SEBHA UNIVERSITY JOURNAL OF PURE & APPLIED SCIENCES VOL.23 NO. 1 2024
DOI: 10.51984/JOPAS.V23I1.2848

مجلة جامعة سبها للعلوم البحتة والتطبيقية
Sebha University Journal of Pure & Applied Sciences

Journal homepage: www.sebhau.edu.ly/journal/jopas

Discover Frequent Patterns from Academic Data Of Student Information System

*Almahdi Alshareef, Hana Safour

Faculty of Information Technology, Sebha University, Libya

Keywords:
Index data mining
Education data mining
Association rules
Apriori algorithm

ABSTRACT

Numerous researchers have explored the realm of data mining in education. The primary goal is knowledge discovery, aiming to support staff in efficiently managing educational units, refining student activities, and ultimately elevating learning outcomes. In this study, we utilize association rules mining, implementing the Apriori algorithm to extract insights from academic datasets sourced from the student information system of Sebha University, Libya. Genuine data is sourced from the cloud server. The algorithm is then applied to unveil relationships among 11 attributes within students' academic records spanning four years. The resulting patterns undergo experimental evaluation, considering support and confidence values. These specific rules are subsequently categorised into four classes and scrutinised for further validation. The proposed method yields valuable patterns pertaining to students' academic progress and retains crucial insights for predicting decisions regarding course additions and drops.

اكتشاف الأنماط المتكررة من البيانات الأكاديمية لنظام معلومات الطلاب

*المهدي محمد الشريف، هناء إبراهيم صفور

كلية تقنية المعلومات، جامعة سبها، ليبيا

الكلمات المفتاحية:
تعدين بيانات الفهرس
التنقيب عن البيانات التعليمية
قواعد الارتباط
خوارزمية *Apriori*

الملخص

قام العديد من الباحثين باستخراج البيانات في البيانات التعليمية المخزنة في قواعد بيانات البرامج التعليمية وأنظمة إدارة التعلم. الهدف هو اكتشاف المعرفة الذي يمكن أن يساعد الموظفين على دعم طلابهم من خلال إدارة الوحدات التعليمية بشكل فعال ، وإعادة تصميم أنشطة الطلاب ، وتحسين نتائج التعلم في النهاية. تتعلق تقنية استخراج البيانات الأساسية باكتشاف الارتباطات المخفية الموجودة في البيانات المخزنة في قواعد بيانات البرامج التعليمية. في هذه الورقة ، نطبق تعدين قواعد الارتباط الذي ينفذ تعدين خوارزمية *Apriori* من مجموعات البيانات الأكاديمية لنظام معلومات الطلاب بجامعة سبها ليبيا ، يتم جمع البيانات الحقيقية من الخادم السحابي. يتم تطبيق خوارزمية النمط لاستخراج العلاقة بين 11 سمة من سمات السجلات الأكاديمية للطالب في فترات مدتها أربع سنوات. يتم تقييم الأنماط بشكل تجريبي من خلال قيم الدعم والثقة بينما يتم تعيين بعض القواعد المحددة إلى أربع فئات وتحليلها لمزيد من التحقق. تنتج الدراسة المقترحة أنماطًا قيّمة من الطلاب الذين يدرسون وتحافظ على المعرفة المهمة للتنبؤ بالإضافة والإسقاط

Introduction

Many researchers have conducted studies on data mining in education, focusing on the data stored in databases of educational software and Learning Management Systems. The primary objective is knowledge discovery to assist staff in effectively managing educational units[1], redesigning student activities, and ultimately enhancing learning outcomes. A fundamental data mining technique involves uncovering hidden associations within the data stored in educational software databases.

In this paper, we employ association rules mining, implementing the Apriori algorithm [2] to mine academic datasets from the student information system of Sebha University, Libya. Real data is collected from the cloud server. The algorithm is applied to extract relationships among 11 attributes of students' academic records over four years. The resulting patterns are experimentally evaluated based on support and confidence values. Specific rules are then mapped to four classes and analyzed for further verification.

