# Developing Classification Model to Investigate the Problem of Computing Students Studies Length

Osman A. Abdalla, Majed M. Aborokbah, and Abdelrahman Osman ELfaki

***Abstract*—Students are the key asset for any higher education institutions and their success in achieving the best academic performance within study planned period are resulting in producing graduates with high educational and professional quality skills. These graduates will become great leaders, manpower and play an important role in country's economic and social development. In most Higher Education Institutions, the students' studies length problem has not been investigated comprehensively despite the seriousness of this problem and its impact in the long and short terms. The aim of this research is to develop mining classification model based on decision tree to support academic administrators in decision making by defining features of linger students which could be used to develop an early warning system that has the ability to predict the students "who" might exceed the planned study length period. Data from faculty of computers and information technology at Uuniversity of Tabuk, KSA, has been collected using a survey method; students from male and female sections are participated in this survey. Then the data is preprocessed, after preprocessing of the data, C4.5 algorithm has been applied to discover the classifications.**

***Index Terms*—Data mining, classification, computer in education, C4.5, linger students commas.**

## I. Introduction

Students are the key asset for any higher education institution. Normally, the students' success is measured and evaluated by achieving the best academic performance within study planned period which is positively affects in graduates' educational and professional quality skills. These graduates will become great leaders, manpower and play an important role in country's economic and social development [1], [2].

In most higher education institutions, the length of study problem has not been investigated comprehensively despite the seriousness of this problem and its impact on the short andlong term. The students who beyond six or seven years in their academic program are called "linger" students [3], this definition exclude medicine domains.

Currently, the higher education in Kingdom of Saudi Arabia (KSA) is growing rapidly through expansion the number of universities and undergraduate students; this may increase the consciousness regarding the quality of education and attained learning outcomes within the planned studies period.

Table I illustrates the number of admitted students for bachelor degree in 2015-2016 according to ministry of education statistics, KSA [4]. Further, the summary statistics about admitted and graduated students from University of Tabuk in 2014-2015 are demonstrated in Table II [5].

From these mentioned tables we can extract that there is a large gap between number of new comer students and graduated student as well as enrolled and graduated students. Although, this big ratio of variant could be justified by an expansion in admission of new students still there is big gap between expected graduate students and actual graduate students.

In the current era, data mining has given a great deal of concern in the wide range application fields such as education [6], finance [7], manufacturing [8], healthcare [9], business [10], telecommunication [11], agriculture [12], and customer relationship management [13]. Furthermore, the most commonly data mining tasks include: classification, clustering, association rules, regression analysis, and decision tree [14], [15]. The field of data mining is growing rapidly and has become one of the most important computing techniques due to its ability to extract and discover meaningful knowledge from a large amount of data and the potentially useful relationships inside these data [16]-[23].

Data mining techniques could be used to build an intelligent classifier model to categorize students based on linger problem which help the education policy makers to take the appropriate decision when required. Further, this model will help in predicting the current students those who show similar behaviors that may lead to linger in their study, and then take the needed precaution to help them in advanced.

The majority of previous studies that are deal with students' studies length problem are used statistical theories to reach the results. Statistical theories are used to prove or disprove predefined assumptions. Hence, in the majority of previous studies, the data had been collected to serve these predefined assumptions. On the other hand, the rest of previous studies have used machine learning to deal with students' studies length problem. Machine learning is concerned with finding hidden relationships between variables without need for predefined assumptions. Machine learning is used for exploring hidden knowledge inside the data. It is hidden because there are no any previous assumptions.

In machine learning, there are two main approaches: classification and clustering. Classification is defined as grouping data based on predefined criteria and clustering is defined as grouping data without any predefined criteria.

Manuscript received September 9, 2018; revised February 18, 2019. This work was funded by the University of Tabuk, Scientific Research Deanship under Grant S-0187-1438.

Osman A. Abdalla and Osman A. Abdalla are with the Information Technology Department, Faculty of Computing and Information Technology, Univesity of Tabuk, Tabuk, CO 71491 KSA (e-mail: o_mohammed@ ut.edu.sa, a.elfaki @ ut.edu.s).

Majed M. Aborokbah is with the Computer Science Department, Faculty of Computing and Information Technology, Univesity of Tabuk, Tabuk, CO 71491 KSA (e-mail: m.aborokbah @ ut.edu.sa).

In this paper, we have developing classification model to investigate the problem of computing Students Studies Length. The aim of this classification model is to extract the features of both types of students: linger and non-linger students. These features will help to defined expected linger student, i.e., this model could be used as an early alarm of problem of students studies length.

