

USING DATA MINING CLASSIFICATION ALGORITHMS TO PREDICT STUDENTS' ACADEMIC PERFORMANCE



Prepared by : Zahraa Salih Mahdi AL-Yasiri
Supervised by :
Dec 2023

Table of Contents

01 Introduction

02 Express the
problem

03 Related work

04 Methodology

05 Implementation
and Evaluation

06 Conclusion and
Future Works

07 Reference

1- Introduction

Education for children, adolescents, young people and adults has become one of the most important focus points for many researchers. Because different educational means and educational assistance coincide with different teaching methods (such as online or offline or in person and ..) It has become common all over the world and people can choose one of these methods or devices and learn different knowledge from a distance or from their own home. One of the most important challenges is to improve the quality of education and improve the performance of students. Therefore, in this research, we will propose a new method to discover the pattern of successful or unsuccessful students in learning different sciences (such as specialized or general university courses).

2- Express The Problem

- Educational Data Mining (EDM) is a new trend in the field of data mining and knowledge discovery in databases (KDD) that emphasizes the extraction of useful patterns and the discovery of useful knowledge of information systems. Education such as reception systems, enrolment systems, course management focuses on. Educational systems (such as Moodle, etc.) and any other system that deals with students at different levels of education, from schools, colleges and universities. Today, they are attracted by various researchers and professors. Researchers in this field focus on discovering useful knowledge to help educational institutions better manage students or to help students better manage the education and products offered and enhance their performance.
- Several different classification methods and techniques are used in the discovery of knowledge and data mining. Each method or technique has its advantages and disadvantages. Therefore, this research uses classification methods to create models for students who have poor or strong educational performance. It was used and then identified the factors affecting student performance and provided guidance to educational authorities and professors to improve the educational performance of students.

3- Research Objectives

Main objectives

To provide a new method for identifying successful and unsuccessful students in end-of-semester exams and to identify the factors affecting their success or failure



Sub-goals

- Using the proposed method to help educational officials and professors to identify students who are likely to be unsuccessful in the final exams.
- Using the proposed method to provide strategies to prevent students from failing and reduce their academic decline in the final exams