The proposed study yields valuable patterns related to students' academic progress and preserves essential knowledge for predicting

*Corresponding author:
E-mail addresses: alm.alshareef@sebhau.edu.ly, (H. Safour) han.sofer@sebhau.edu.ly

Article History : Received 26 November 2023 - Received in revised form 28 February 2024 - Accepted 18 March 2024

course add and drop decisions[3].

It is one of the essential tasks in data mining to discover frequent patterns from previous datasets [4]. Patterns represent a frequent occurrence of serial events in a database, constituting a common form of data that may contain crucial knowledge to be uncovered. Finding frequent patterns can provide interesting insights to experts in various domains, especially in areas such as alarm log analysis, financial events, and stock trend relationship analysis.

Education is a fundamental element of a country's progress, and data mining proves to be highly useful in this context. Data mining, in this sense, refers to the method of uncovering hidden, previously unknown, and potentially significant tasks within a vast amount of data [5].

Educational institutions can utilize data mining to extract valuable information from their databases, a field known as Educational Data Mining (EDM) [6]. EDM requires the application of innovative approaches derived from statistics, automated learning, psychological measurement, scientific computing, etc., to better understand students. Educational Data Mining offers a range of technologies to support the education system. The methodology followed in Educational Data Mining mirrors the traditional methods of data mining, involving understanding the environment to be handled, compiling, cleaning, and arranging data, selecting applicable techniques, interpreting results, and verifying the validity of applied technologies.

EDM is employed to extract knowledge patterns from the academic data in the student database of Sebha University (SUL) and to determine optimal solutions. This information may assist in predicting the courses students should add or drop each semester and recommend suitable courses for new students.

The Student Information System [7] is one of the recently implemented electronic systems at SUL, assisting students and teaching staff in managing the educational process. It encompasses tasks such as student registration, course addition, examinations, attendance, absences, grade monitoring, and graduation.

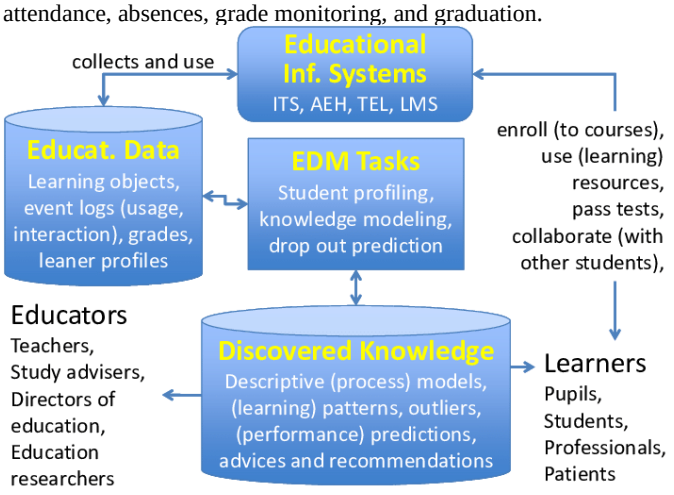

Fig. 1: Education data mining process

In the education system Fig 1 despite the existence of this electronic system, students, especially newcomers, still face challenges related to adding and dropping courses and choosing appropriate courses at the start of university life. With the College Hierarchy Law preventing students from studying a course based on a previous one, students need guidance to navigate these decisions without issues affecting their morale and academic performance.

To address this, some colleges have academic assistants with extensive experience guiding students on the right path and suitable course selection based on their college records.

However, the electronic system presents challenges as students require immediate assistance without seeking an academic advisor. This paper explores the possibility of studying and analyzing previous student records (specifically, records of adding and dropping courses over several years) to identify problems faced by former students and patterns that can help prevent these problems in the future.

The study focuses on mining frequent patterns in academic student data (ASD) collected from Sebha University's cloud server (SUCS). The primary goal is to identify frequent patterns in ASD, categorized into four main events: Fail (F), Pass (P), Good (G), and Very Good (V). We apply a frequent algorithm to find the most common patterns during different time periods. This paper is organized as follows: Section 2 presents related work on association rule mining, Section 3 explains the proposed algorithm, and Sections 4 and 5 cover the experimental results and conclusions, respectively.

