Check for updates

DOI: https://doi.org/10.34069/AI/2024.74.02.22

How to Cite:

Alhamad, I.A., & Singh, H.P. (2024). Predicting dropout at master level using educational data mining: A case of public health students in Saudi Arabia. *Amazonia Investiga*, *13*(74), 264-275. https://doi.org/10.34069/AI/2024.74.02.22

Predicting dropout at master level using educational data mining: A case of public health students in Saudi Arabia

التنبؤ بالتسرب في مستوى الماجستير باستخدام التنقيب في البيانات التعليمية: حالة طلاب الصحة العامة في المملكة العربية السعودية

Received: January 7, 2024 Accepted: February 25, 2024

Written by:

Ibrahim Abdullah Alhamad[1]

https://orcid.org/0000-0001-7099-0335

Harman Preet Singh[2]

https://orcid.org/0000-0003-4297-0016

Abstract

Student dropout and its economic and social consequences are significant issues in developing countries. Students who drop out experience reduced employment prospects and encounter social stigma. While early dropout prediction can assist in mitigating the consequences, it remains a considerable challenge. The present research employed a data mining approach to predict dropout of public health master-level students in Saudi Arabia, a developing nation that has invested considerable resources to promote higher education. The research model focused on three fundamental determinants of students' dropout: individual, institutional, and academic. The study analysis on a dataset of 150 students revealed that all three determinants predicted student dropout. The results indicated that students with low academic performance who received an academic warning were likelier to drop out. Freshmen with poor academic achievement were particularly at risk of dropping out of college. Students between 31 and 36 years old who attended technical courses as a subject specialization could also dropout. The research contributes to the literature by suggesting that universities should consider these individual, institutional, and academic determinants to develop their dropout prevention strategies. This study has ramifications for university administrators in developing nations, such as Saudi Arabia, who can establish dropout

خلاصة

يعد تسرب الطلاب وعواقبه الاقتصادية والاجتماعية من القضايا المهمة في البلدان النامية. يعاني الطلاب الذين يتركون المدرسة من انخفاض فرص العمل ويواجهون ضغوطا اجتماعية. وفي حين أن التنبؤ المبكر بالتسرب يمكن أن يساعد في تخفيف العواقب، إلا أنه لا يزال يمثل تحديًا كبيرًا. استخدم البحث الحالي أسلوب استخراج البيانات للتنبؤ بتسرب طلاب درجة الماجستير في الصحة العامة في المملكة العربية السعودية، وهي دولة نامية استثمرت موارد كبيرة لتعزيز التعليم العالي. ركز النموذج على ثلاث محددات أساسية لتسرب الطلاب: الفردية والمؤسسية والأكاديمية. كشف تحليل الدراسة على مجموعة البيانات المكونة من 150 طالبًا أن المحددات الثلاثة جميعها تنبأت بتسرب الطلاب. أشارت النتائج إلى أن الطلاب ذوي الأداء الأكاديمي المنخفض والذين تلقوا تحذيرًا أكاديميًا كانوا أكثر عرضة للتسرب. كان الطلاب الجدد ذوي التحصيل الأكاديمي الضعيف معرضين بشكل خاص لخطر التسرب من الكلية. يمكن أيضًا للطلاب الذين تتراوح أعمارهم بين 31 و36 عامًا والذين حضروا الدورات الفنية المتخصصة أن يتسربوا من الدراسة. يساهم البحث في الأدبيات من خلال اقتراح أن الجامعات يجب أن تأخذ في الاعتبار هذه المحددات الفردية والمؤسسية والأكاديمية لتطوير استراتيجياتها لمنع التسرب. ولهذه الدراسة تداعيات على قيادات الجامعات في الدول النامية، مثل المملكة العربية السعودية، الذين يمكنهم إنشاء برامج لمنع التسرب بناءً على المحددات التي كشفت عنها هذه الدراسة.

[1] Department of Management and Information Systems, College of Business Administration, University of Ha'il, Kingdom of Saudi Arabia. WoS Researcher ID: HII-6863-2022

[2] Department of Management and Information Systems, College of Business Administration, University of Ha'il, Kingdom of Saudi Arabia. WoS Researcher ID: B-7160-2012

prevention programs based on the determinants revealed in this study.

Keywords: Dropout, educational data mining, higher education institutions, public health, Saudi Arabia, Saudi Vision 2030.

الكلمات المفتاحية: التسرب، التنقيب في البيانات التعليمية، مؤسسات التعليم العالي، الصحة العامة، المملكة العربية السعودية، رؤية السعودية 2030.

Introduction

Higher education is a vital constituent for the progress of a country. Higher education is particularly important for developing countries like Saudi Arabia (Singh et al., 2011a). In 1932, when the Kingdom was established, Saudi Arabia had no university. King Saud University was the inaugural university founded in Saudi Arabia in 1957 (Al-Omar, 2023). Saudi Arabia undertook an ambitious effort in the 1990s to establish a vast array of educational facilities, recognizing the importance of education to the growth of the nation (Singh & Chand, 2012). In Saudi Arabia, the proportion of education spending to GDP climbed from 5.3% in 1985 to 7.3% in 2018 (Singh et al., 2022a).

This illustrates that Saudi Arabia invested extensively in education to enhance human capital and improve its human resources in accordance with the ideals of the Vision 2030 government program (Singh & Alodaynan, 2023). Currently, Saudi Arabia has 67 public and private universities and colleges that deliver higher education to students (CGIJ, 2023).