4- Related works

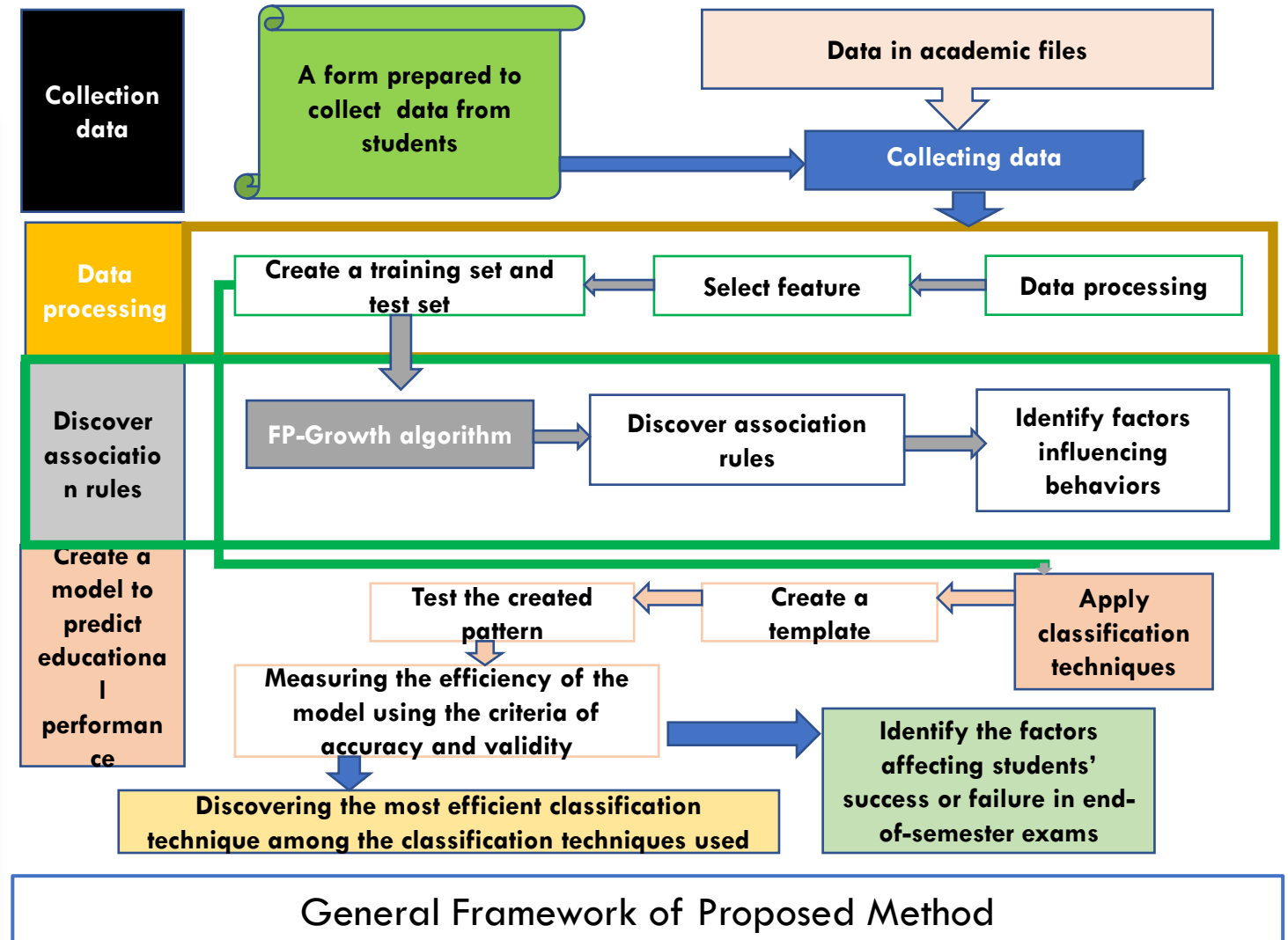
- Anupam Khan and his colleagues they carried out research that is specific and limited in nature and mainly focuses on studies that try to identify the predictor or model of student performance.
- Ramaswami and colleagues adopted a pilot survey method to generate a database that was built from a primary source and a secondary source. While primary data was collected from ordinary students, secondary data was collected from the school and the Office of the Chief Educational Officer (CEO).
- Moriana and her colleagues had a possible impact on extracurricular activities such as study-related (taught or private classes, computers) and/or related to sports (in and out-of-home games), they studied the academic performance of high school students in Spain.
- Ahuja and her colleagues compared the performance of different clustering and classification algorithms applied to the same training dataset.

4- Related works

- Asif and his colleagues used data mining methods to study the performance of undergraduate students and focused on two aspects of student performance, which are: First, the ,predicting student academic achievement at the end of a four-year study program .Second, the study of typical improvements and their combination with predictive outcomes.
- Slater and his colleagues reviewed some of the most widely used, accessible and powerful tools available to researchers interested in conducting EDM/LA research.
- Tismy and colleagues in a study ,A new system suggested that the proposed system is a web-based application that uses a simple Bayesian mining technique to extract useful information.

5- Methodology

To predict the results of students' end-of-semester exams, we will use different data mining methods in this thesis. Since each of the ranking techniques has abilities and weaknesses, we will use a combination of classification techniques in this thesis. Finally, we can choose the technique that works best in data analysis and forecasting the results of students' end-of-semester exams. The general framework of the proposed method in this thesis is presented in Figure



5- Methodology

The data and information collection data stage

In this thesis, by studying multiple studies conducted by different researchers, we create a form for collecting data from students. Form created online or in person at the disposal of the target group (students are placed so that we can collect their views and opinions, along with the numerous characteristics and factors affecting their life and educational status.

The form created in this thesis is shown in Figure .

Dear student, please answer the following questions carefully.