TABLE I: STATISTICS OF ADMITTED STUDENTS FOR BACHELOR DEGREE IN 2015-2016, KSA

| Study Type | Available seats | | | Admitted seats | | | Vacant seats | | | Admission Percentage | Vacant Percentage |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Male | Female | Total | Male | Female | Total | Male | Female | Total | | |
| **Regular** | 102,6 | 121,1 | 223,7 | 90,5 | 112,1 | 202,6 | 112,0 | 10,0 | 21,8 | 90.3 | 9.7 |
| **Affiliation** | 22,9 | 16,4 | 39,2 | 13,6 | 6,9 | 20,5 | 15,0 | 10,7 | 25,6 | 44.5 | 55.5 |
| **Parallel** | 1,9 | 820 | 2,7 | 2,0 | 1,0 | 3,1 | 268,0 | 227 | 495,0 | 86.2 | 13.8 |
| **e-learning** | 9,4 | 6,1 | 15,5 | 0 | 0 | 0 | 9,4 | 6,1 | 15,9 | 0.00 | 100.0 |
| **Total** | 136,8 | 144,4 | 281,2 | 106,2 | 120,0 | 226,225 | 36,5 | 27,0 | 63,4 | 78.1 | 21.9 |

TABLE II: SUMMARY STATISTICS FOR ADMITTED AND GRADUATED STUDENT FROM UNIVERSITY OF TABUK IN 2014-2015

| Gender | New comer students | Enrolled | Graduated last year |
|---|---|---|---|
| Male | 4,2 | 11,4 | 1,4 |
| Female | 4,9 | 18,2 | 2,8 |
| Total | 9,1 | 29,6 | 4,2 |

This model has been developed on five steps: 1) investigated the related works to design the questionnaire. 2) Collected the filled questionnaires. 3) Develop classification model based on C4.5 algorithm, the collected questionnaires have been classified in two groups: linger students and non-linger students. 5) Finally, features of both classes have been extracted. Applying classification model to define the most affective factors lead to the length of study problem can contribute in improving quality education and academic performance.

This paper is organized as follows. Section II reviews of literature related to application of data mining in length of studies problem. The method and design steps used in the study are described in the Section III. The results are discussed in Section IV. Section V concludes the paper and gives recommendations for future direction.

## II. LITERATURE REVIEW

In literature, there have been very few studies conducted with the aim of discovering the factors lead in studies length problem and graduation dropout rate, N. Zavale *et al.* [24] utilize a logistic regression analysis to measure and explain the graduation rate at Eduardo Mondlane University (UEM), Mozambique. The data of undergraduate students from 2003 to 2006 are collected to measure the graduation rate, where 8 independent variables (6 variables as student-specific and 2 concerning socio-demographic profile) are selected as core factors affecting in graduation rate. The findings of paper shows that: Firstly (i) More than 25% percent of students graduate within the prescribed duration of programs (ii) 30 percent of students graduate within the duration +2 additional years. (iii) Less than 10 percent of students graduate within the duration +4 years (iv) 40 percent of students do not graduate within the allowed maximum time, i.e., program duration +4 years. Secondly, from the 8 selected variables the following variables are significantly affect the graduation rates of UEM students: gender, admission grade, number of failed courses, program duration, and regime.

The investigation about length of studies problem based on clustering and the extraction of association rules has been conducted in the Greek Higher Education by P. Belsis et al. [3]. The students' data are collected through questionnaires in the lab classes. The study focuses on students who "linger" a low priority for registration in the lab classes and with limiting the number of times of attending lab based courses. Interesting results and rules are obtained and discussed. E. Katsikas, and T. Panagiotidis [25] focus on the symptom of long duration of studies in Greece. This study employs administrative and survey data to assess the relationship between students' socioeconomic background and educational outcomes. Regression and quantile regression methods are applied to examine the relationship between students' status i.e. working and non-working. The results show that the working students have not been not attained lower grades than non-working peers; the negative effect of the studies length on grades is not related to both students status, and the impacts of both students status are similarly.

The study aims to examine the influence of family financial assets during the time of student's college enrollment has been conducted by M. Zhan and D. Lanesskog [26]. The data is gathered from the National Longitudinal Survey of Youth. The findings show that the family assets are positively impacted on the chances of college graduation among the students. S. Robertson *et al.* [27] explore issues and processes of measuring progression and graduation rates in an RN-to-BSN (BSN stands for Registered Nurse and RN stands for Bachelor of Science in Nursing) population and to identify factors that facilitate/hinder their successful progression to work toward establishing benchmarks for success. Using the data collected from 14 California schools of nursing with RN-to-BSN programs, RN-to-BSN students were identified as generally older, married, and going to school part-time while working and juggling family responsibilities.