Students' academic performance during university education is a turning point in their career (Singh et al., 2013). Students obtain employment opportunities depending on their academic performance at the university level (Singh et al., 2022b). The improved performance of students has a favorable effect on the reputation of higher education institutions (HEIs) (Singh & Alhamad, 2022a). However, student dropout negatively influences the employment prospects of the affected students and seriously impacts a nation’s scarce resources (Alhamuddin et al., 2023). Typically, student dropout negatively affects the reputation of the institution and leads to strategic and monetary loss for the nation (Singh et al., 2011b). Further, dropout reduces employment possibilities and leads to societal stigma, resulting in negative social and economic effects for the students (Ye et al., 2022). This is especially critical in the case of public health students, who, following graduation, play a crucial role in educating the population about diverse health issues and their management. The dropout of public health students diminishes the likelihood that governments, particularly in developing nations (Kumari & Singh, 2021) (like Saudi Arabia), will produce sufficient health educators in the future. The dropout of public health students from HEIs is critical for Saudi Arabia, as it has invested significant resources to manage public health (Alam et al., 2022).

Prior research has thoroughly investigated the issue of higher education student dropout. However, most of the prior higher education research has primarily focused on the dropout of undergraduate university students (Baalmann, 2023; Abdulghani et al., 2023). Some studies have examined master's student dropout (Dlungwane & Voce, 2020; Rotem et al., 2020), however, there is a shortage of empirical research on the dropout of public health students, especially in Saudi Arabia. The current study is built on earlier studies but differs from the existing ones in context (Saudi Arabia) and focus (public health).

The current study also employs data mining techniques (Singh & Alhamad, 2022b) to identify the predictive factors affecting dropout of students, which is unique in the Saudi context.

The rest of the paper is arranged as follows. The second section deals with the study objectives. The third section is dedicated to the literature review on educational data mining and students’ dropout. The fourth section deals with the methods employed in this study. The fifth section is related to the results and discussion. The sixth section presents the conclusion of the study. The last section is dedicated to the limitations and future research directions.

Objectives of study

This study aims to develop a predictive model using educational data mining to predict students’ dropout. The following are the objectives of this study:

- To identify the important attributes that help to predict the students’ dropout.

- To identify an appropriate data mining algorithm suitable for building a students' dropout predictive model.
- To apply the selected algorithm to train, test, and build the students' dropout model.
- To suggest strategies to reduce the dropout rate of students.

Literature review on educational data mining and students' dropout

Higher education is deemed essential for a country's progress, particularly in developing nations (Alhulail & Singh, 2023). Early dropout prediction is crucial to mitigate its detrimental impact on the standard of higher education. The commitment of the faculty (Robert, 2023) and technology (Singh et al., 2011c) can play a vital role in the early identification of students at the risk of dropping out (Alkhalil, 2021; Alazemi, 2023). Student dropout rates may negatively impact individuals in developing nations such as Saudi Arabia (Ibeaheem et al., 2018), as it diminishes their job opportunities in nations with constraints on economic resources (Singh & Alwaqaa, 2023).

Educational data mining utilizes education-related data to produce data mining results (Singh et al., 2023; Alhamad & Singh, 2021). In line with the prior research (Gentle et al., 2012; Singh & Alhamad, 2021; Singh & Alhulail, 2023), we employ predictive modelling in this study. Consequently, we undertake a literature review on using educational data mining (EDM) and other empirical methods to examine critical factors affecting student dropout.

Zhang et al. (2010) leveraged EDM to identify key student dropout determinants and improve retention. They conducted the study by collecting institutional data from Thames Valley University students in the United Kingdom (UK). They employed classification algorithms like naïve bayes, decision trees, and support vector machine (SVM) to identify the key predictive factors. The study results indicated that the students' average course marks were a key determinant of their dropout. Further, the study informed that first-year students are more likely to drop out.

Bharadwaj & Pal (2011) employed EDM to conduct a study using institution-specific internal factors to analyze and predict students' retention at VBS Purvanchal University, Jaunpur, India. In their study, the classification method of data mining was used to analyze the performance of 50 MCA students using the classification approach of data mining. From the university database, the study used data such as attendance, class quizzes, seminars, and assignment grades to predict the students' performance. The study employed ID3, C4.5, and ADT decision tree classifiers to predict student performance. Applying the model to the student data of incoming freshmen revealed that the ADT algorithm generated a concise and precise prediction list for student retention. In this study, the decision tree classification technique was tested and found to be highly accurate for improving student retention.

Bharadwaj & Pal (2012) employed EDM to conduct a study based on external factors to predict the retention of students at RML Awadh University, Faizabad, India. Data from 300 students from 5 different degree colleges studying Bachelor of Computer Applications (BCA) was collected for this study. A prediction model for student performance enhancement was developed utilizing the data mining classification method. Bayesian classification model was applied to 17 data attributes of the students. Students' senior secondary exam marks, dwelling location, teaching medium, specialized courses, mother's qualification, family income, and family status were found to be highly connected with their dropout.