- 1- Your gender? a) woman (b) man
- 2- Your age?
(a) under 18 years (b) between 18 and 22 years (c) between 22 and 27 years (d) more than 28 years
- 3- Your degree? (a) first year (b) second year (c) third year (d) fourth year
- 4- Education level of parents? (A) Sub-diploma (b) Post-graduate diploma (c) Bachelor's degree (D) Post-graduate - doctorate and post -doctorate
- 5- Average GPA of your previous semesters? (a) 18-20 (b) 16-18 (c) 14-16 (d) 14-16 and 10-14
- 6- Proportion of online classroom teaching hours with the volume of courses? (A) Excellent (B) High (C) Moderate in low and up to moderate
- 7- How enjoyable is online training? (A) Too much (B) Much (C) Medium D Little and not enjoyable
- 8- How enjoyable are face -to-face trainings? (A) Too much (B) Much (C) Medium or low and not enjoyable
- 9- The average time spent on online training during the day and night? (a) 3 to 5 hours (b) 2 to 3 hours (c) 1 to 2 hours (d) under 1 hour
- 10- The average time spent for face -to-face training during the day and night? (a) 3 to 5 hours (b) 2 to 3 hours (c) 1 to 2 hours (d) under 1 hour
- 11- The average time spent on social networks during the day and night? a) 3 to 5 hours (b) 2 to 3 hours (c) 1 to 2 hours (d) under one hour
- 12- The average time spent on the Internet for scientific discussions day and night? a) 3 to 5 hours (b) 2 to 3 hours (c) 1 to 2 hours (d) under one hour
- 13- Do professors use slides and books related to the classroom? (A) Very much (B) Much (C) Moderate (D) Very little and rarely or not at all
- 14- Participating in discussion groups on the Internet? a) very much (b) a lot (c) moderate or little and rarely or none
- 15- The amount of student participation in classroom teaching? (A) Very good (B) Good (C) Average (D) Few and very few
- 16- How many virtual friends do you have compared to your real friends? (a) very much (b) much (c) medium (d) little and very little or infrequently
- 17- Audio content used for teaching? (a) very much (b) much (c) medium (d) little and very little or infrequently
- 18- video content used for teaching? a) very much (b) a lot (c) medium (d) little and very little or infrequently
- 19- The amount of homework given for home by the teacher? (a) very much (b) much (c) medium (d) little and very little or infrequently
- 20- How satisfied are you with social interactions in virtual education? (A) Very much (B) Much (C) Average Low and very little or rarely
- 21- The satisfaction level of using video conference for teaching or interactions in online classrooms? a) very much b) a lot (c) average (d) Little, very little or infrequently
- 22- Using the Internet to exchange course information with classmates? (A) Very much (B) Much (C) Medium (D) Little and very little or very often
- 23- How useful is the use of online classroom facilities? (A) Very much (B) Much (C) Average (D) Little and very little or rarely
- 24- How useful is the use of online classroom facilities? a) very high (b) high (c) average (d) little and very little or rarely
- 25- The effect of internet speed and bandwidth on e -learning system? (a) very much (b) much (c) medium (d) little and very little or infrequently

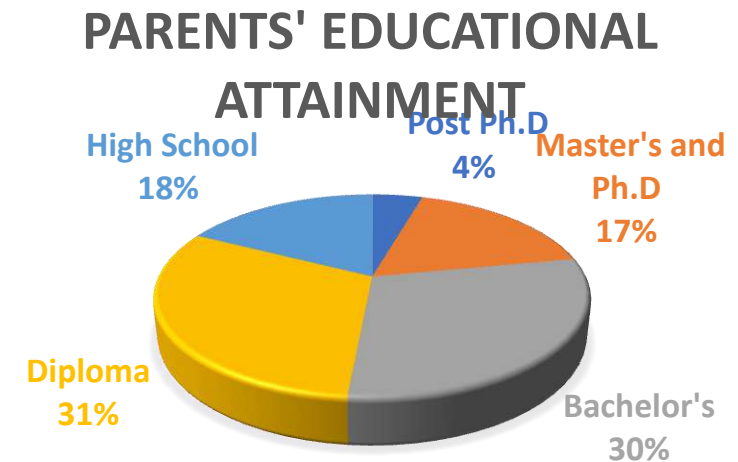
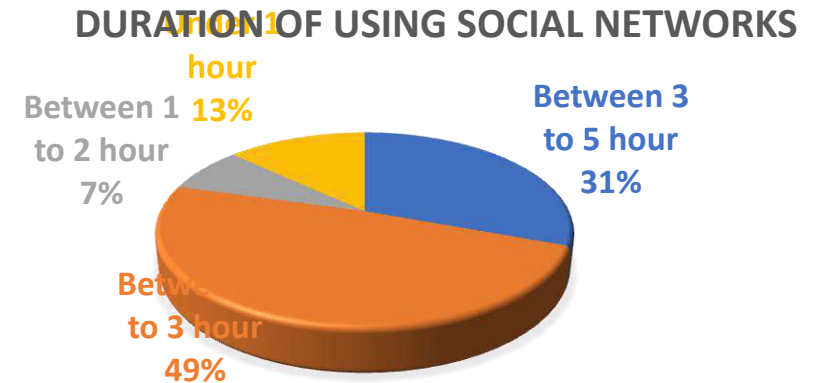
Form Created In This thesis