The work in [28] deal with this question: Do expenditures other than instructional expenditures impact graduation and persistence rates in American higher education? to answer this question a simulation based on panel data from institutional level and a variety of econometric approaches are used to analyze the influence graduation rates of undergraduate students. The most important result shows that expenditures for student service effect on graduation rate and the effects are higher for students with lower entrance test scores and higher Pell Grant expenditures.

From the above literatures, two groups of previous studies have been discussed. First group was focused on measurement of graduation rate as way to determine the gap between the numbers enrolled and graduated students, and the other group was focused on investigation the factors affecting study length problem. Table III(A) and Table III(B) summarize some of the main related studies.

Our work is focused on investigating the factors affecting study length problem similar to [24] but with covering most of related factors.

TABLE III(A): SUMMARY OF RELATED WORKS

| # | Author(s) | Year | Objective(s) | Sample | Method(s) |
|---|---|---|---|---|---|
| 1 | N. Zavale *et al.* | 2017 | • Measuring the graduation rate from statistical data<br>• Test the statistical significance of the selected variables in affecting the graduation rate | Data from student records system | Statistical analysis<br>And Stepwise logistic regression analysis |
| 2 | P. Belsis *et al.* | 2014 | • Defining linger students | Survey Students about lab attendance | Clustering and the mining of Association rules |
| 3 | E. Katsikas, and T. Panagiotidis | 2011 | Defining the symptom of long duration of studies | Collecting data from student records<br>and Survey | Regression and quantile regression |

TABLE III(B): SUMMARY OF RELATED WORKS

| # | Author(s) | Year | Objective(s) | Sample | Method(s) |
|---|---|---|---|---|---|
| 4 | M. Zhan, D. and Lanesskog | 2014 | • Investigates the relationships of family assets and debt with college graduation.<br>• Estimating models that control for a wide range of student, parental, and institutional variables | Survey | Regression analyses |
| 5 | S. Robertson *et al*. | 2010 | • Exploring issues and processes of measuring progression and graduation rates in an RN-to-BSN population<br>• Identify factors that facilitate/hinder their successful progression | Collecting data from the nursing education literature and survey | Benchmarking |
| 6 | D. A Webber, and R. Ehrenberg | 2009 | Analyze whether non-instructional expenditure categories influence graduation and first-year persistence rates of undergraduate students | Collecting data from several systems | unconditional quantile Regression model |

## III. DATA OF FACTORS AFFECTING STUDY LENGTH PROBLEM

The questionnaire method was adopted to collect empirical data for this study. The questionnaire was targeted computing students in two academic departments in faculty of computers and information technology, university of Tabuk, KSA. Table IV(A) to Table IV(F) in appendixes section show the questionnaire variables which are issued based on four perspectives: Linger Information, basic information, personal information, academic information, and learning and teaching Information.

## IV. DEVELOPING THE CLASSIFICATION MODEL

In order to investigate the factors affecting length study problem a questionnaire was distributed. In total, more than 250 male and female students are targeted to participate in this study. 175 students completed the survey for a response rate of 70%, which is considered as a very good response rate. Then the classification processes were performed by means of the data mining software RapidMiner 5.3.013.

The C4.5 algorithm was used to generate decision tree to mining the gathered data, and after the processes of constructing the complete decision tree, eight rules have been generated. Four rules deduce linger students and four rules deduce non-linger students. Table V shows rules deduce linger students. Table VI shows rules deduce the non-linger students.

In the following, features of two classes have been presented. Each class has group of subclasses. The combined features of each subclass represent either linger student or non-linger student. First features of linger students, and then followed by features of non-loner students. Table V(A), V(B), V(A), and V(B) represents the results of our classification model.

TABLE V(A): RULES DEDUCE LINGER STUDENTS

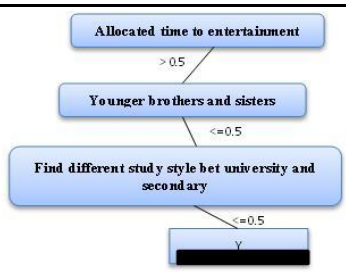
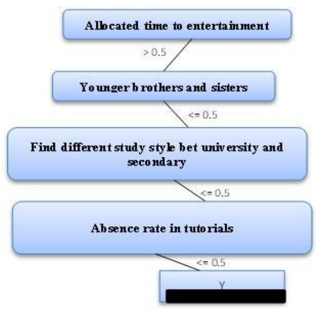
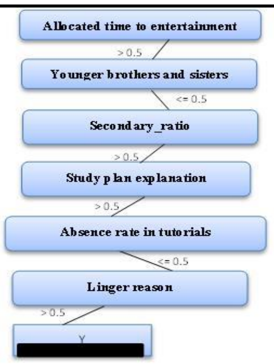
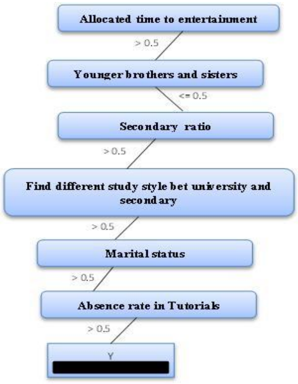
| # | Rule statement | Tree of rule |
|---|---|---|
| **1** | If (Allocated time to entertainment = No) and (Younger brothers and sisters = Yes) and (Find different study style bet university and secondary = Yes) then Linger |  |
| **2** | If (Allocated time to Entertainment = N0) and (Younger brothers and sisters = yes) and (Find different study style bet university and secondary = yes) and (low absence rate in Tutorials) then Linger |  |