Yadav et al. (2012) employed EDM to predict university students' retention using machine learning classifiers. They utilized decision tree algorithms like ID3, C4.5, and ADT to a dataset of selected students from VBS Purvanchal University in Jaunpur, India. The study informed that students' grades play a key role in their continuation of studies. Pal (2012) conducted another study to predict the dropout rate of engineering students at VBS Purvanchal University, Jaunpur, India. They employed decision tree algorithms such as ID3, C4.5, CART, and ADT to identify the key predictive factors influencing the dropout of students. The results of this study revealed that students' high school and secondary school grades, family income, and mother's education level impacted their dropout.

Márquez-Vera et al. (2015) employed EDM to predict early dropout among high school students in Mexico. They proposed a methodology and an ICRM2 algorithm to improve the accuracy of early dropout prediction. They compared the efficacy of their ICRM2 algorithm with traditional data mining algorithms. The study results informed that the developed ICRM2 algorithm could predict dropout within four to six weeks of the start of the course and could be employed to detect early dropout of students. The study results also suggested that students' grades, absenteeism, and motivation level influenced their dropout.

Casanova et al. (2018) utilized EDM to identify the key determinants of first-year students' attrition at a public university in Portugal. The study employed IBM SPSS software version 24 to create a decision tree for predicting the key determinants of students' dropout. The study results revealed that students' academic performance played a key role in determining their dropout. Specifically, low-achieving first-year students were at a higher risk of dropping out. Further factors such as sex, course type, and education level of the mother also affected the dropout of first-year university students.

A study was carried out by Ahmed et al. (2018) in Bahrain to examine the relationship between academic support and student involvement. The study focused on how academic elements influence students' academic engagement and performance. The study findings indicated that increased academic engagement among students contributes to their retention in higher education. Therefore, students who receive academic support and are academically engaged are less likely to drop out.

Rodríguez-Muñiz et al. (2019) leveraged EDM to identify the profiles of students at risk of dropping out at the University of Oviedo, Spain. They employed data mining algorithms such as CART, C4.5, bayesnet, RF, and SVM to identify the risk profiles. The study results informed that students were at a higher risk of dropping out in the first year of university. Specifically, the first-year students with low academic performance were at a higher risk of dropping out. The other factors that influenced students' dropout included students full-time or part-time education, age, and attendance.

Mubarak et al. (2020) utilized EDM to predict the dropout of students studying online at the Open University, UK. They employed data mining algorithms such as SVM, random forest, and decision trees to develop the early prediction model. The study's results indicated that the likelihood of a student dropping out of an online course decreased as their participation in course activities increased. The study also indicated that the study models were 82 percent accurate in predicting the chance of students dropping out during the second week of the online course.

Singh & Alhulail (2022) conducted a study on dropout of students in teacher-training colleges in Ethiopia. The authors employed a four-step logistic regression approach to predict the critical determinants of students dropout in the context of the least-developed country (Ethiopia). The study found that academic variables, such as academic performance and higher education aspirations, have a crucial role in affecting student dropout rates. The study found that cultural influences, family education level, and economic considerations did not significantly affect dropout decisions in Ethiopian teacher training colleges.

Hashim et al. (2024) did an empirical study to evaluate the factors influencing student dropout rates in Malaysia. They analyzed a database of over 100,000 students to identify the dropout determinants. The study indicated that academic characteristics, students' gender, and family financial level influence student dropout rates.

The literature review reveals that individual, institutional, academic, and economic factors predict dropout of students. The individual factor comprises students' age, gender, and parents' education. The institutional factor comprises the student's field of study and the type of course. The academic factor encompasses students' marks, GPA, CGPA, and attendance. The economic factor encompasses family income and the availability of a job for the student. The current study aims to evaluate the impact of these factors on the dropout rate of public health masters' students in Saudi Arabia.

The literature review indicates a lack of empirical studies in Saudi Arabia that have applied dropout factors identified in the literature, including individual, institutional, academic, and economic aspects. Further, the majority of dropout studies utilizing EDM have been undertaken outside Saudi Arabia. In addition, there is a dearth of dropout research employing EDM on students of public health, particularly at the master's level. Therefore, the data lies unused in Saudi universities and is not mined to uncover hidden knowledge regarding the causes of master's students dropping out of public health programs. It is crucial to analyze the factors influencing the dropout of public health students in Saudi Arabia, especially at the master's level, given Saudi Arabia's substantial investment in higher education. The insufficient research in this field warranted the necessity of doing this study. This study aims to fill this research gap.

Methods

In this study, we adapted the six stages of the Cios et al. (2000) model to predict the dropout of public health students.

Stage 1 – Understand Problem Domain

In this stage, we reviewed the literature on EDM to understand the important factors influencing dropout of students. This stage clarified EDM goals in line with Cios et al. (2007). This study considered the individual, institutional, and academic determinants of public health masters' students at the University of Ha'il (UoH), Saudi Arabia. We did not consider the economic factors as information such as income level was not retained by the UoH, Saudi Arabia.

Stage 2 – Understand Data

At this stage, we collected academic data from the UoH, Saudi Arabia. We obtained 161 records for students pursuing a master's in public health. The UoH kept the data of the public health students related to individual, institutional, and academic determinants. The factors corresponding to individual determinants were student ID number, national ID number, name, mobile number, email address, gender, and date of birth. The students' ID number, national ID number, name, mobile number, and email address could have no influence on their dropout, therefore we ignored them. The factor corresponding to institutional determinant was the subject specialty offered to public health masters' students. The factors corresponding to academic determinants were cumulative GPA (CumGPA) (0 to 4), academic warnings (0 to 2), number of credit hours required (45), number of credit hours passed (0 to 45), and number of credit hours remaining (0 to 45). The number of credit hours remaining can be computed as a difference between the number of credit hours required and passed; therefore, we ignored the number of credit hours remaining attribute.