6- Implementation and Evaluation

To implement the proposed method in this thesis, we use Rapid Miner software. For this purpose, we will perform the following steps.

- Use data collection methods
- Preprocessing of data
- Conducting initial analyses
- Operations performed to discover associative rules
- Operation performed to create models to identify successful and unsuccessful students
- Evaluation of the proposed method

* Conducting initial analyses



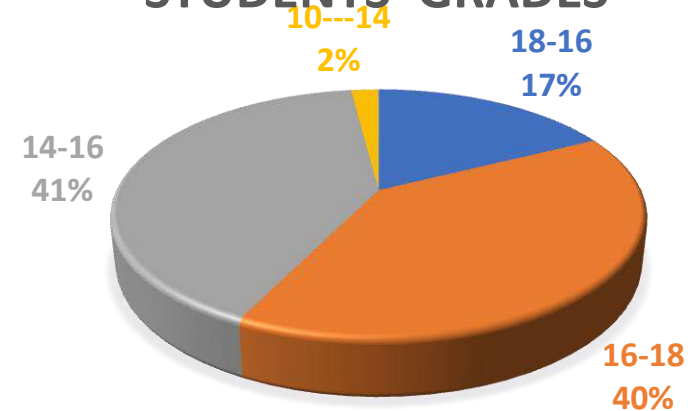
6- Implementation and Evaluation

To implement the proposed method in this thesis, we use Rapid Miner software. For this purpose, we will perform the following steps.

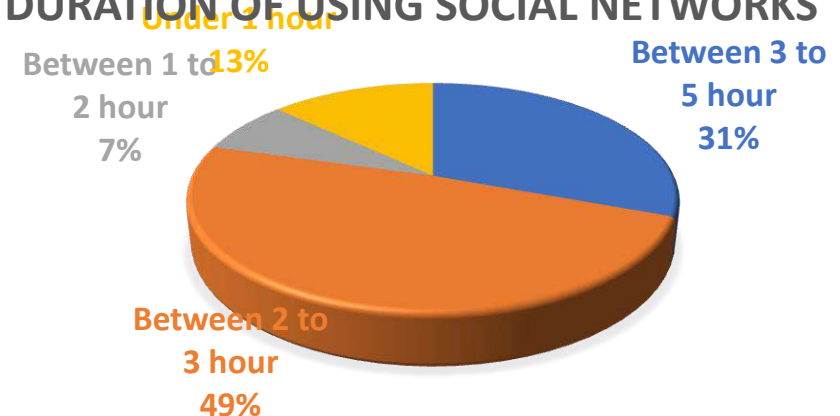
- Use data collection methods
- Preprocessing of data
- Conducting initial analyses
- Operations performed to discover associative rules
- Operation performed to create models to identify successful and unsuccessful students
- Evaluation of the proposed method

* Conducting initial analyses

STUDENTS' GRADES



DURATION OF USING SOCIAL NETWORKS



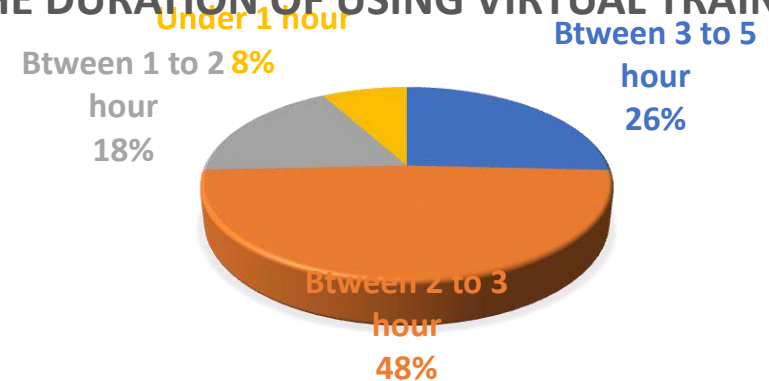
6- Implementation and Evaluation

To implement the proposed method in this thesis, we use Rapid Miner software. For this purpose, we will perform the following steps.