Related work

Researchers have conducted extensive studies on observing the performance of university students since the end of the last century. Many studies have focused on applying machine learning algorithms to discover knowledge from university databases, as outlined below:

In[8], a data mining technique in higher education was implemented to examine various faculty factors affecting the quality of productive students in the global marketplace. The findings highlighted the crucial role of quality staff in promoting higher education, emphasizing the need for efforts to improve student performance in lower classrooms.

[9] introduces a model for modeling and predicting academic performance based on academic records and forum participation. Three data mining techniques (Naïve Bayes, Neural Network, and Decision Tree) were employed, with Naïve Bayes outperforming others, achieving an overall prediction accuracy of 86%. This study aims to assist teachers in enhancing academic performance.

Similarly, [10] introduces a new student performance prediction model with additional features, incorporating various aspects of student records and demographic background. Rule-Based proved to be the most effective among other techniques, achieving an accuracy of up to 71.3%. The model successfully predicted the success level of first-year students.

In addition, [7] presents another data mining technique aiming to discover patterns through the use of algorithms on educational data. The study focuses on enhancing decision-making by identifying at-risk students, reducing student repetition rates, increasing success rates, and improving learning outcomes. The proposed method successfully detected 25 different types of survey data, revealing high-frequency patterns that could improve overall student performance.

This paper [8] introduces a framework for predicting learner performance in higher education institutions by employing suitable data mining techniques. The data utilized is primarily collected from the Learning Management System (LMS) and other systems within the organization. The aim of this framework is to assist institutions in making informed decisions regarding learners' performance. Various clustering algorithms are proposed for application to the data collected from diverse activities within the LMS.

This research [11] investigates the application of association rules after collecting student data over the years from the Learning Management System (LMS). Interestingness metrics and other relevant student preferences are captured, considering several association rule algorithms. Visual presentations of rules and their corresponding results will be demonstrated in terms of performance metrics, highlighting their suitability in eLearning environments.

Further, [10] utilizes the Apriori and K-means algorithms in educational databases to extract and define students' academic performance based on various criteria. This study aids in predicting academic trends and patterns, allowing for the classification of students into good, medium, or weak groups and helping lecturers identify hidden patterns in student learning.

[2] implements the Apriori algorithm to discover meaningful patterns in the relationship between mathematics scores and problem-solving patterns of students. The study analyzes the correlation of solution patterns to specific problems through association rule search.

Another association algorithm is introduced in [12], proposing the FP-Growth algorithm to find association rules between placement test questions and grades in a programming fundamental course. The study analyzes common knowledge or skills of students who achieve good grades, suggesting essential skills for computer science students.

This study [13] aims to explore various attributes related to academic students' achievements. The subjects of the study were 3,509 graduated students from the Applied Science Private University in Jordan, encompassing all department programs of the Faculty of Information Technology. The objective is to analyze students' profiles and identify their relation to and association with their university GPA (Grade Point Average) by selecting four main attributes: department, secondary certificate grade, secondary certificate city, and gender. The study then seeks to determine the highest impacting value for each

attribute and generate main association rules using data mining classifiers such as Decision Tree, Naïve Bayes, and association rules. Ultimately, the results of this study may assist current and incoming faculty students in predicting their potential GPA, enabling them to monitor and enhance their performance.

This paper [14] focuses on association rule mining technique to elicit pragmatic information from the primary data collected. The work carried here explores the relationship between student's intellectual performance and their psychological factors with the use of Apriori algorithm. It aims at predicting and improving academic performance of students by reshaping their psychological parameters.

Given the success of association rules mining in education data, particularly the Apriori algorithm, this study gives attention to applying it to find patterns that could help students at Sebha University overcome the problem of choosing optimal courses to start their academic journey.

Preprocessing Data

The data were collected from the ASD of Sebha University, College of Science, over a period of up to four academic years, focusing solely on result data. The following table lists the attributes extracted from the student data.