Stage 3 – Prepare Data

In this stage, we evaluated the dataset for missing values and outliers (Han et al., 2012). We found 11 missing values in the dataset; therefore, we ignored those records. There were no outliers found in the dataset. The selected dataset consisted of 150 records. The dataset values should be converted from continuous to categorical for data mining tasks (Liu et al. (2002). For this purpose, first, we derived age, cumulative GPA, and percentage of credit hours completed (CRPER) attributes. We derived the age from the date of birth. The CRPER was derived as a proportion of the number of credit hours passed to the number of required credit hours. We converted the cumulative GPA from a number to a percentage, where a cumulative GPA of 4 denoted 100%. Next, we transformed the age, specialty, cumulative GPA, and CRPER attributes from continuous to categorical.

Table 1 portrays the criteria for data transformation.

Table 1.
Criteria for Data Transformation

Attribute(s)	Continuous Value(s)	Transformed Value(s)
Age	24 to 30	A1
	31 to 36	A2
	37 to 42	A3
Specialty	Electronic Health	S1
	Health Informatics	S2
	Health Service Management	S3
	Hospital and Health Services	S4
	Occupational Health	S5
	Public Health	S6
Cumulative GPA (CumGPA)	>95%	A+
	90% to 94.99%	A
	85% to 89.99%	B+
	80% to 84.99%	B
	75% to 79.99%	C+
	70% to 74.99%	C
	65% to 69.99%	D+
	60% to 64.99%	D
	<60%	F
Percentage of credit hours completed (CRPER)	0% to 25%	CR1
	26% to 50%	CR2
	51% to 75%	CR3
	76% to 100%	CR4

(Source: Authors design)

We transformed the age into three categories. Students aged 24 to 30 were transformed as A1, 31 to 36 as A2, and 37 to 42 as A3. The six subject specialties were given to code from S1 to S6. The cumulative GPA was transformed as per the grading criteria employed by the UoH. Accordingly, cumulative GPA >=95% was transformed as A+, >=90% as A, >=85% as B+, >=80% as B, >=75% as C+, >=70% as C, >=65% as D+, >=60% as D, and <60% as F. The percentage of credit hours completed, ranging from 0% to 25%, was transformed as CR1, 26% to 50% as CR2, 51% to 75% as CR3, and 76% to 100% as CR4.

The final dataset consisted of six non-class attributes and one class attribute. The non-class attributes were gender (male or female), age (A1 to A3), specialty (S1 to S6), cumulative GPA (A+ to F), warnings (0 to 2), and percentage of credit hours completed (CR1 to CR4). The class attribute was status (active or dropout).

Finally, we converted the dataset into Weka software's understandable arff format for data mining.

Stage 4 – Mine Data

In this stage, we employed the classification approach of data mining to predict key determinants of public health master students at the UoH, Saudi Arabia. Prior to experimentation, we balanced the data based on the class attribute (active or dropout) utilizing Weka software's resample filter. Before balancing, the dataset's active and dropout cases were 140 and 10, respectively. Post-balancing, the active and dropout cases in the dataset were 75 each. We conducted experimentation by employing two tree-based and two rule-based algorithms. In accordance with Refaeilzadeh et al. (2009), we did experimentation utilizing ten-fold cross-validation to make sure that the study findings were valid. Table 2 portrays the results of the experiments.

Table 2.
Experimentation Results

Algorithm	**Accuracy-%**	**TPR**	**FPR**	**Precision**	**Recall**	**F-Measure**
J-48	90.67%	0.907	0.093	0.909	0.907	0.907
REP-Tree	86.67%	0.867	0.133	0.869	0.867	0.866
J-Rip	86.00%	0.860	0.140	0.862	0.860	0.860
PART	95.33%	0.953	0.047	0.957	0.953	0.953

Note: The meaning of TPR is true positive rate, whereas FPR is false positive rate.
(Source: Authors design)

The Table 2 experimentation results depict that the PART algorithm performs better than the other three algorithms in terms of accuracy-% (95.33%), TP-rate (0.953), precision (0.957), recall (0.953), and F-measure (0.953). The FP-rate (0.047) of PART is also lower than the other three algorithms. In view of these results, we selected the PART algorithm to extract rules. Only the rules corresponding to the dropout class, having coverage above 5%, and accuracy over 90% were chosen. Table 3 depicts the chosen rules.

Table 3.
PART Rules

S.No.	**Rule(s)**	**Coverage**	**Accuracy**
1	IF (Warnings = 1) AND (CumGPA = F) THEN (Class = Dropout) (36.0/1.0)	24%	97.22%
2	IF (CRPER = CR1) AND (CumGPA = D+) THEN (Class = Dropout) (21.0/1.0)	14%	95.24%
3	IF (CRPER = CR1) AND (Age = A2) THEN (Class = Dropout) (18.0/1.0)	12%	94.44%
4	IF (CumGPA = D+) AND (CRPER = CR1) THEN (Class = Dropout) (15.0/1.0)	10%	93.33%
5	IF (CumGPA = D+) AND (Age = A2) AND (Specialty = S1) THEN (Class = Dropout) (12.0/1.0)	8%	91.67%

(Source: Authors design)

Rule 1 states that if a student has received one academic warning and the cumulative GPA belongs to the F category (means less than 60% or 2.4), then the student is likely to dropout. This rule covered 36 out of 150 instances and gave incorrect result once. Accordingly, the rule coverage and accuracy are 24% and 97.22%, respectively.