TABLE V(B): RULES DEDUCE LINGER STUDENTS

| # | Rule statement | Tree of rule |
|---|---|---|
| **3** | If (Allocated time to Entertainment = N0) and (Younger brothers and sisters = yes) and (Secondary_ratio > 70) and (Study Plan Explanation = no)and (high absence rate in Tutorials) and (linger reason = course nature) then linger |  |
| **4** | If (Allocated time to Entertainment = No) and (Younger brothers and sisters = yes) and (secondary_ratio > 70) and (find different study style bet university and secondary = no) and (Marital status = no) and (low absence rate in Tutorials) then linger |  |

Features of Linger Students Class: Table V can be summarized in the following sub classes.

**Sub-Class one:** They ignore to allocate specific time for entertainment, they have young sisters and brothers, and they feel there is different study style between university and secondary school.

**Sub-Class two:** They ignore to allocate specific time for entertainment, they have young sisters and brothers, they feel there is different study style between university and secondary school, and they have low absence rate in Tutorials.

TABLE VI(A): RULES DEDUCE NON-LINGER STUDENTS

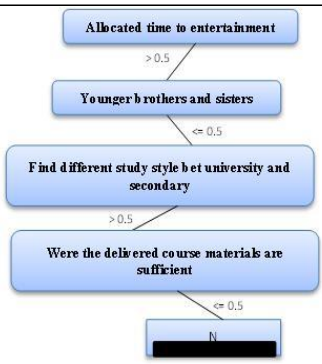
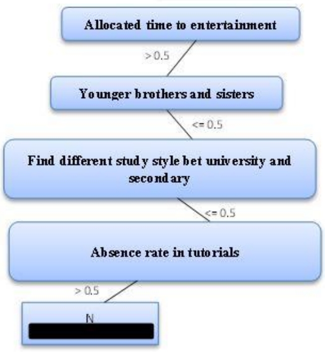
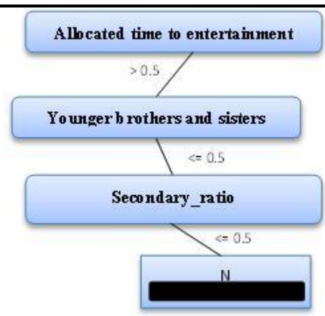
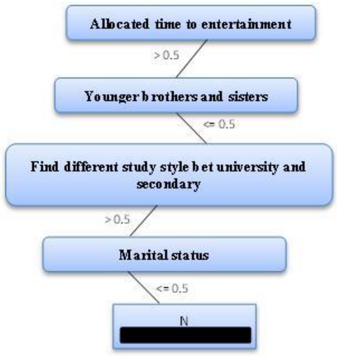
| # | Rule statement | Tree of rule |
|---|---|---|
| 1 | If (Allocated time to Entertainment = N0) and (Younger brothers and sisters = yes) and (Find different study style bet university and secondary = no) and (Were the delivered course materials are sufficient = yes) then not Linger |  |
| 2 | If (Allocated time to Entertainment = N0) and (Younger brothers and sisters = yes) and (Find different study style bet university and secondary = yes) and (high absence rate in Tutorials) then not Linger |  |

TABLE VI(B): RULES DEDUCE NON-LINGER STUDENTS

| # | Rule statement | Tree of rule |
|---|---|---|
| **3** | If (Allocated time to Entertainment = N0) and (Younger brothers and sisters = yes) and (Secondary_ratio < 70) then not Linger |  |
| **4** | If (Allocated time to Entertainment = No) and (Younger brothers and sisters = yes) and (Secondary_ratio > 70) and (find different study style bet university and secondary = no) and (Marital status = yes) then not Linger |  |

**Sub-class three:** They ignore to allocate specific time for entertainment, they have young sisters and brothers, they have secondary ratio greater than 70, they do not get explanation for their study plan, they have high absence rate in Tutorials, and they think that they are linger students due to the nature of their courses.