Table 1: Shows attributes of student data

No	Attributes
1	Referince No
2	Student_No
3	Semster
4	Course_name
5	Stu_smester_No
6	Mid_exam_grade
7	final_exam_grade
8	Total_grade
9	GPG
10	Course_add_date
11	Nationalty

After retrieving data from the ASD, the data processing phase commenced. Initially, the data underwent cleaning to remove repetitions, noise, and missing values. Additionally, segmentation and transformation algorithms in data tracking were employed to refine the dataset. The table below illustrates the data's appearance before the processing and selection stages.

Table 2: Example of original data

Depertment	faculty	Gender	Course name	Course code	Course Type
computer	faculty sciences	male	Internet programming	CS702	major
computer	faculty sciences	male	Operation system	CS801	major
computer	faculty sciences	male	Introduction To Arti	CS808	major
computer	faculty sciences	male	Linear Algebra I	MA604	supporting
computer	faculty sciences	male	Linear Algebra I	MA604	major
computer	faculty sciences	male	Linear Algebra I	MA604	supporting
computer	faculty sciences	male	Electronics I	PH605	supporting
computer	faculty sciences	male	Electronics I	PH605	major
computer	faculty sciences	male	Compiler Constructio	CS802	major
Depertment	faculty	Gender	Course name	Course code	Course Type

Proposed Algorithm

This algorithm is utilized to examine the relationship between data from the Academic Student Data (ASD) of Sebha University Cloud Server (SUCS). The algorithm presented in [2] is employed to discover association rules between attributes. The proposed algorithm takes as input the attributes representing the ASD. The output of the algorithm consists of the most frequent patterns that meet the minimum support and minimum confidence thresholds predefined by the user. The Apriori algorithm is designated as Algorithm 1 and is outlined below. The process for discovering frequent patterns is summarized in the following steps:

1. *Read ASD data sets (student records).*
2. *Scan the ASD to get the support for each frequent patterns.*
3. *Generate set of frequent patterns.*
4. *Generate a set of candidate k-frequent patterns.*
5. *Use Apriori property to prune the unfrequented k- patterns from this set.*
6. *Scan the ASD to get the conf with each candidate k- frequent patterns in the final set.*
7. *Compare Sup with and get a set of frequent k-frequent patterns.*
8. *Generate all nonempty sub events.*
9. *Output the rule where each rule satisfies the min-conf.*

Input :	***processed ASD data with class (students records) Min.con Min.sup***
Output: Association rules	
1:	***Begin***
2:	***For i=1,..,n***
3:	***L_i=Count item frequency***
4:	***End***
5:	***For (k=2;L_{K-1} ≠[13];k++} do begin***
6:	***C_k=apriori-gen(L_{k-1}); new candidates***
7:	***∀transactions t ϵ D do begin***
8:	***C_t=subset(C_k,t); candidates in transaction***
9:	***c.count++;determine support***
10:	***end***
11:	***L_k={c ϵ C_k | c.count ≥ min.sup} create new set***
12:	***end***

Algorithm 1: Apriori algorithm

Experiment AND Results

This section describes the experimental results from the Apriori algorithm and processed ASD. The Apriori algorithm is a machine-learning algorithm. Table 3 describe the number of attributes that involved in the experiment.

Table 3: Description of student academic attributes

No	Attributes	Abbreviation	Example
1	Depertment	D	Biotechnology
2	faculty	F	Science
3	Gender	G	Male-female
4	Social status	SS	Single-married
5	Course Type	CT	Major-supporting-cultural
6	Course state	CS	Compulsory-selective
7	Semster	S	Fall - Spring
8	Stu_smester_No	SN	1^{st}-2end…..eight
9	Course name	CN	Eng-arab-math …..chem
10	Course side	CS	Theoretical-practical
11	Grade	C	Fail-pass-good-v.good-exellent

TABLE 3 shows an example of generated rules (fixed at 300) for 37135 students records . The confidence is also fixed (at 0.7) and the support is changed nine times from 0.1 as min- sup to 0.9 as max.sup. The above table shows that for all the attributes (student attributes)

almost all appeared in the rules; the best rules are chosen where confidence equals 0.6, support equals 0.7 and the number of rules equals 300.

