident about your goal?	Yes	No	Average
21. Do you like your classes or university?	Yes	No	Average
22. Do you engage with social extracurricular activities or cultural programs?	Yes	No	
23. How many time do you spend on social media?	0-2	3-5	6-10
24. Have you any bad addiction ?	Yes	No	
25. Have you any contribution to your family?	Yes	No	
26. Do you have any serious health problems ?	Yes	No	
27. Do you satisfied with your campus environment?	Yes	No	

☺☺ Thanks ☺☺

References

[1] KhawarShakeel and Naveed AnwerButt, Educational Data Mining to Reduce Student Dropout Rate by Using Classification",Conference Paper**:** May 2015.

[2]Latif A, Choudhary AI and Hammayun AA," Economic Effects of Student Dropouts: A Comparative Study" J Glob Econ 2015.

[3] Miguel Gil, Norma Reyes, MaríaJuárez, Emmanuel Espitia, Julio Mosqueda and Myriam Soria," Predicting Early Students with High Risk to Drop Out of University using a Neural Network-Based Approach",ICCGI 2013, The Eighth International Multi-Conference on Computing in the Global Information Technology.

[4] Dr. Saurabh Pal," Mining Educational Data Using Classification to Decrease Dropout Rate of Students", INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY SCIENCES AND ENGINEERING, VOL. 3, NO. 5, MAY 2012

[5] Satyarthi, K. (2009), 50pc Dropout at Primary Level in Bangladesh, The Daily Star, January, 2009

[6] IRIN, (2007), Bangladesh: High rate of school dropouts, online version available at: http://www.irinnews.org/Report.aspx?, Report Id=75937 (viewed on 5 May 2015)

[7] DU (2010), Office Record, Office of the Controller, University of Dhaka, Nilkhet, Dhaka-1000

[8] Islam, M.N. and Pavel, T. (2014), Factors Contributing Towards Dropouts at Undergraduate Level: An Analysis, *ASA University Review*, Vol. 5 No. 1

[9] BM Sajjad Hossain," Dropout at Tertiary Education in Bangladesh: Configurations and Determinants",Feni University Journal, Vol. 01, No. 01, August 2016, ISSN [2518-3869]

Prediction of student dropout Using Data Mining

ORIGINALITY REPORT

17%	%	%	17%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	2%
2	Submitted to Higher Education Commission Pakistan Student Paper	2%
3	Submitted to University of Nottingham Student Paper	2%
4	Submitted to Botswana International University of Science and Technology Student Paper	2%
5	Submitted to Universiti Utara Malaysia Student Paper	1%
6	Submitted to British University In Dubai Student Paper	1%
7	Submitted to Hellenic Open University Student Paper	1%
8	Submitted to Southern New Hampshire University - Continuing Education Student Paper	1%

9	Submitted to University of the East Student Paper	1%
10	Submitted to University of Kufa Student Paper	<1%
11	Submitted to University of Nizwa Student Paper	<1%
12	Submitted to Heriot-Watt University Student Paper	<1%
13	Submitted to Universiti Teknologi MARA Student Paper	<1%
14	Submitted to Associatie K.U.Leuven Student Paper	<1%
15	Submitted to National College of Ireland Student Paper	<1%
16	Submitted to Sheffield Hallam University Student Paper	<1%
17	Submitted to University of Wales Swansea Student Paper	<1%
18	Submitted to Central Queensland University Student Paper	<1%
19	Submitted to Institute of Technology Blanchardstown Student Paper	<1%
	Submitted to Northcentral	

20	Student Paper	<1%
21	Submitted to Laureate Higher Education Group Student Paper	<1%
22	Submitted to Middlesex University Student Paper	<1%
23	Submitted to University of Strathclyde Student Paper	<1%
24	Submitted to Birla Institute of Technology Student Paper	<1%
25	Submitted to UT, Dallas Student Paper	<1%
26	Submitted to Savitribai Phule Pune University Student Paper	<1%
27	Submitted to London School of Marketing Student Paper	<1%
28	Submitted to Universitas Islam Indonesia Student Paper	<1%
29	Submitted to University of Kent at Canterbury Student Paper	<1%