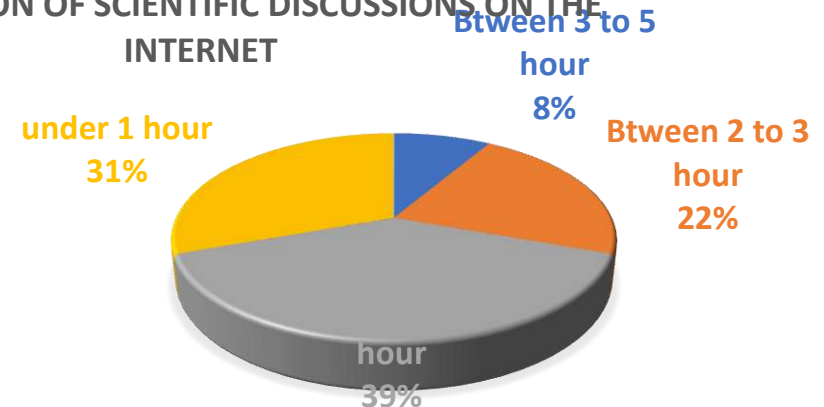
- Use data collection methods
- Preprocessing of data
- Conducting initial analyses
- Operations performed to discover associative rules
- Operation performed to create models to identify successful and unsuccessful students
- Evaluation of the proposed method