Table 4: N example of an experiment with 300 rules

Conf	Min.sup	Number of rules	Generated rules
0.1	0.1	300	All
0.2	0.2	300	All
0.3	0.3	300	All
0.4	0.4	300	All
0.5	0.5	300	All
0.6	0.6	300	All
0.7	0.7	300	All

The experiment is performed on 10 departments of faculty of science. The Apriori algorithm is run with a fixed number of rules (300) and with confidence (0.1) and support (0.7). The rules are divided by using the course grade as the class, where each grade is specified as the class.

i) Frequent patterns level 1

Fig.2: shows an example of frequent patterns in level 1 that generated when two attributes had relationship and satisfied the min-confidence and min-support.

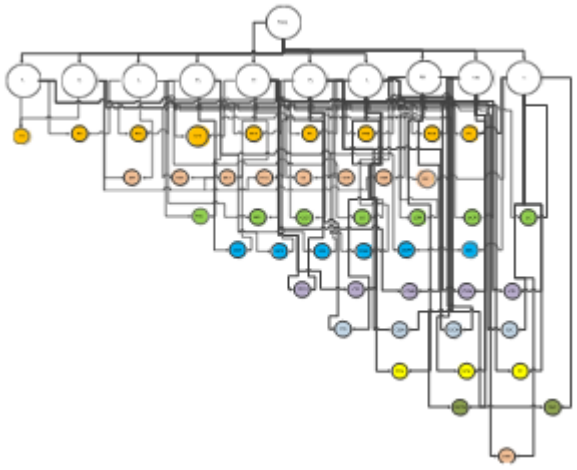

Fig.2 : Frequent patterns level 1

Then , apriori reducing the Number of candidates and generates best reules for output.

j) Frequent patterns level 2

Fig.3 shows an example of frequent patterns in level 2 that generated when three attributes had relationship and satisfied the min-confidence and min-support.

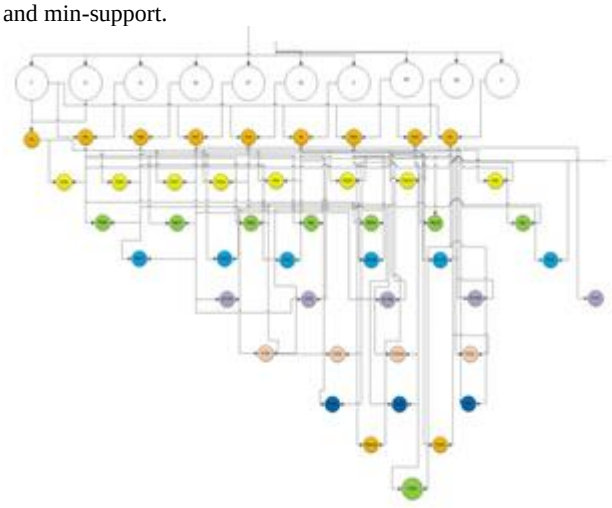

Fig.3: Frequent patterns level 2

Normal patterns are patterns that more frequently occur and are usually known by experts in a particular field. Appendix 1 and 2 explain the output of the Apriori algorithm.

Appendixes show a sample of the patterns that were extracted for different departments. Each pattern consists of five facts. one pattern represented the class F that related to students who could not pass their courses in different departments and semester No, meanwhile other patterns indicate facts related to students who pass their courses in different departments. Those patterns could help new students to take right courses in optima time and semester.

Conclusion

The pattern discovery approach, based on data mining techniques, specifically the Apriori algorithm, was employed to extract interesting patterns from student academic data. The novel attempts have shown promising discoveries in this domain. In this study, new and intriguing patterns of student data are uncovered through frequent pattern mining techniques. A series of patterns is detected, explaining the overall student performance pattern over the four years. Some patterns can be used as rules for prediction, while others can aid in detecting abnormal or surprising patterns. These patterns can be utilized to assist students and Sebha University in building prediction models.

References

[1]- Malini, H. and L. Suresh. Data Mining in Higher Education System and the Quality of Faculty Affecting Students Academic Performance: A Systematic Review. 2018.

