Journal of Theoretical and Applied Information Technology
31st March 2021. Vol.99. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

RATIONALIZING THE CHOICE OF ACADEMIC MAJOR FOR UNDERGRADUATES USING CLASSIFICATION ALGORITHMS

ELMASRY, MOHAMED ABBAS
National Egyptian E-Leaning University (EELU), Giza, Egypt.
E-mail: malmasry@eelu.edu.eg

ABSTRACT

Lately, there has been a massive increase in the amount of data stored in educational databases. These databases include important knowledge of student score records.
Educational data mining is used to study the data available in the educational field and extract unseen implications from it.
From the moment a student enrolls in higher education, he needs guidance and assistance in decision making. Choosing a major is one of the important decisions for students to be qualified for. The decision to choose a major should be based on a rationale suitable for the student's cognitive skills. It should be noted that choosing a major is also important because it is often related to the type of job the student will occupy after graduation.
All of these considerations require detailed research so that we can scientifically guide students, especially the most vulnerable students, to encourage them to realize their fullest potential in the major that best suits them.
So, to what extent can the current academic advising system help students make this crucial decision?
Academic advisors need a clear form that sheds light on how to guide students in choosing the right major for them. Looking at the rate of change in majors for students of the faculty of Business Administration of the National Egyptian E-Learning University (EELU), it was found to be a very high rate during the six-year period covered by the study, reaching 48% of the students, which reflects the lack of clarity of vision for these students while they make a decision about the major.
Classifying student data would contribute to making decisions about making rational decisions regarding major selection, and achieve the best results.
In this paper, we used classification algorithms for a dataset of business administration students at the National Egyptian E-Learning University in a way that enables academics to predict the student's GPA in different majors, so that students can use them to make informed decisions about the appropriate major choice.
The model used in this study predicted the appropriate major with an average recall value of 89.83%.
Faculty members can benefit from these results in taking the necessary actions to improve the student's academic performance. This study may reduce dropout rates and improve student performance in the recommended majors.

Keywords: *Educational Data Mining, EDM, Classification Algorithms, Prediction, Major Selection, Undergraduates, Higher Education, Academic Performance.*

1. INTRODUCTION

The carefully studied selection of academic major is a critical factor in the success and superiority of students and in building their future.

In a Gallup Poll - Education Consumer Pulse [16] it turns out that more than 36% of alumni in the United States indicated that they would have changed majors if they had the opportunity to do so again.

An important question revolves around how students obtain advice on choosing a major, and what do students find most valuable in this regard?

A survey was conducted via Moodle for Business Administration students at National Egyptian E-learning University (EELU) who chose the major.

The aim of the survey was to determine the reasons for students choosing the major.

312 students out of 340 answered the survey, and the response rate was 91%, and the answers were as follows:

Table 1: Sources of information for decision-making in choosing a major

Sources	Number of students	Percentage %
Informal social sources: such as friends and relatives.	216	69.24%
Official Sources: Academic Advising by Doctors and Teaching Assistants	34	10.89%
Expectations of available employment areas	55	17.63%
Other sources	7	2.24%
Total	312	100%

It is evident from Table (1) that 69.24% of students made a decision to choose a major based on informal social sources, and thus it is considered the main source of decision-making, while only 10.89% of students chose a major decision based on the main source, which is the academic advising. About 17.63% chose the major based on their expectations about the future career, these expectations may be true as well as wrong. Overall, the results indicate an urgent need to help students choose a major in an informed manner.

By reviewing the cases of changing majors, it was found that 162 students out of 340 students changed the major after choosing it, i.e. 48%, which is a high percentage close to half.

Another survey was conducted on Moodle for students who changed their major to determine the reasons for the change. 151 students out of 162 answered the survey, the response rate was 93.2% , with the results as in the following table:

Table 2: Reasons for changing students' major after being selected

Reasons for changing Major	Number of students	Percentage %
Academic deterioration in major courses	122	80.79%
Personal conviction of the importance of another major	15	9.93%
Follow friends and colleagues	11	7.29%
Other reasons	3	1.99%
Total	151	1

The high percentage of changing major indicates that the decision to choose a major was ill-considered from the beginning. The problem deepens with the delay in the decision to change the major, as students start studying according to a major, and then discover over time that they are unable to continue in the major and then decide to change the major. This delay in making the decision to change the major after a period of study leads to delaying the graduation of students; in addition to increasing financial burdens as a result of studying courses belong to another major. These financial burdens constitute a crisis for students with limited income.

It is evident from Table (2) that 80.79%, which is the vast majority, made a decision to change the major due academic deterioration in courses of previously selected major.

As it is evident from the above, the change of major is due to a poor choice from the beginning, in addition to choosing a major that does not suit the students' tendencies, which was evident from their academic deterioration in the major courses that they changed later.

In the same context, educational data mining has emerged as a way to enhance academic advising, improve students' academic choices, and increase their academic performance. Therefore, this paper seeks to prepare academic advising to aid in appropriate selections for the major by mining students' data and extracting useful results. We can identify patterns and trends in students' grades by using decision tree in order to predict their future performance and make academic decisions that help

improve their performance in various majors.