**Sub-class four:** They ignore to allocate specific time for entertainment, they have young sisters and brothers, they have secondary ratio greater than 70, they feel there is different study style between university and secondary school, they are not married, and they have low absence rate in Tutorials.

Features of Non-Linger Students Class: Table VI cloud be summarized in the following sub classes.

**Sub-class five:** They ignore to allocate specific time for entertainment, they have young sisters and brothers, they feel there is no different study style between university and secondary school, and they feel that the delivered course materials were sufficient.

**Sub-class six:** They ignore to allocate specific time for entertainment, they have young sisters and brothers, they feel there is different study style between university and secondary school, and they have high absence rate in Tutorials.

**Sub-class seven:** They ignore to allocate specific time for entertainment, they have young sisters and brothers, and they have secondary ratio less than 70.

**Sub-class eight:** They ignore to allocate specific time for entertainment, they have young sisters and brothers, and they have secondary ratio less than 70, they feel there is no different study style between university and secondary school, and they are married.

## V. DISCUSSION AND CONCLUSION

In this paper, C4.5 algorithm has been used on student data to classify students based on their study length. Features of linger students have been deduced in four sub classes and same for non-linger students. As conclusion, linger student could be described by combination of factors such as in sub-class one, sub-class two, sub-class three, or sub-class four. Linger student could be any one of these subclasses. For instance, from sub-class one, the students who have these three characteristics together in the same time: 1) there is no schedule for entertainment, 2) they have young sisters and brothers, 3) and feel there is different study style between university and secondary school, those students with these characteristics have high potentially to be linger students.

On the other hand, non-linger students also could be described by combination of f characteristics, such as in sub-class five, sub-class six, sub-class seven, or sub-class eight. For instance, from sub-class six, the students who have these four characteristics together in the same time: 1) there is no schedule for entertainment, 2) they have young sisters and brothers, 3) feel there is different study style between university and secondary school, and 4) they have high absence rate in Tutorials, those students with these characteristics have high potentially to be non-linger students.

The previous two examples illustrate features or characteristics of linger and non-linger students. There are strong similarity between these two classes, for instance, in the previous examples the only difference between linger and non-longer students is that non-linger has one more characteristic which is “high absence rate in Tutorials”. Although characteristic which is “high absence rate in Tutorials” seems strange to be labeled with non-linger student but this is which has been extracted from the data.

In the following, we have used First Order Logic (FOL) for mathematical representing the results from Tables V and VI.

**Suppose:**

A = Allocated time to Entertainment; B = Has younger brothers and sisters; C = Find different study style bet university and secondary; D = Low absence rate in Tutorials;

E = Secondary_ratio > 70; F= Study Plan Explanation; G= Marital status; H = The delivered course materials are sufficient; I= Secondary_ratio < 70; J= linger reason = course nature;

$$\forall A,B,C: (\sim A) \wedge B \wedge C \Longrightarrow \text{Linger} \quad (1)$$

$$\forall A,B,C,D:(\sim A) \wedge B \wedge C \wedge D \Longrightarrow \text{Linger} \quad (2)$$

$$\forall A,B,E,D:(\sim A) \wedge B \wedge E \wedge (\sim F) \wedge (\sim D) \wedge \mathbf{J} \Longrightarrow \text{Linger} \quad (3)$$

$$\forall A,B,C,E,D,G: (\sim A) \wedge B \wedge E \wedge (\sim C) \wedge (\sim G) \wedge D \Longrightarrow \text{Linger} \quad (4)$$

$$\forall A,B,C,H:(\sim A) \wedge B \wedge (\sim C) \wedge H \Longrightarrow \sim(\text{Linger}) \quad (5)$$

$$\forall A,B,C,D:(\sim A) \wedge B \wedge C \wedge (\sim D) \Longrightarrow \sim(\text{Linger}) \quad (6)$$

$$\forall A,B,I:(\sim A) \wedge B \wedge I \Longrightarrow \sim(\text{Linger}) \quad (7)$$

$$\forall A,B,C,E,G: (\sim A) \wedge B \wedge E \wedge (\sim C) \wedge G \Longrightarrow \sim(\text{Linger}) \quad (8)$$

The above eight equations summarized mathematically the results from tables five and six. Equations one to four could be combined as:

$$\forall A,B,C,D,E,G,J:((\sim A) \wedge B \wedge C) \vee ((\sim A) \wedge B \wedge C \wedge D) \vee ((\sim A) \wedge B \wedge E \wedge (\sim F) \wedge (\sim D) \wedge \mathbf{J}) \vee ((\sim A) \wedge B \wedge E \wedge (\sim C) \wedge (\sim G) \wedge D) \Longrightarrow \text{Linger} \quad (A)$$