Rule 2 informs that if the percentage of credit hours completed by the student belongs to the CR1 category, (means between 0% to 25%) and cumulative GPA belongs to D+ category (means between 65% (2.4) to 69.99% (2.6)), then the student is likely to dropout. This rule covered 21 out of 150 instances and gave incorrect result once. Accordingly, the rule coverage and accuracy are 14% and 95.24%, respectively.

Rule 3 asserts that if the percentage of credit hours completed by the student belongs to the CR1 category (means between 0% to 25%) and the age belongs to the A2 category (means between 31 and 36 years), then the student will likely drop out. This rule covered 18 out of 150 instances and gave incorrect result once. Accordingly, the rule coverage and accuracy are 12% and 94.44%, respectively.

Rule 4 states that if the cumulative GPA belongs to the D+ category (means between 65% (2.4) to 69.99% (2.6)) and the percentage of credit hours completed by the student belongs to the CR1 category (means between 0% to 25%), then the student is likely to dropout. This rule covered 15 out of 150 instances and gave incorrect result once. Accordingly, the rule coverage and accuracy are 10% and 93.33%, respectively.

Rule 5 informs that if the cumulative GPA belongs to the D+ category (means between 65% (2.4) to 69.99% (2.6)) and age belongs to the A2 category (means between 31 and 36 years), then the student is likely to dropout. This rule covered 12 out of 150 instances and gave incorrect result once. Accordingly, the rule coverage and accuracy are 8% and 91.67%, respectively.

Stage 5 – Evaluate Knowledge

In this stage, we discussed the chosen rules with the domain experts. The domain experts agreed with all the five rules. Consequently, we accepted all the five rules.

Stage 6 – Utilize Knowledge

In this stage, we utilized the five rules to identify the determinants of health science masters' students' dropout at the UoH.

Table 4.
Critical Dropout Determinants

Rule	Age	Specialty	CumGPA	Warnings	CRPER	Class
R-1			F	1		Dropout
R-2			D+		CR1	Dropout
R-3	A2				CR1	Dropout
R-4			D+		CR1	Dropout
R-5	A2	S1	D+			Dropout
	Individual	Institutional	Academic			Dropout

(Source: Authors design)

Table 4 depicts that individual, institutional, and academic determinants affect public health university student dropout in Saudi Arabia. Students who have a low cumulative GPA (less than 2.6) are at risk of dropping out. Students with a GPA of less than 2.4 and who have received an academic warning are especially likely to drop out. The dropout risk is higher in the case of students who have completed less than 25% of their credit hours. The freshmen who have a low GPA and belong to the age group 31 to 36 years are likely to dropout. Also, the low-performing students aged 31 to 36 years and have taken the subject specialty of electronic health could drop out.

Results and discussion

According to the current study findings, student attrition has multiple causes, such as individual, institutional, and academic determinants. The study results show that a student's age is an individual determinant of dropout. In this regard, the current study's findings are consistent with those of Rodríguez-Muñiz et al. (2019), who found that a student's age is one of the factors influencing their dropout. However, the current study results contribute to the literature by indicating that low-performing public health master's students between the ages of 31 and 36 are more likely to drop out.

The current study results reveal that the subject specialty taken by the students could be a determinant of their dropout. This could be a significant determinant of dropout for students aged 31 to 36 whose academic performance is below average. In this regard, the current study's findings resonate with Bharadwaj & Pal (2012), who stated that the subject specialty taken by students could influence their dropout. However, the present study contributes to the body of knowledge by revealing that students between the ages of 31 and 36 who have taken a subject specialty in electronic health could drop out.

The present study further informs that student's cumulative GPA, academic warnings, and percentage of credit hours completed are the academic determinants of their dropout. This study indicates that students with low academic performance are at risk of dropping out. This result is in alignment with Casanova et al. (2018), Ahmed et al. (2018), Mubarak et al. (2020), Singh & Alhulail (2022), and Hashim et al. (2024) who anticipated that students with low academic performance could dropout. The current study enriches the body of knowledge by suggesting that public health master students with a cumulative GPA of less than 2.60 are at risk of dropping out. The students could be given academic warnings due to their low attendance in class. This result is aligned with Bharadwaj & Pal (2011) and Rodríguez-Muñiz et al. (2019), who found that students with low-class attendance may drop out. The present study also suggests that low-performing students who have completed fewer than 25 percent of credit hours (freshmen) are more likely to dropout. This result is aligned with Zhang et al. (2010), Casanova et al. (2018), and Rodríguez-Muñiz et al. (2019), who found that low-performing freshmen have a greater probability of dropping out.

The study findings also depict a relationship between individual, institutional, and academic determinants of dropout. Age is an individual determinant of dropout, the subject specialty is an institutional determinant of dropout, whereas cumulative GPA, academic warnings, and percentage of credit hours completed are academic determinants of dropout. The results of the study indicate that masters' of public health students between the ages of 31 and 36, who are specializing in electronic health and exhibit low academic performance in terms of CGPA and attendance, are more likely to drop out.