This paper aims to guide students' decisions in choosing their major through designing and building a supervised machine learning model whose results provide a clearer vision for activating the role of academic advising in helping students choose their major.

To achieve this purpose, the paper will focus on classifying of a large volume of educational data which can be processed and analyzed in various ways, The data used in the analysis relates to six years of the Faculty of Business Administration at the Egyptian E-learning University, while the grades are all grades of exams and the coursework of the semesters including quizzes, assignments, midterms and attendance, as well as in the form of pre-processing of data.

2. RELATED WORKS

Students' choice of major has a big impact on their way of life, but many questions remain unanswered about how students choose their majors and why they choose those majors specifically.

Previous research indicates that personal interests, career expectations, parental influence, and the academic atmosphere are important factors in determining students' choice of major.

However, there is limited research on how advising influences students' decision to pursue studies in a particular major.

In a study by Gallup and Strada [16] to find out how students choose a major, identify which sources provide guidance to students, and to what extent students benefit from this advice, a survey was conducted of 350 graduates in the United States about their academic experience and choice of their majors.

The results show that students received advice from a variety of sources. The results also show that students find some resources more useful than others. This study was expanded to include about 2,2087 responses from US residents who asked where they obtained guidance on academic specialization. They were also asked to evaluate the advice they received in this regard.

The study indicates that although informal social networks were the most used source of support for choosing a major, reliance on other sources of advice has changed over time. However, the extent of reliance on these sources varies according to the time in which students obtained their higher education; the following figure shows the source of advice, according to the completion decade.

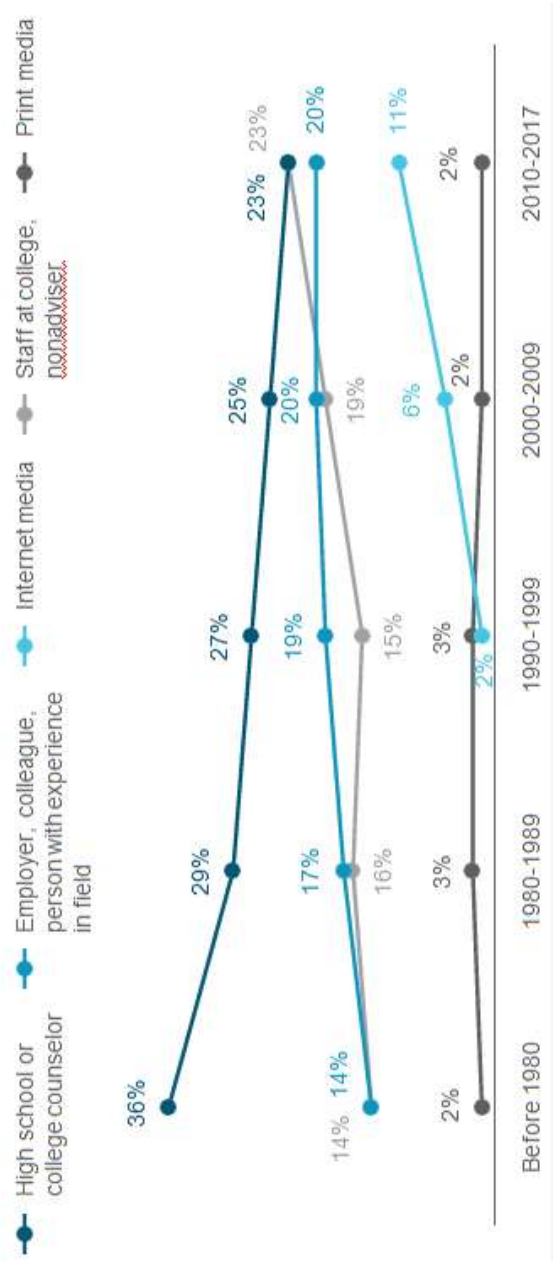

Figure 1: Source of advice, by decade of completion [16]

According to the study, the first source was the official source; it includes both academic advisors and academic media. The official source is primarily concerned with providing guidance to students on their academic options.

The second source was the informal social network, which mainly consists of friends who influence the student, family, and personalities who have an impact in society. These resources are the culmination of an informal network of information and advice provided to students, but they are not resources originally intended to provide students guidance on their academic choices. The study found that about 55% of students receive guidance on choosing their majors from informal social networking sources. The study also indicated that students whose parents obtained a bachelor's degree (60%) or a post graduate degree (65%) receive support in the decision to choose their major through informal social networks.

The third source was the non-formal school, which includes high school teachers, faculty members and those of similar rank, and they have a basic educational role, but their role in selecting students for their majors is an advisory role. The study found that about 18% of students receive guidance on choosing their majors from the high school teachers and faculty members.

The fourth source was informal work; in relation to students working alongside their studies, where experience gained on the job and advising of colleagues have a role in students' decision to choose a major. T he study found that older students are more likely to receive support in choosing their major from work sources.