Equations five to eight could be combined as:

$$\forall A,B,C,D,E,G,I,H: ((\sim A) \wedge B \wedge (\sim C) \wedge H) \vee ((\sim A) \wedge B \wedge C \wedge (\sim D)) \vee ((\sim A) \wedge B \wedge I) \vee ((\sim A) \wedge B \wedge E \wedge (\sim C) \wedge G) \Longrightarrow \sim(\text{Linger}) \quad (B)$$

From equations (A) and (B) we can see the following observations

- Students do not allocate time for entertainment which reflects that students do not have interesting in time management.
- The position of student in his family is critical factor.
- Linger students have alienated feeling, which means they found difficulty to cope with university environment.
- The impact of tutorials attendance rate in linger students is unclear. In sub-class 2 there is low absence rate feature while in sub class three there is high absence rate feature. This ambiguity might be happen due to subjectivity of the questionnaire. The questionnaire is left free to students to answer based on their opinion. In fact, after analyzing the feedback we found that it is better to use an external source of information, such as formal records of attendance.
- Linger students need more effort in study plan explanation.
- The impact of secondary school ratio is unclear. This ambiguity might be happen due to subjectivity of the questionnaire.

This study will help in improving the quality of education and decision making by determining early the students "who" might exceed the planned study length period. We recommend future studies to use directed questionnaire and support it by independent and external source of information to avoid to biases.

## APPENDIX

TABLE IV(A): VARIABLES DESCRIPTION

| Variable | Description | Possible Values | Perspective |
|---|---|---|---|
| Linger | The student is linger or not | Yes, No | Linger Information |
| Linger Source | The source of delaying in study | Personal matter, Courses nature, Teaching methods, university environment, Tabuk environment | |
| Graduation Semester | The expected graduation semester | Current semester, Next semester, After next semester | |
| Preparatory Year | Student face an academic problem or not | Yes, No | |
| Previous Linger | Student is delayed previously or not | Yes, No | |
| Course(s) That Caused Problems | The course(s) that caused academic problem | Programming, Mathematics, English, | |

TABLE IV(B): VARIABLES DESCRIPTION

| Variable | Description | Possible Values | Perspective |
|---|---|---|---|
| Gender | Gender of student | Male-Female | Basic information |
| Age | Age of student | Less than or equal 20 years, Greater than 20 years | |
| Origin | Origin place of student | Inside Tabuk- outside Tabuk | |
| Current Accommodation | The current accommodation of students | With Family- Single | |
| Marital Status | The marital Status | Married, Single, divorced, widower | |
| Offspring | No of children | None, 1, 2, 3, more than 3 | |
| Family Position | The student is oldest brother or not | Yes, No | |
| Siblings | No. of sisters and brothers | None, 1, 2, 3, more than 3 | |
| Father Qualifications | Qualifications of the student's father | Illiterate, Primary School, Middle School, Secondary School, University, Master Degree, PhD Degree | |

TABLE IV(C): VARIABLES DESCRIPTION

| Variable | Description | Possible Values | Perspective |
|---|---|---|---|
| Mother Qualifications | Qualifications of the student's mother | Illiterate, Primary School, Middle School, Secondary School, University, Master Degree, PhD Degree | Basic information |
| Father Alive | Student father is alive or not | Yes, No | |
| Father Job | The job of the student's father | Service, Educational, Commercial, Military, Freelance, Retired | |
| Mother Job | The job of the student's mother | Housewife, Service, Educational, Commercial, Retired | |
| Family Income | The monthly income is enough or not | Yes, No | |

TABLE IV(D): VARIABLES DESCRIPTION

| Variable | Description | Possible Values | Perspective |
|---|---|---|---|
| Smoking | The student is smoking or not | Smoking, Not smoking | Personal information |
| Chronic Diseases | Student has a chronic diseases or not | Yes, No | |
| Vision State | The vision of state of student is using glasses or not | Glasses, No Glasses | |
| Physical Disability | The student has Physical disability or not | Yes, No | |
| Transportation to The University | The type of transportation that the student used to go to the university | Shared Transportation, Bus, Private car | |
| Place of Break Time | The student place in break time | Library, Cafeteria, Outside university | |
| Time Plan | Student has a time plan or not | Yes, No | |
| Lessons Review Time Commitment | student has a commit of review lessons time | Yes, No | |
| Entertainment Time | Spent time in entertainment game | Less than or equal 2 hours, 4 hours, Grater than 4 hours | |
| Enthusiasm for Study | Students has an enthusiasm feel for the study or not | Yes, No | |
| A source of Enthusiasm | The source of no enthusiasm feeling | Courses nature, Teaching methods, university environment, Tabuk environment. | |
| Future Goal(s) | Student has a future goal(s) or not | Yes, No | |
| Relation with Outstanding Students | Student has a relation with outstanding colleagues or not | Yes, No | |
| Future Position Type | The desired position of student in future | Related to computer, Not related to computer | |