Conclusions

The issue of student dropout is a complex issue due to the long-term effects it can have on a student's life and career. Student dropout is a serious issue for Saudi Arabia as it has invested significant resources to promote its education sector. Particularly, the dropout of public health masters' students is more serious as it reduces the supply of adequate health educators for the country. In addition, Saudi Arabia has invested large resources to upgrade its education and health sectors as part of its Vision 2030 program (Saudi Gazette, 2020); hence, Saudi educational institutions should have dropout prevention measures in place.

In this study, we employed the data mining's classification approach to identify the key determinants of student dropout in the public health master's program at the UoH. We adapted the Cios et al. (2000) model to conduct the data mining analysis. We analyzed 150 student records to identify the key determinants of students' dropout. The analysis was conducted using two tree-based (J-48 and Rep-Tree) and two rule-based algorithms (J-Rip and PART). The PART method was used to extract the rules because it performed the best among the four algorithms. The extracted knowledge informed that individual (age), institutional (specialty), and academic (cumulative GPA, academic warnings, and percentage of credit hours completed) determinants affect public health university student dropout in Saudi Arabia.

The study reveals that low-performing students who have been issued an academic warning are likelier to drop out. Freshmen with low academic performance are especially vulnerable to dropping out of university. The study also suggests that students between the ages of 31 and 36 who have studied technical courses as a subject specialization (e.g., electronic health) may drop out.

To develop dropout prevention strategies, universities should consider these individual, institutional, and academic determinants. The universities should pay close attention to the academic performance of the freshmen. The freshmen whose performance is below average should be offered remedial classes. Further, they should be given academic counseling to discourage them from dropping out. The universities should also monitor the performance of the students in technical courses (e.g., electronic health). The students who perform below average in the technical courses should also be given remedial classes to improve their performance. The universities should increase the practical content of the technical courses to improve students' understanding of these courses.

Limitations and future research

The present study has shortcomings that can be resolved in future research. The first limitation pertains to the sample size of 161 public health master's students. We obtained data from the UoH, where a sample of 161 students pursuing a master's degree in public health was available. Future research can opt for a larger sample size to enhance the generalizability of the research findings. The current study exclusively utilized data from a single public university in Saudi Arabia. Future research can gather data from multiple public and private universities in Saudi Arabia to extrapolate the research results. Future research can also gather data from universities in other Gulf Cooperation Council countries due to their analogous political, economic, and social characteristics to Saudi Arabia.

The second limitation is the selection of study variables. The study focused on identifying factors contributing to student dropout, such as individual, institutional, and academic variables. The current research did not consider economic factors because student economic data from the UoH was unavailable. Subsequent studies can gather this data using questionnaires to enhance the existing research results.

The current study used data mining techniques to investigate the UoH public health masters' student dropout rates. The data mining technique has its advantages but relies on classified data rather than continuous data. The regression analysis utilizes continuous data to uncover subtle insights. Future research can utilize regression analysis as a robustness check on the current research findings.

Lastly, the current study focused on master's students in public health. Future research can select both undergraduate and graduate students majoring in public health to extend the research findings to the field of public health. Future research can select students from various disciplines, such as Applied Sciences and Engineering, to compare dropout profiles and draw generalized research findings.

Bibliographic references

Abdulghani, H. M., Alanazi, K., Alotaibi, R., Alsubeeh, N. A., Ahmad, T., & Haque, S. (2023). Prevalence of potential dropout thoughts and their influential factors among Saudi medical students. *SAGE Open, 13*(1), 215824402211469. https://doi.org/10.1177/21582440221146966

Ahmed, U., Umrani, W. A., Qureshi, M. A., & Samad, A. (2018). Examining the links between teachers support, academic efficacy, academic resilience, and student engagement in Bahrain. *International Journal of Advanced and Applied Sciences*, 5(9), 39-46. https://doi.org/10.21833/ijaas.2018.09.008

Alam, F., Singh, H. P., & Singh, A. (2022). Economic Growth in Saudi Arabia through Sectoral Reallocation of Government Expenditures. *SAGE Open,* 12(4), 1-13. https://doi.org/10.1177/21582440221127158

Alazemi, N. F. S. A. (2023). The use of information technology applications by faculty members at the college of basic education in the public authority for applied education and training in the state of Kuwait. *International Journal of Advanced and Applied Sciences*, 10(3), 136-142. https://doi.org/10.21833/ijaas.2023.03.018

Alhamad, I. A., & Singh, H. P. (2021). Decoding Significant and Trivial Factors Influencing Online Hotel Ratings: The Case of Saudi Arabia's Makkah City. *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies*, 12(7), 12A7H, 1-11. https://doi.org/10.14456/ITJEMAST.2021.134

Alhamuddin, A., Inten, D. N., Mulyani, D., Suganda, A. D., Juhji, J., Prachagool, V., & Nuangchalerm, P. (2023). Multiple intelligence-based differential learning on critical thinking skills of higher education students. *International Journal of Advanced and Applied Sciences*, 10(8), 132-139. https://doi.org/10.21833/ijaas.2023.08.015

Alhulail, H. N., & Singh, H. P. (2023). Impact of multimedia technology on university students learning agility and creativity. *Amazonia Investiga*, *12*(70), 189-199. https://doi.org/10.34069/AI/2023.70.10.17