The study indicated that 47% of female students receive advice from official sources of advice in choosing a major, compared to 41% for male students. As for the probability of female students receiving advice in choosing a major from their social network, it is 53% compared to 56% for male students.

Table 3: Evaluation of each Source category in providing advice for major selection [16]

Source of advice	% Say source helpful/ very helpful	Source category
Family	75%	Informal social network
College counselor	66%	Formal
Friend	71%	Informal social network
Staff at college, nonadviser	79%	Informal school-based
High school teacher	78%	Informal school-based
High school counselor	61%	Formal
Person with experience in field	84%	Informal work-based
Employer or coworker	82%	Informal work-based
Internet media	68%	Formal
Print media	69%	Formal
Other	77%	-
Don't know	79%	-
Did not receive advice	62%	-
Military	81%	Informal work-based
Community leader	84%	Informal social network
High school coach	79%	Informal school-based

Among the most important findings of the study is that the most common advice about students 'choice of their majors is not always useful, the sources that

can provide valuable guidance to students on choosing their majors are the least used sources. As a result, this study raises the alarm about how to rely on those sources in making decisions, and stressing the need for more research that supports the role of academic advising to help students make rational decisions when choosing a major.

Looking at another angle of literature review, which is data mining, it can be said that data mining in higher education is a modern research field and this research field has gained popularity due to the great value that its results can achieve in the field of university education.

Data mining can be used in the educational field to deepen our understanding of the learning process in order to focus on identifying, extracting and evaluating variables related to student learning.

Han and Kamber [17] used data mining software that allowed users to analyze data from various dimensions, categorize it and extract meaningful relationships that were identified during the mining process.

Calkins and Welki study [10] indicated that factors affecting students 'choice of majors include prior interest in the subject, the reputation of the university, and the influence of parents or other important figures in students' lives.

The study found that 55% of respondents believe that the teaching reputation of faculty members in the department was a very important factor in their choice of economics major.

The study of Baker et al. [3] was concerned with the role of colleges in preparing students for the job market by influencing the choice of majors. The study found that less than 15% of a sample of community college students in California, USA, pursues a range of majors precisely according to the needs of the labor market. As students tend to pursue majors that lead to jobs with higher salaries, the study found that students believe that salaries are 13% higher than they are in reality. The study also indicated that estimates of the impact of expected labor market results show that a 10% salary increase is associated with an increase, With a 14 to 18% probability of choosing a specific major.

As for the study Bordon and Fu [8], they found that forcing students to choose a major, as is currently happening in Chile, gives a higher probability of the student's inability to conform to the major, the authors also concluded that moving to another model that allows choosing a major like the one in the United States will increase students 'well-being by 1%, this improvement, albeit slight, is effective for people with low incomes.

As for the study Bridet and Leighton [9], it examined the relationship of early selection for a major, which allows for more credit hours to focus on specific key skills that are suitable for professions associated with specific specializations, while allowing for a longer exploration period that allows students a longer time to get to know their capabilities that allow them to have a competitive advantage. . The study found that the shift from an open selection model, such as what is available in USA, to an early major decision required, as is available in some European countries, leads to a decrease in the expected profits of the alumni by 1.5 %.

Magali et al. [24] examined the impact of expected salaries for jobs related to different majors, where they found that the expected profits had only a weak effect on students' decision-making in choosing their majors. The study also found that students 'choices had other considerations such as the student's ability and preferences for a particular type of study.

The study of Wiswall and Zafar [36] concludes that the earnings expectations, as well as students 'convictions about their abilities, perceptions of their abilities, have a major role in students' choice of majors. Despite this, the study found that students with affordable financial status have different reasons for choosing the major, such as the non-financial aspects of professions related to certain specialties and the pleasure of study courses.

Hijazi and Naqvi [19] studied the student performance by choosing a sample of 300 students (males and females) in a Pakistani university. The hypothesis that was stated as "the student's behavior towards attendance in the classroom, the hours they spend daily on study, family income and mother's education is largely framed with the student's performance." The study also found that factors such as the student's family income and the mother's level of education are closely related to the student's academic performance.

Galit [15] prepared a case study using student data in analyzing their learning behavior in order to

predict results and warn students at risk of failing well before their final exams.

Shaeela et al. [31] used k-means clustering algorithm to predict student learning activities. The study found that the information generated after implementing the data mining is useful for both the instructors and the students.

Bharadwaj and Pal [6] studies the student performance using a sample of 300 students of 5 different degrees studying Bachelor of Computer Applications, and found that factors such as the student's grade in the high school exam, the location of living, method of teaching, and the family's annual income are closely correlated to the academic performance of the student.

Bindiya V., Avittathur Un, K. Poulose [7] Clustered the students data of six consecutive years to characterize performance patterns. The study found that there is a direct relative correlation between attendance and the overall performance of students and that performance decreases with lower attendance. With the exception of the second semester, all university grades are equal or just below the university marks for the first semester. The first semester mark can be considered as an indication of what can be expected from a student in the coming semesters. The study also found that the internal assessment scores are a clear indication that can be used to predict future student academic performance.