TABLE IV(E): VARIABLES DESCRIPTION

| Variable | Description | Possible Values | Perspective |
|---|---|---|---|
| University Entry Ratio | The average ratio for the university entry ratio | 80%-89%, 70%-79%, 60%-69%, 50%-59%, 40%-49% | Academic information |
| Current GPA | The current cumulative average of student | More than 4.5, 3.5-4.4, 2.5-3.4, 2-2.4, Less than 2 | |
| Current Level | The current level of student in the university | First, Second, Third, Fourth | |
| GPA | The current cumulative average of student | More than 4.5, 3.5-4.4, 2.5-3.4, 2-2.4, Less than 2 | |
| Current Major | The current major of the student in the faculty | Computer Science, Information Technology | |
| Major Satisfaction | The level of student satisfaction about major | Bad, Medium, Good | |
| Dissatisfaction Source | The source(s) of dissatisfaction | Courses nature, Teaching methods, university environment, Tabuk environment. | |
| Major Selection | Student who select the major or not | Yes, No | |

TABLE IV(F): VARIABLES DESCRIPTION

| Variable | Description | Possible Values | Perspective |
|---|---|---|---|
| Lectures Absent Rate | The student absent rate of lectures is high or not | Yes, No | Learning and teachings information |
| Labs Absent Rate | The student absent rate of labs is high or not | Yes, No | |
| Tutorials Absent Rate | The student absent rate of tutorials is high or not | Yes, No | |
| Courses Objectives Clarity | The courses objectives are clear or not | Yes, No | |
| Teaching Methods and Efficiency | The teaching methods are enough and effective or not | Yes, No | |
| Course's Materials and textboxes Availability | The courses materials and textboxes are available and effective or not | Yes, No | |
| Classrooms, Labs and Library | The classrooms, labs and library are equipped or not | Yes, No | |
| Lectures, Labs and Tutorials Time | The time of Lectures, labs and tutorials time are appropriate | Yes, No | |
| Lectures, Labs and Tutorials Time Sufficiently | The time of Lectures, labs and tutorials time are sufficient | Yes, No | |

## REFERENCES


[1] A. Norhidayah, J. Kamaruzaman, A. Syukriah, M. Najah, and S. Andin, 'The factors influencing students' performance at Universiti Teknologi MARA Kedah, Malaysia," Canadian Research & Development Center of Sciences and Cultures, vol. 3, no. 4, December 2009.

[2] I. Mushtaq and S. Khan, "Factors affecting students' academic performance," *Global Journal of Management and Business Research*, vol. 12, no. 9, June 2012.

[3] P. Belsis, I. Chalaris, M. Chalaris, Ch. Skourlas, and A. Tsolakidis, "The analysis of the length of studies in higher education based on