[2]- Park, S. and Y.B. Park. Analysis of Association Between Students' Mathematics Test Results Using Association Rule Mining. in 2018 International Conference on Platform Technology and Service (PlatCon). 2018.

[3]- Rimpy, A. Dhankhar, and K. Solanki. Educational Data Mining tools and Techniques used for Prediction of Student's Performance: A Study. in 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO). 2022.

[4]- Kujur, A.G.P. and R.K. Tiwari. Study and design of decision-making models for education system using artificial intelligence. in 2021 4th International Conference on Recent Trends in Computer Science and Technology (ICRTCST). 2022.

[5]- Wook, M., et al. Predicting NDUM Student's Academic Performance Using Data Mining Techniques. in 2009 Second International Conference on Computer and Electrical Engineering. 2009.

[6]- Pal, S., Framework for Predicting Students' Performance–Review of Techniques. Indian Journal of Computer Science, 2017. 2: p. 28.

[7]- Alshareef, A., et al. Mining survey data on university students to determine trends in the selection of majors. in 2015 Science and Information Conference (SAI). 2015.

[8]- Job, M.A. and J. Pandey. Academic Performance Analysis Framework for Higher Education by Applying Data Mining Techniques. in 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO). 2020.

[9]- Mueen, A., B. Zafar, and U. Manzoor, Modeling and Predicting Students' Academic Performance Using Data Mining Techniques. International Journal of Modern Education and Computer Science, 2016. 11: p. 36-42.

[10]- Hashim, A., W. Akeel, and A. Khalaf, Student Performance Prediction Model based on Supervised Machine Learning Algorithms. IOP Conference Series: Materials Science and Engineering, 2020. 928: p. 032019.

[11]- Poonsirivong, K. and C. Jittawiriaynukoon. Big data analytics using association rules in eLearning. in 2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA). 2018.

[12]- Pianprasit, P., P. Seesai, and S. Rimcharoen. Association rule mining for analyzing placement test of computer science students. in 2017 2nd International Conference on Information Technology (INCIT). 2017.

[13]- Alkayed, M., et al. Factors Analysis Affecting Academic Achievement of Undergraduate Student: A study on Faculty of Information Technology Students at Applied Science Private University. in 2023 International Conference on Business Analytics for Technology and Security (ICBATS). 2023.

[14]- Burman, I., et al. Mining to Discover Association of Psychological Factors with Student Academic Performance. in 2019 4th International Conference on Information Systems and Computer Networks (ISCON). 2019.

Appendix

Example of Discovered frequent patterns with min_conf=0.7 in class F

Dept	Gender	Course name	Student semester No	Score
Botany	female	Fundamental of Organic Chemistry Lecture I	Eight	failed
Zoology	female	introduction to computer science	One	failed
Botany	male	General Mathematics I	One	failed
Zoology	female	General Chemistry I	Two	failed
Botany	female	General Chemistry I	One	failed
Zoology	male	General physics	One	failed
Biotech nology	female	General Zoology	Three	failed
Botany	female	Arabic A	One	failed
geology	male	General Mathematics I	Two	failed
comput er	female	Calculus IV	Twelve	failed
Botany	female	History of libya	Four	failed
chemist ry	female	Botany	One	failed
Botany	male	Arabic A	One	failed

Example of Discovered frequent patterns with min_conf=0.7 in class P

Dept	Gender	Course name	Student semester No	Score
Botany	female	Arabic A	One	pass
Zoology	female	Physical Geology I	One	pass
Biotechnolo gy	female	Arabic B	Two	pass
Biotechnolo gy	female	Organic Chemist Practical	Three	pass

Example of Discovered frequent patterns with min_conf=0.8 in class P

Dept	Gender	Course name	Student semester No	Score
Botany	Male	Arabic A	One	pass
Zoology	Male	Physical Geology I	One	pass
Biotechnolo gy	Male	Arabic B	Two	pass
Biotechnolo gy	Male	Organic Chemist Practical	Three	pass
computer	Male	Arabic B	Two	pass
geology	Male	Physical Geology I	Two	Pass