Hedayetul-Islam Md., Mahfuza H [18] used of data mining process in the student's database using k-means clustering algorithm and decision tree technique to predict student learning activities. The study indicated that the use of data mining and data clustering technique improves student performance; reduces the failure rate and improves the quality of education.

Al-barrak & Al-razgan, [1] suggested using a J48 algorithm to predict final GPA of the students. This study used data of 236 students who graduated in the academic year 2012-2013 from Faculty of Computer Sciences at King Saud University. The paper attempted to determine which courses of previous semesters have a direct impact on final GPA. In result, it was found out that Java1, Database Principles, Software Engineering I, Information Security, Computer Ethics, and Project 2 are the most important courses that affect the final GPA of the students.

From the above, it is evident from the literature review that there are many studies that have applied educational data mining for purposes such as predicting students' GPA, improving students achievement, and reducing learning dropout, but there is a research gap regarding the use of educational technology tools such as data mining to assist students in making a decision to choose their major. To the best of our knowledge, there is no study directed to the purpose of predicting the appropriate academic major for the undergraduate student using classification algorithm prior to our current study.

3. Educational Data Mining (EDM)

Educational Data Mining (EDM) primarily develops methods for exploring the unique types of data that are produced within educational institutions [4]. It is based on the application of data mining techniques possessed by the educational entities to answer many important educational questions [30].

The following figure shows the main areas related to Educational Data Mining.

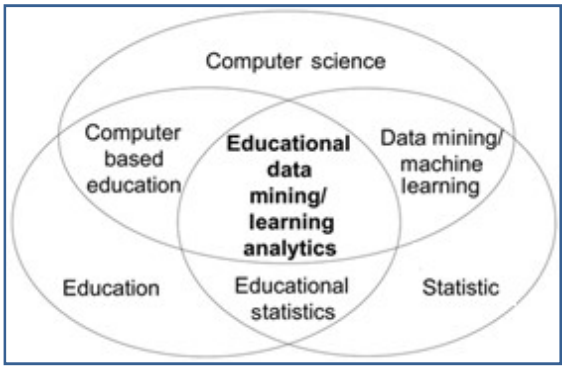

Figure 2: Key domains related to Educational Data Mining [4]

In this context, the academic analytics is concerned with identifying, collecting and analyzing academic program activities, such as courses, undergraduate programs, postgraduate programs, scientific research, etc.

The Clustering of basic student's data and academic data should generate useful knowledge for academic staff. Basic data may include: age, gender,

nationality, etc. Academic data may include: quizzes grades, assignments grades, midterms grades, and final scores.

There is a direct relationship between the amount of data available and the knowledge obtained. The more data that is available, the more knowledge we obtain.

The more time and effort was given in the pre-processing stage. The more results of the analysis are of greater benefit to the entire educational system, including students, instructors, teaching assistants, and administration [7].

Pre-processing of data and using of data mining techniques are aimed at obtaining high-quality learning and enhancing the quality of educational decisions.

The major way to obtain the highest quality in the higher educational system is by searching for knowledge from academic data to determine the key factors affecting student performance.

The resulting knowledge can be used by the academic decision-makers in the area of enhancing academic decisions, major selection, improving student academic performance, reducing failure rate [6].

Data mining, also called discovery of knowledge in large data sets refers to mining or extracting knowledge from a large amount of data.

The steps for data mining can be shown in the following figure:

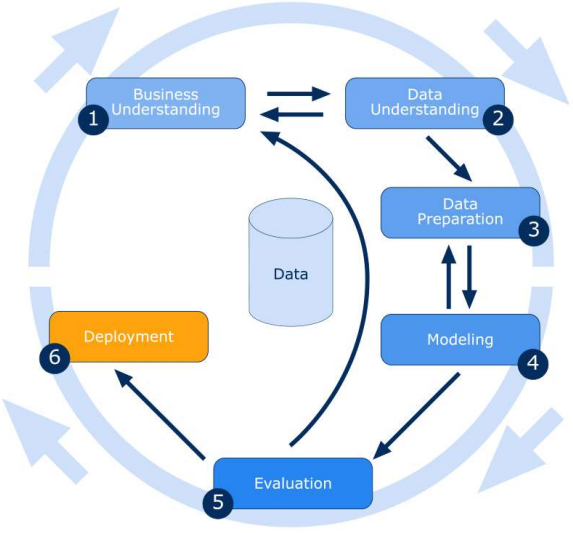

Figure 3: Steps for data mining,

The mining of academic data depends on many techniques like neural networks, decision tree, clustering, k-nearest neighbor and many other techniques.

Algorithms and models of data mining are used to work on a large amount of data to discover hidden or invisible patterns and relationships that aid decision-making processes.

Data mining is a five-step process:

- Identify source information.
- Choose the data points that need to be analyzed.
- Extract relevant information from the data.
- Determine the key values from the extracted data set.
- Interpreting and reporting results.

There are two types of data analysis that can be used to extract models that describe important classes or to predict future data trends. These two types are Classification and Prediction.

Data Classification is a form of analysis that creates a model which describes important class variables.