Alkhalil, A. (2021). Decision support model to adopt big data analytics in higher education systems. *International Journal of Advanced and Applied Sciences*, 8(6), 67-78. https://doi.org/10.21833/ijaas.2021.06.008

Al-Omar, B. (2023). About King Saud University. King Saud University Accessed

December 12, 2023. from https://tinyurl.com/y5cek94f

Baalmann, T. (2023). Health-Related Quality of Life, Success Probability and Students' Dropout Intentions: Evidence from a German Longitudinal Study. *Research in Higher Education*. 65. https://doi.org/10.1007/s11162-023-09738-7

Bharadwaj, B. K., & Pal, S. (2011). Mining educational data to analyze students performance. *International Journal of Advanced Computer Science and Applications*, *2*(6). https://doi.org/10.14569/ijacsa.2011.020609

Bharadwaj, B. K., & Pal, S. (2012). Data Mining: A prediction for performance improvement using classification. *International Journal of Computer Science and Information Security*, *9*(4), 1-5. https://tinyurl.com/mvzvun6v

Casanova, J. R., Cervero, A., Núñez, J. C., Almeida, L. S., & Bernardo, A. (2018). Factors that determine the persistence and dropout of university students. *Psicothema*, *30*(4), 408-414. https://doi.org/10.7334/psicothema2018.155

CGIJ. (2023). *List of universities and colleges in Saudi Arabia*. Consulate General of India, Jeddah. Retrieved December 16, 2023, from http://www.cgijeddah.com/listofuniversity.pdf

Cios, K. J., Pedrycz, W., Swiniarski, R. W., & Kurgan, L. A. (2007). *Data Mining: A Knowledge Discovery Approach* (1st ed.). New York, NY: Springer US. https://doi.org/10.1007/978-0-387-36795-8

Cios, K. J., Teresinska, A., Konieczna, S., Potocka, J., & Sharma, S. (2000). A knowledge discovery approach to diagnosing myocardial perfusion. *IEEE Engineering in Medicine and Biology Magazine*, *19*(4), 17-25. https://doi.org/10.1109/51.853478

Dlungwane, T., & Voce, A. (2020). Exploring student persistence to completion in a Master of Public Health programme in South Africa. *African Journal of Health Professions Education*, *12*(1), 17. https://doi.org/10.7196/ajhpe.2020.v12i1.1183

Gentle, J. E., Härdle, W. K., & Mori, Y. (Eds.). (2012). *Handbook of Computational Statistics: Concepts and Methods* (2nd ed., Springer Handbooks of Computational Statistics). Heidelberg, Germany: Springer-Verlag. https://doi.org/10.1007/978-3-642-21551-3

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Waltham, MA: Morgan Kaufmann. https://doi.org/10.1016/C2009-0-61819-5

Hashim, R. A., Lim, H. E., Jafar, M. F., Shanmugam, S. K. S., & Bukhari, N. (2024). Statistical identification of predictors of dropout in secondary education: evidence from Malaysia. *Journal of the Asia Pacific Economy*, 1-27. https://doi.org/10.1080/13547860.2024.2306673

Ibeaheem, H. A., Elawady, S., & Ragmoun, W. (2018). Saudi Universities and higher education skills on Saudi Arabia. *International Journal of Higher Education Management*, *04*(02), 1-14. https://doi.org/10.24052/ijhem/v04n02/art05

Kumari, R., & Singh, H. P. (2022). Role of Incident Reporting System in Healthcare Management: A Case of Multispeciality Tertiary Hospital in India. *International Journal of Information Movement*, 6(IX), 12-18. https://tinyurl.com/38pejmwb

Liu, H., Hussain, F., Tan, C. L., & Dash, M. (2002). Discretization: An Enabling Technique. *Data Mining and Knowledge Discovery*, *6*, 393-423. https://doi.org/10.1023/a:1016304305535

Márquez-Vera, C., Cano, A., Romero, C., Noaman, A. Y., Fardoun, H. M., & Ventura, S. (2015). Early Dropout Prediction using Data Mining: A Case Study with High School Students. *Expert Systems*, *33*(1), 107-124. https://doi.org/10.1111/exsy.12135

Mubarak, A. A., Cao, H., & Zhang, W. (2020). Prediction of students' early dropout based on their interaction logs in online learning environment. *Interactive Learning Environments*, *30*(8), 1414-1433. https://doi.org/10.1080/10494820.2020.1727529

Pal, S. (2012). Mining Educational Data to Reduce Dropout Rates of Engineering Students. *International Journal of Information Engineering and Electronic Business*, *4*(2), 1-7. https://doi.org/10.5815/ijieeb.2012.02.01

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-Validation. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of Database Systems* (pp. 532–538). Springer, Boston, MA. https://doi.org/10.1007/978-0-387-39940-9_565

Robert, K. J. B. (2023). Faculty commitment and performance in Montfortian educational institutions: Basis for a faculty development program. *International Journal of Advanced and Applied Sciences*, 10(2), 113-127. https://doi.org/10.21833/ijaas.2023.02.015

Rodríguez-Muñiz, L. J., Bernardo, A. B., Esteban, M., & Díaz, I. (2019). Dropout and transfer paths: What are the risky profiles when analyzing university persistence with machine learning techniques? *Plos One, 14*(6), 1-20. https://doi.org/10.1371/journal.pone.0218796