* Conducting initial analyses

THE DURATION OF USING VIRTUAL TRAINING



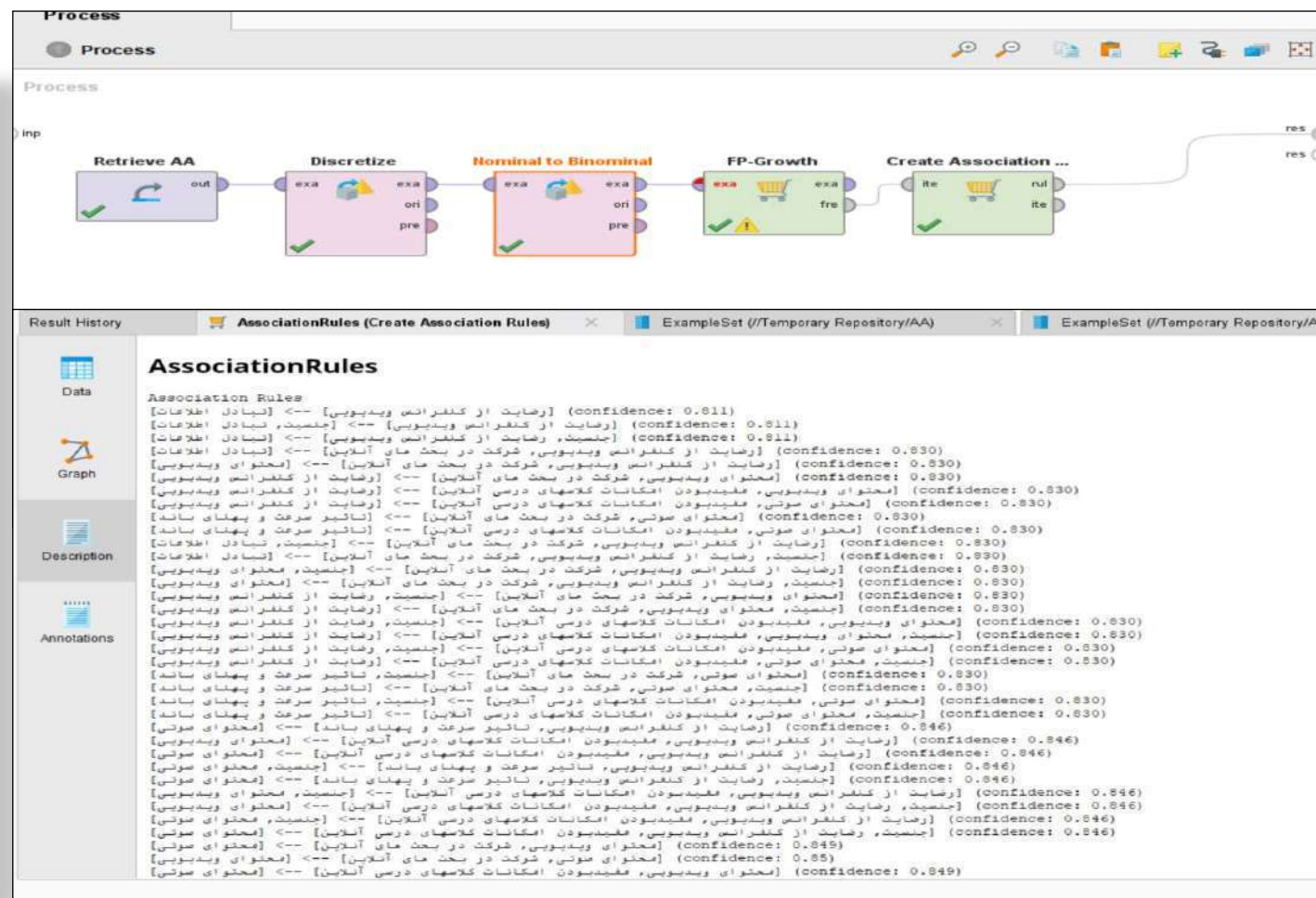
THE DURATION OF SCIENTIFIC DISCUSSIONS ON THE INTERNET



6- Implementation and Evaluation

Operations performed to discover associative rules

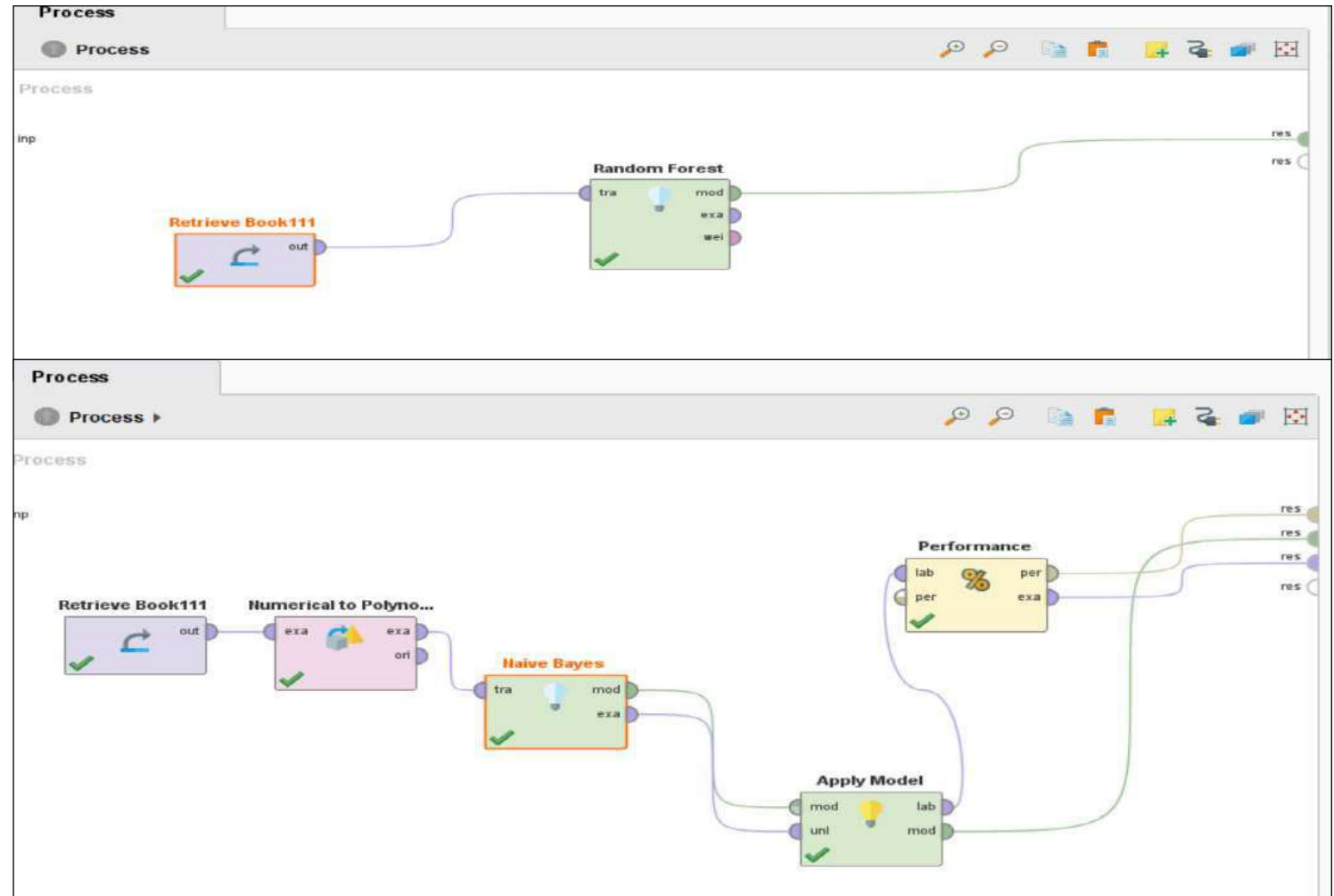
The training dataset was loaded into Rapid miner software and we create associative rules using the operations and operators shown in Figure