clustering and the extraction of association rules," *Dia-Social And Behavioral Sciences*, vol. 147, no. 25, pp. 567–575, August 2014.
[4] Ministry of Education (April 10, 2016). Higher Education Statistics. [Online]. Available: http://www.moe.gov.sa/ar/AcceptedStatus/Pages/default.aspx
[5] Ministry of Education. (April 10, 2016). Higher Education Statistics, University of Tabuk. [Online]. Available: https://www.moe.gov.sa/ar/Ministry/Deputy-Ministry-for-Planning-and-Information-affairs/HESC/Ehsaat/Pages/default.aspx-affairs/HESC/Ehsaat/Pages/default.aspx
[6] E. B. Costa, B. Fonseca, M. A. Santana *et al*., "Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses," *Computers in Human Behavior*, vol. 73, pp. 247-256, 2017.
[7] L. Zhang, L. L. Zhang, W. L. Teng, and Y. B. Chen, "Based on information fusion technique with data mining in the application of finance early-warning," *Procedia Computer Science*, vol. 17, pp. 695-703, 2013.
[8] M. S. Packianather, A. Davies, S. Harraden, S. Soman, and J. White, "Data mining techniques applied to a manufacturing SME," *Procedia CIRP*, vol. 62, pp. 123-128, 2017.
[9] N. Jothi and N. A. Rashid, "Data mining in healthcare – a review," *Procedia Computer Science*, vol. 72, pp. 306-313, 2015.
[10] J. Peral, A. Maté, and M. Marco, "Application of data mining techniques to identify relevant key performance indicators," *Computer Standards & Interfaces*, vol. 54, part 2, November 2017, pp. 76-85
[11] [11] A. Keramati. R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari, and U. Abbasi, "Improved churn prediction in telecommunication industry using data mining techniques," *Applied Soft Computing*, vol. 24, pp. 994-1012, November 2014.
[12] S. Hira and P. S. Deshpande, "data analysis using multidimensional modeling, statistical analysis and data mining on agriculture parameters," *Procedia Computer Science*, vol. 54, 2015, pp. 431-439.
[13] T. F. Bahari and M. S. Elayidom, "An efficient CRM data mining framework for the prediction of customer behaviour," *Procedia Computer Science*, pp. 725-731, vol. 46, 2015.
[14] A. Dursun and M. Caber, "Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis," *Tourism Management Perspectives*, vol. 18, pp. 153–160, April 2016.
[15] C.-H. Chen and Y.-S. Chen, "Classifying the segmentation of customer value via RFM model and RS theory," *Expert Systems with Applications*, vol. 36, pp. 4176–4184, 2009.
[16] S. Natek and M. Zwilling, "Student data mining solution–knowledge management system related to higher education institutions," *Expert Systems with Applications*, vol. 41, issue 14, pp. 6400–6407, October 15, 2014.
[17] K.-M. Osei-Bryson, "Towards supporting expert evaluation of clustering results using a data mining process model," *Information Sciences*, 2010, vol. 180, no. 3, pp. 414–431.
[18] A. P. Sinha and H. Zhao, "Incorporating domain knowledge into data mining classifiers: An application in indirect lending," *Decision Support System*, vol. 46, no.1, pp. 287–299, 2008.
[19] G. K. F. Tso and K. K. W. Yau, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks," *Energy*, vol. 32, pp. 1761–1768, 2007.
[20] S. Wan and T. C. Lei, "A knowledge-based decision support system to analyze the debris-flow problems at Chen-Yu-Lan River, Taiwan," *Knowledge-Based System*, vol. 22, issue 8, pp. 580–588, 2009.
[21] S. H. Zanakis and I. B. Fernandez, "Competitiveness of nations: A knowledge discovery examination," *European Journal of Operational Research*, vol. 166, issue 1, pp. 185–211, 2005.
[22] Z.Y. Zhuang, L. Churilov, F. Burstein, and K. Sikaris, "Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners," *European Journal of Operational Research*, vol. 195, pp. 662–675, 2009.
[23] I. Becerra-Fernandez, A. Gonzales, and R. Sabherwal, "Knowledge management, challenges, solutions, and technologies," Pearson, Prentice Hall, 2004.
[24] N. Zavale, L. Santos, and L. Manuel, "Decision-making in African universities demands rigorous data: Evidence from graduation rates at Eduardo Mondlane University in Mozambique," *International Journal of Educational Development*, vol. 52, pp. 122–134, 2017.
[25] E. Katsikas and T. Panagiotidis, "Student status and academic performance: Accounting for the symptom of long duration of studies in Greece," *Studies in Educational Evaluation*, vol. 37, pp. 152-161, 2011.
[26] M. Zhan and D. Lanesskog, "The impact of family assets and debt on college graduation," *Children and Youth Services Review*, vol. 43, pp. 67–74, 2014.
[27] S. Robertson, C. W. Canary, M. Orr, P. Herberg, and D. N. Rutledge, "Factors related to progression and graduation rates for RN-to-bachelor of science in nursing programs: Searching for realistic benchmarks," *Journal of Professional Nursing*, vol. 26, issue 2, pp. 99–107, March 2010.
[28] D. A Webber and R. Ehrenberg, "Do expenditures other than instructional expenditures affect graduation and persistence rates in American higher education?" Working Paper, Cornell Higher Education Research Institute (CHERI), 2009.

**Osman A. Abdalla** is an assistant professor at the University of Tabuk, Saudi Arabia. Previously, he was an assistant professor at the University of Science and Technology, Sudan. He has participated in many funded projects and research related to artificial intelligence in Malaysia, Sudan and KSA. The fields of his interests' include artificial intelligence, modeling and data mining. He has published many papers that covered his current areas of interest.

**Majed Aborokbah** is the dean in the Faculty of Computing and Information Technology, University of Tabuk, Saudi Arabia. He has received the BSc from Taif University, Saudi Arabia, the Msc. from Bradford University, UK and the PhD from De Montfort University, U.K. His research is in the areas of software engineering, human computer iInteraction and software security. He has published many papers in international journals and conferences, has organized various workshops and conferences in his research areas interest.

**Abdelrahman Osman Elfaki** currently services at the Department of Information Technology at University of Tabuk, Saudi Arabia. Previously, he served at the University of Science and Technology, Malaysia. He published many papers under various computer science topics. He is a member of Center of Robotics and Artificial Intelligence and group leader at programming club at University of Tabuk, Saudi Arabia.