Rotem, N., Yair, G., & Shustak, E. (2020). Dropping out of master's degrees: objective predictors and subjective reasons. *Higher Education Research and Development, 40*(5), 1070-1084. https://doi.org/10.1080/07294360.2020.1799951

Saudi Gazette. (2020, May 20). *Full text of Saudi Arabia's Vision 2030 | Al Arabiya English*. https://tinyurl.com/4nt5pxan

Singh, A., Singh, H. P., Alam, F., & Agrawal, V. (2022b). Role of Education, Training, and E-Learning in Sustainable Employment Generation and Social Empowerment in Saudi Arabia. *Sustainability, 14*(14), 8822. https://doi.org/10.3390/su14148822

Singh, H. P., & Alhamad, I. A. (2022a). Influence of National Culture on Perspectives and Factors Affecting Student Dropout: A Comparative Study of Australia, Saudi Arabia, and Ethiopia. *Archives of Business Research, 10*(11), 287-300. https://doi.org/10.14738/abr.1011.13508

Singh, H. P., & Alhamad, I. A. (2022b). A Data Mining Approach to Predict Key Factors Impacting University Students Dropout in a Least Developed Economy. *Archives of Business Research, 10*(12), 48-59. https://doi.org/10.14738/abr.1012.13556

Singh, H. P., & Alhulail, H. N. (2022). Predicting Student-Teachers Dropout Risk and Early Identification: A Four-Step Logistic Regression Approach. *IEEE Access, 10*, 6470-6482. https://doi.org/10.1109/access.2022.3141992

Singh, H. P., & Alodaynan, A. M. M. (2023). The role of educational technology in developing the cognitive and communicative skills of university students: A Saudi Arabian case. *International Journal of Advanced and Applied Sciences, 10*(7), 157-164. https://doi.org/10.21833/ijaas.2023.07.017

Singh, H. P., & Chand, P. (2012). ICT Education: Challenges and Opportunities. In D. Parimala (Ed.), *Role of Teachers in Changing Context: Policy and Practice* (1st ed., pp. 255–263). Kanishka Publishers, Distributors. https://tinyurl.com/3cvexykt

Singh, H. P., Agarwal, A., & Das, J. K. (2013). Implementation of E-Learning in Adult Education: A Roadmap. *Mumukshu Journal of Humanities, 5*(1), 229-232. https://tinyurl.com/yfcws7rw

Singh, H. P., Alshallaqi, M., & Altamimi, M. (2023). Predicting Critical Factors Impacting Hotel Online Ratings: A Comparison of Religious and Commercial Destinations in Saudi Arabia. *Sustainability, 15*(15), 11998. https://doi.org/10.3390/su151511998

Singh, H. P., Jindal, S., & Kaurav, R. P. S. (2011a). *Adult Education and E-Learning. In Proceedings of the National Conference on Turbulent Business Environment: The Road Ahead*. Rohini, Delhi, India; Gitarattan International Business School (giBS). https://tinyurl.com/ru9dhne7

Singh, H. P., Jindal, S., & Samim, S. A. (2011b). A Critical Study on Adoption of E-Learning for Development of Human Resources in Developing Countries. *Mumukshu Journal of Humanities*, 3(3), 116-120.

Singh, H. P., Jindal, S., & Samim, S. A. (2011c). Role of Human Resource Information System in Banking Industry of Developing Countries. *Special Issue of the International Journal of the Computer, the Internet and Management*, 19(SP1), 59.1-59.5. https://bit.ly/3coQmWw

Singh, H., & Alhulail, H. N. (2023). Information Technology Governance and Corporate Boards' Relationship with Companies' Performance and Earnings Management: A Longitudinal Approach. *Sustainability, 15*(8), 6492. https://doi.org/10.3390/su15086492

Singh, H., Singh, A., Alam, F., & Agrawal, V. (2022a). Impact of Sustainable Development Goals on Economic Growth in Saudi Arabia: Role of Education and Training. *Sustainability, 14*(21), 14119. https://doi.org/10.3390/su142114119

Singh, H.P., & Alhamad, I. A. (2021). Deciphering Key Factors Impacting Online Hotel Ratings Through the Lens of Two-Factor Theory: A Case of Hotels in Makkah City of Saudi Arabia. *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies*, 12(8), 12A8M, 1-12. https://doi.org/10.14456/ITJEMAST.2021.160

Singh, H.P., & Alwaqaa, M. A. M. (2023). The educational technology's impact on youth creativity and innovation: A case of Ha'il region of Saudi Arabia. *Amazonia Investiga, 12*(66), 144-154. https://doi.org/10.34069/AI/2023.66.06.14

Yadav, S. K., Bharadwaj, B., & Pal, S. (2012). Mining Education Data to Predict Student's Retention: A Comparative Study. *International Journal of Computer Science* *and Information Security*, *10*(2), 113-117. https://tinyurl.com/4jndx6dt

Ye, X., Zhai, M., Feng, L., Xie, A., Wang, W., & Wu, H. (2022). Still want to be a doctor? Medical student dropout in the era of COVID-19. *Journal of Economic Behavior and Organization, 195*, 122-139. https://doi.org/10.1016/j.jebo.2021.12.034

Zhang, Y., Oussena, S., Clark, T., & Kim, H. (2010). Use data mining to improve student retention in higher education - A case study. In ICEIS - *12th International Conference on Enterprise Information Systems*. Retrieved December 27, 2023, from http://shura.shu.ac.uk/11970/