6- Implementation and Evaluation

Operation performed to create models to identify successful and unsuccessful students

- Using Random Forest Classification Technique
- Using the Naive Bayes classification technique



6- Implementation and Evaluation

- Evaluation of the proposed method

Accuracy of Random Forest Method

Table View Plot View

accuracy: 93.50%

	true b	true c	true a	true d	class precision
pred. b	79	0	9	0	89.77%
pred. c	0	82	0	4	95.35%
pred. a	0	0	26	0	100.00%
pred. d	0	0	0	0	0.00%
class recall	100.00%	100.00%	74.29%	0.00%	

Accuracy of the Decision tree method

accuracy: 63.41%

	true b	true c	true a	true d	class precision
pred. b	9	2	1	0	75.00%
pred. c	6	14	3	0	60.87%
pred. a	1	1	3	1	50.00%
pred. d	0	0	0	0	0.00%
class recall	56.25%	82.35%	42.86%	0.00%	

Accuracy of the Naive Bayes Method

accuracy: 100.00%

	true ب	true ج	true ايف	true الف	true د	class precision
pred. ب	79	0	0	0	0	100.00%
pred. ج	0	51	0	0	0	100.00%
pred. ايف	0	0	18	0	0	100.00%
pred. الف	0	0	0	35	0	100.00%
pred. د	0	0	0	0	17	100.00%
class recall	100.00%	100.00%	100.00%	100.00%	100.00%	

6- Implementation and Evaluation

Comparison of the efficiency of different methods in predicting student GPA at the end of the semester

Technique/critical	Accuracy	Recall	F-Measure
Decision tree	63.4%	63.3%	63.1 %
Random forest	93.5%	93.1 %	93.2%
Naïve Bayes	100%	99.9%	99.9%

7- Conclusion and Future Works

Conclusion

Education from the very birth of mankind was an important part of his life and maturity, and men tried in different ways to bring their efforts and knowledge to the people. Leave again. One of the most important issues in education is that the power of understanding and understanding of all educators is not the same. On the other hand, it is seen that some students or students fail for different reasons in the exams at the end of the semester and therefore both their life and life loss. The capital of the government and the family is wasted, so in this research we have tried to investigate the factors affecting the success or failure of students and then using Different classification techniques try to suggest a model for predicting students' academic status. The results of the evaluation of the proposed method indicate that Naive Bayes method has better performance in predicting the academic status of students in the final exams.

7- Conclusion and Future Works

Future Works

In the future, we will try to increase both the number of data collection samples and use other classification techniques to identify students at risk of failure in the final exams.



THANKS