Decision tree is one of the popular predictive modeling approaches used in statistics. It is a data mining technique used to create classification models. It can be used to explain various variables such as the percentage of student grades in assignments, quizzes, midterms and so on. By using a decision tree, we can visualize decisions that make it easy to understand.

Classification is a very simple method and it is mostly used for data extraction. It is the process of data mining that predicts the item membership of the data tuples [17]. In similar cases, classification methods can be used to predict the final grade of the graduate student and there are four grades (Fail, Pass, Good, Very Good and Excellent). Normal Experiment runs as follows: the data analysts runs all the data on classifier then generate rules to predict the next data to predict the proper data class.

The Rule-based classifier makes use of a set of IF-THEN rules for classification. The Rule-based classifier extracts a set of rules that display relationships among fields of the data set and the

class label.

Classification and prediction generate a set of IF-THEN rules for data classification. The rules are easier for field experts to understand the relationships between data.

C&R Tree classifier is a data classification technique used to classify data for the purpose of prediction which allow the use of these relations to classify new rows [17].

There are two stages of the classification procedure:

• Developing the model.

• Evaluating the model using the testing data.

•Statistical based algorithms: Statistical procedures are normally having an accurate fundamental probability model which provides chances of being in each class rather than just a simple classification.

•Correlation Analysis: It is a statistical method aimed at measuring the degree of correlation between two continuous variables that are measured in a numerical manner (such as the correlation of a person's age and his weight).

•Regression Analysis: This method is used to demonstrate how the independent variable is related to the dependent variable in a numerical way.

•Simple Approach: According to this method each class is represented by its center, thus the new element may become a member of the class with an opportunity to obtain the greatest possible value from similarity.

•K nearest neighbors: It is a non-parametric method whose idea is to use distance measurement in a way that allows storing all available cases, and every time a new case is inserted, it can be classified according to the distance function.

•Decision tree based algorithms: Decision Tree Mining is a type of data mining method for developing classification models. Classification models are designed as a tree-like structure, and this type of mining falls under the category of supervised class learning.

The classification method passes through two stages:

1-Build a tree known as a decision tree

2- Implementation of the decision tree at the database level

Data mining can be done by combining Decision tree induction with data storage techniques. A decision tree is used for clustering, classification, and prediction purposes as a predictive node technique. A question is posed to differentiate the root and each internal node in the decision tree. The arcs emanating for each node represent the possible answers to the question related to it. Each leaf node is a prediction for solving the problem under study.

The decision tree is used as a predictive model to move from observations about an item that ends at inferences about the item's target value. The decision tree is also used to develop data models that predict the class or value labels for appropriate decision making.

The basic algorithm used in decision tree induction is an algorithm that builds decision trees in a recursive, iterative, divide-and-conquer fashion. The decision tree algorithm is based on building a decision tree from specific training data.

• Neural Network based algorithms: Its idea is to build a model that provides a format for representing the data. At the time of tuple classification, the attributes of that tuple are forwarded in the form of a graph.

•Rule-based algorithms: rely on a classification procedure according to: (if then else) method for classifying data.

4. PRACTICAL DATA MINING

Data Collection and Preprocessing

We collected a dataset of the records and transcripts of 340 graduate students from the Faculty of Business Administration at EELU in the period from 2012 to 2018. We faced some issues while collecting and preparing this dataset as listed below:

•There were a group of students who started their programs in other universities and moved to EELU. We only have the courses they passed, but there is no

information about the attempts they made in these courses or other courses.

•Complete the missing data from the transcripts to make sure that all the students passed at least 52 courses which are required to complete the program.

•Complete the missing data from the transcripts to make sure that all the students completed the mandatory 5 courses needed for their graduation major.

Multi-Label Model for Major Recommendation:

The objective of this task is to design and build a supervised machine learning model to help the students select their major/department. In the Faculty of Business Administration at EELU, the students take mandatory general courses for a complete 5 levels/semesters. At the beginning of Semester 6, they should decide which major they will continue their studies in out of three majors; Marketing, Accounting and Finance. This decision is very difficult and most of the time depends on the student's study group preferences as the student himself has no clues or factors to build his decision on. So, we need to build a classification model to study the historical data of 340 graduate students from the EELU from 2012 to 2018 to recommend the proper major for the new students.

For more details, we collected the students' grades in the general courses (from Semester 1 to Semester 5) as our main dataset. In addition, we include three more values for each student, which are: Final Grade (graduation grade), Major (the student's graduation major) and Change Major (flag indicating if the student changes his/her major before graduation or not). In our model, we need to provide the new students with a complete recommendation about the suitable major and their expected graduation grade.

For this task, we have two target values to train our model on, which are the Major and the Final Grade. So, when it comes to the implementation, we decided to select an ensemble technique to build and train our model for both target values. We also used **RapidMiner®** tool for all the implementation steps as shown in Figure (4).

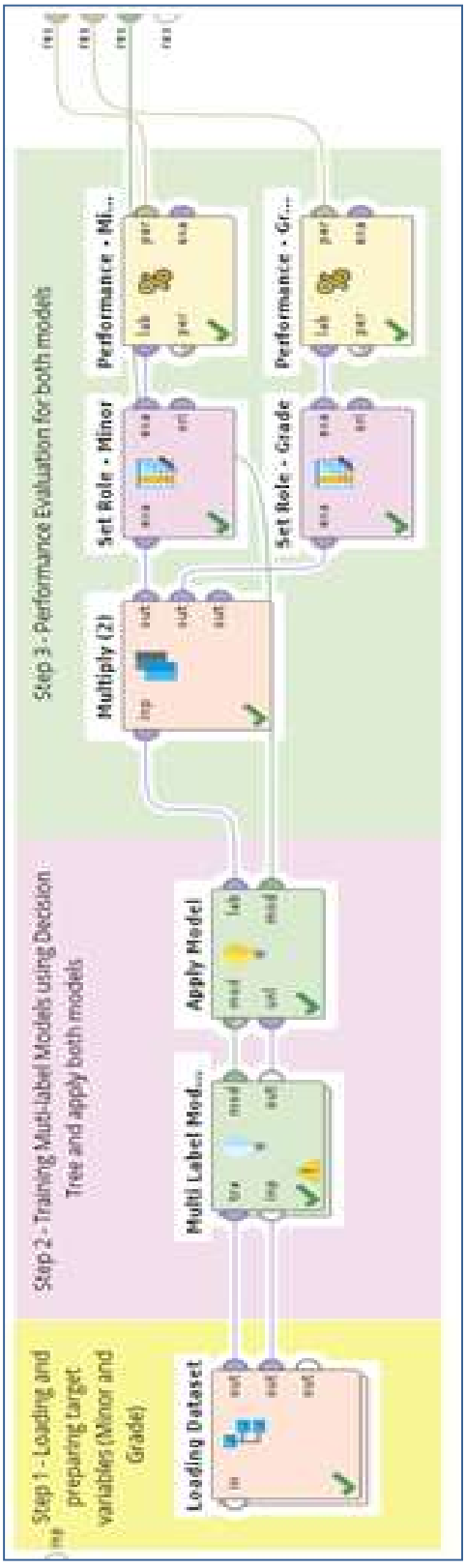

Figure 4 Multi-Label Model Implementation in Rapid Miner

As shown in Figure (4), the process starts with loading the dataset containing the students' grades in the general courses (up to Semester 5) and select the Major and Final Grade to be the target values for the classifier. Next, we used the Multi-Label ensemble operator which employed a Decision Tree classifier to map the students' grades into each of the target values. Finally, we apply the trained model on the dataset to predict the Major and Final Grade for each student. The performance for both target values is evaluated by comparing the true values with the predicted values for each student. The result confusion matrix for both Major and Final Grade targets are presented in Table 4 and Table 5, respectively.

Table 4: Confusion Matrix - Major Target

True / Predicted	Marketing	Accounting	Finance
Marketing	163	9	1
Accounting	5	143	3
Finance	1	0	15
Class Recall (%)	**96.45%**	**94.08%**	**78.95%**

As shown in Table (4), the classifier has a very good performance in predicting the student's major as overall with accuracy of 94.41%. For the individual majors, the model can detect the Marketing and Accounting majors with recall value of 96.45% and 94.08% respectively. For the Finance major, the model could not obtain such high performance and achieved a recall value of 78.95% only due the relatively small number of students graduated from this major compared to the other two majors. As an overall measure, the model can predict/recommend the correct major with average recall value of 89.83%.

Table 5: Confusion Matrix –Final Grade Target

True / Predicted	Pass	Good	Very Good	Excellent
Pass	92	0	0	0
Good	4	123	1	1
Very Good	0	2	76	1
Excellent	0	0	0	40
Class Recall (%)	95.83 %	98.40 %	98.70 %	95.24 %

As shown in Table (5), the classifier has an excellent performance in predicting the student's grade with overall accuracy of 97.04%. The model can predict the student's final grade efficiently especially the Good and Very Good grade as the model achieved very high recall values of 98.40% and 98.70%, respectively. The other two grades, Pass and Excellent, can be predicted also with recall values of 95.83% and 95.24%, respectively. In general, the model achieved an average recall value of 97.04% while considering all the Final Grade target values.

Based on our proposed model implementation and results, we can notice that the Decision Tree classifier was able to handle this classification problem very efficiently, especially for the Final Grade prediction. The performance of the Major prediction was also very good, but for the case of Finance major, the performance was little weak due to the unbalanced students' distribution as this major is relatively new and only 18 students graduated from it.

5. CONCLUSION AND FUTURE WORKS

Several studies have examined the factors of choosing a major from an economic point of view. These studies have linked choosing a major with an aspiration for higher-paying jobs. On the other hand, the surveys directed at students through this study showed that the majority of changing major cases were the result of deteriorating academic performance in the courses of the previous major. There is a gap in studies that search for choosing an academic major through mining educational data to

discover student's interests and strengths and direct them towards a major that suits their abilities.

Therefore, this paper investigated the use of educational data mining to support academic advising in assisting students in decision-making for choosing an appropriate major.

The case study involved students of business administration at The National Egyptian E-Learning University, as they have a variety of majors to choose from, including: Management, Marketing, Accounting, Management Information Systems, E-Commerce, and Finance and Investment.

Since the students specialize in majors starting from the sixth semester, there is ample opportunity to assess their academic performance and their progress through data mining to predict the most appropriate majors and advise them on the selection of majors through the academic advising process.

Significantly, students are very reluctant to select a major, discover over time that their choices were wrong, and request a change in majors, and this change major occurred by 48% of total graduated students (162 out of 340) during the six-year period covered by the study, from the academic year 2012 to 2018.

In this study, we make use of data mining process in EELU student's database using clustering algorithm and decision tree technique in order to predict the most appropriate major for students.

Tree classifier was able to handle this classification problem very efficiently, especially for the Final Grade prediction. The performance of the Major prediction was also very good,

The model used predicted the correct major with an average recall value of 89.83%.

It should be noted that all students require support to varying degrees in making the decision to choose a major, but low-income students are the most beneficial in choosing a major because changing it later will lead to more financial burdens that they may not be able to meet.

We expect to deploy this model in the Students management Information System (MIS) in the faculty as it not only gives recommendations, but also provides some confidence of these recommendations benefiting from the explainability feature of the Decision Tree models.

We hope that the information created after implementing the data mining and data clustering technique is useful for students through the academic advising service they receive.

This work may enhance student performance; reduce failure ratio in majors; and limit the late change of major and the consequent academic delay. This is done through making appropriate guidance decisions at the right time to improve the quality of education.

For future work, we hope that data mining will take place in the high school detailed transcripts during the stage of admission in the university to predict the study type that best suits their abilities and competencies.

REFERENCES:

[1] Al-barrak & Al-razgan, "Predicting Students Final GPA Using Decision Trees: A Case Study", *International Journal of Information and Education Technology*, 6(7), 2016, 528-533. Article (CrossRef Link)

[2] Arcidiacono, Peter, Esteban Aucejo, and Ken Spenner, "What Happens After Enrollment? An Analysis of the Time Path of Racial Di§erences in GPA and Major Choice", *IZA Journal of Labor Economics*, 1(5) 33(8), 2012, 547-564. Article (CrossRef Link)

[3] Baker, Rachel & Bettinger, Eric & Jacob, Brian & Marinescu, Ioana, "The Effect of Labor Market Information on Community College Students' Major Choice", *Economics of Education*, 2018. Article (Web Link)

[4] Bakhshinategh, Behdad & Zaïane, Osmar & Elatia, Samira & Ipperciel, Donald, "data mining applications and tasks: A survey of the last 10 years. Education and Information Technologies", 2018. Article (CrossRef Link)

[5] Baradwaj, B. and Pal, S, "Mining Educational Data to Analyze Students Performance", *International Journal of Advanced Computer Science and Applications*, vol. 2. 2011. Article (CrossRef Link)

[6] Bharadwaj B.K. and Pal S, "Data Mining: A prediction for performance improvement using classification", *International Journal of Computer Science and Information Security*

(IJCSIS), Vol. 9, No. 4, 2011, pp. 136-140. Article (Web Link)

[7] Bindiya V., Avittathur Un, K. Poulose, "Clustering Student Data to Characterize Performance Patterns", *International Journal of Advanced Computer Science and Application,* 2011. Article (Web Link)

[8] Bordon, Paola, and Chao Fu. "College-major choice to college-then-major choice", *The Review of Economic Studies*, Vol. 82, No. 4, 2015, pp. 1247-1288. Article (CrossRef Link)

[9] Bridet, Luc, and Margaret Alice Leighton, "The major decision: Labor market implications of the timing of specialization in college", *School of Economics & Finance Discussion*, Paper , no. 1510 , University of St Andrews , St Andrews , 2015, pp. 1-74 . Article (CrossRef Link)

[10] Calkins, L. N., & Welki, A, Factors that influence choice of major: Why some students never consider economics", *International Journal of Social Economics*, 33(8), 2006, 547-564. Article (CrossRef Link)

[11] Chandra, E. and Nandhini, K, "Knowledge Mining from Student Data", *European Journal of Scientific Research*, 2010. Article (CrossRef Link)

[12] David H.et el., "Principles of Data Mining", *The MIT Press*, 2001. Article (CrossRef Link)

[13] Davies, Peter, Jean Mangan, Amanda Hughes, and Kim Slack, "Labour market motivation and undergraduates' choice of degree subject", *British Educational Research Journal*, Vol. 39, No. 2, 2013, pp. 361-382. (Web Link)

[14] El-Halees, A, "Mining Students Data to Analyze Learning Behavior: A Case Study',. The 2008 international Arab Conference of Information Technology", *Conference Proceedings, (ACIT2008), University of Sfax, Tunisia*, 2008. Article (CrossRef Link)

[15] Galit.et.al, "Examining online learning processes based on log files analysis: a case study. Research, Reflection and Innovations in Integrating ICT in Education", 2007. Article (CrossRef Link)

[16] Gallup and Strada Education Network, "Major Influence: Where Students Get Valued Advice on What to Study in College", 2017. Article (Web Link)

[17] Han, J. and Kamber, M., "Data Mining: Concepts and Techniques", 2nd edition. *The Morgan Kaufmann Series*, 2006. Article (Web Link)

[18] Hedayetul-Islam Md., Mahfuza H, "An Approach of Improving Student's Academic Performance by using K-means clustering algorithm and Decision tree", *(IJACSA) International*, 2012. Article (CrossRef Link)

[19] Hijazi, S. T. and Naqvi, "Factors affecting student's performance: A Case of Private Colleges", *Bangladesh e-Journal of Sociology*, Vol. 3, No. 1, 2006. Article (CrossRef Link)

[20] Kamal, M. S, Chowdhury, L & Nimmy, S. F, "New Dropout Prediction for Intelligent System", *International Journal of Computer Applications*, (0975 – 8887), 42(16), 2012, PP. 26-31. Article (Web Link)

[21] Kantardzic, M, "DATA MINING Concepts, Models, Methods, and Algorithms", *John Wiley & Sons, Inc*, 2019. Article (CrossRef Link)

[22] Long, Mark C, Dan Goldhaber, and Nick Huntington-Klein, "Do completed college majors respond to changes in wages?", *Economics of Education Review*, Vol. 49, 2015, pp. 1-14. Article (CrossRef Link)

[23] Luppino, Marc, and Richard Sander, "College major peer effects and attrition from the sciences", *IZA Journal of Labor Economics*, Vol. 4, No. 1, 2015, pp. 1-23. Article (CrossRef Link)

[24] Magali B, Denis F., Arnaud M, "Choosing the Field of Study in Postsecondary Education: Do Expected Earnings Matter? ", *The Review of Economics and Statistics,* MIT Press, vol. 94(1), 2012, PP. 334-347. Article (Web Link)

[25] Nahid Hajizadeh, Marzieh Ahmadzadeh, "of factors that affect students' academic performance - Data Mining Approach", *International Journal of advanced studies in Computer Science and Engineering*, *IJASCSE*, Volume 3, Issue 8, 2014. Article (CrossRef Link)

[26] Necla G¨und¨uza and Ernest Fokou´e, "Pattern Discovery in Students' Evaluations of Professors: A Statistical Data Mining Approach", *The Journal of Applied Statistics,* Vol. 215, 2015. Article (CrossRef Link)

[27] Oyelade, Jelili & Oladipupo, Olufunke & Obagbuwa, Ibidun, "Application of k Means Clustering algorithm for prediction of Students Academic Performance", *International Journal of Computer Science and Information Security*. 7. IJCSIS2010, vol.7, No.1, 2010. Article (CrossRef Link)

[28] Pandey, M & Sharma, V. K, "A Decision Tree Algorithm Pertaining to the Student

Performance Analysis and Prediction", *International Journal of Computer Applications*, (0975 – 8887), 61(13), 2013, PP. 1-5. Article (Web Link)

[29] Romero, C. a, "Educational data Minin, Survey from 1995 to 2005", *Expert Systems with Applications*, 2007. Article (CrossRef Link)

[30] Romero, Cristóbal & Ventura, Sebastian, "Data Mining in Education", *Wiley Interdisciplinary Reviews*, Volume3, Issue1, 2013, pp. 12-27. Article (CrossRef Link)

[31] Shaeela Ayesha & Tasleem Mustafa & Ahsan Raza Sattar & M.Inayat Khan, "Data Mining Model for Higher Education System", *European Journal of Scientific,* 2010. Article (CrossRef Link)

[32] Shannaq, B. , Rafael, Y. and Alexandro, V, "Student Relationship in Higher Education Using Data Mining Techniques", *Global Journal of Computer Science,* 2010. Article (Web Link)

[33] Tufféry, S, "Data Mining and Statistics for Decision Making", *Wiley & Sons*, 2011. (Web Link)

[34] Vashishta, S, "Efficient Retrieval of Text for Biomedical Domain using Data Mining Algorithm", *IJACSA - International Journal of Advanced Computer Science and Applications*, 2(4), 2011, PP. 77-80. Article (CrossRef Link)

[35] Vuda Sreenivasaro & Capt.Genetu , "Improving academic performance of student of defence university based on data warehousing and data mining", *Global Journal of computer science and technology*, v.12, Issue2, Version.1, 2012. Article (Web Link)

[36] Wiswall, Matthew, and Basit Zafar, "Determinants of college major choice: Identification using an information experiment", *The Review of Economic Studies*, Vol. 82, No. 2, 2015, pp. 791-824. Article (CrossRef Link)

[37] Zafar, Basit, "College major choice and the gender gap", *Journal of Human Resources*, Vol. 48, No. 3, 2013, pp. 545-595. Article (CrossRef Link)

[38] ZhaoHui. Maclennan.J, "Data Mining with SQL Server 2005", *Wiley Publishing, Inc. Journal of Advanced Computer*, 2005. Article (CrossRef Link)