مجلّة جامعة عدن للعلوم الانسانيّة والاجتماعية
EJUA-HS Vol. 4 No. 4 (2023)
https://doi.org/10.47372/ejua-hs.2023.4.316
ISSN: 2708-6275

RESEARCH ARTICLE

THE RELATIONSHIP BETWEEN PREVIOUS QUALIFICATION CHARACTERISTICS AND UNIVERSITY GRADUATES USING ASSOCIATION RULES

Roiss Mohammed Salem Alhutaish[1,*], Anees Abdullah Shafal Ali[2], Khalid Kaied Shafal Ali[3]

[1]*Dept. of Information Technology, Aden Community Aden, Aden- Yemen.*
[2]*Dept. of computer Science, Faculty of Computer & Information Technology, Abyan University, Abyan- Yemen*
[3]*Dept. of computer, Faculty of Education, Aden University, Aden- Yemen*

***Corresponding author Roiss Mohammed Salem Alhutaish; E-mail: roiss2000@hotmail.com**

Received: 16 September 2023 / Accepted 02 November 2023 / Published online: 31 December 2023

Abstract

This study investigates the correlation factors between previous qualification and graduates of private universities using association rules mining. The experiments were conducted on a real dataset comprising graduates from three private universities. FP-Growth was utilized to create association rules in the experiments. The performance assessment criteria were a minimum support of 0.10 and a minimum confidence of 0.95. The findings suggest a strong relationship between high-achieving pharmacy graduates and whose previous education was a diploma in pharmacy.

Keywords: Association rules mining, Minimum support, Minimum confidence.

Introduction

Data mining is the process of discovering hidden, potentially useful, valid, and comprehensible knowledge from large databases (Borgelt, 2005). Data mining involves various techniques that transform raw data into valuable information, including clustering, statistical learning, association rule mining, and classification. (Mandave, Mane, & Patil, 2013). The fundamental goal of data mining is to recognize patterns, provide comprehensive information, and make accurate predictions.(Agrawal, Imieliński, & Swami, 1993). This is achieved through the process of analyzing raw data and summarizing it into a cohesive and useful format of information. Association rule Information mining is regarded as one of the most vital techniques in data mining. The primary objective of association mining is to discover interesting relationships and correlations among various items (Kotsiantis & Kanellopoulos, 2006).

The rest of the paper is organized as follows: Section 2 provides an overview of the related works. Section 3 describes the methodology, including a formal definition of evaluation metrics. Section 4 presents the experiments and discusses the results, while Section 5 offers a summary of the conclusions.

Related work

This section discusses some related work concerning association rules. Agrawal et al. proposed association rules by discovering correlations between products in large-scale transaction data recorded by point-of-sale (POS) systems in supermarkets.(Agrawal, Imielinski, & Swami, 1993). Abhishek and Anita(Jang, Yang, Park, & Kim, 2021) proposed a comparative study between three algorithms in association rules mining. The comparison was based on four criteria: accuracy, execution time, database scan, and memory usage. Eclat and FP-growth outperformed the Apriori algorithm in terms of execution time and memory usage for generating candidate item sets. YanRong (Liu, Wang, Miao, & Ren, 2022) proposed a method that includes certain

constraint conditions. They utilize a clustering matrix and a pruning strategy that combines the association rule data mining algorithm. This approach reduces the generation of candidate item sets and alleviates the I/O time consumption issue resulting from multiple database scans, thereby improving the algorithm's execution efficiency. The dataset of the Coronavirus pandemic has encouraged researchers to investigate factors influencing Coronavirus 2019 vaccinations through association rule mining. Wipawan and Pita proposed an analysis of the factors influencing vaccinations (Buathong & Jarupunphol, 2023). They used a questionnaire comprising 26 questions, which was conducted online among 403 participants residing in Phuket, Thailand. Among these participants, 377 people were vaccinated. Apriori and FP-Growth were employed, with performance assessment criteria of a minimum support of 0.5 and 0.6, and a minimum confidence of 0.9 and 1.0. The experimental results indicate that females with knowledge about vaccines are concerned about severe illness or death, as well as experiencing pain, nausea, vomiting, numbness, and hemiplegia.

Methodology

Before commencing the data mining process, several stages need to be completed. The first stage is data preparation, followed by the second stage, which involves data transformation. The third stage focuses on implementing Association Rules Mining, where FP-Growth and the creation of association rules are applied. The final stage involves evaluation. In the preparation phase, select the relevant datasets for use. This article utilizes a real dataset collected from the High Education Private Education Administration. This stage involves the cleaning of data, aiming to correct or remove unnecessary values, and the removal of any tuple lacking essential information. Additionally, we construct and integrate the data.

In data mining, datasets are categorized into two types. The first type is Quantitative, which comprises three aspects: Numeric, Discrete, and Continuous. The second type is Qualitative, which also has three aspects: Nominal, Ordinal, and Binary. During this stage, we convert the dataset's nominal attributes into binary attributes. Association rule mining is regarded as one of the most significant techniques in data mining. The primary objective of utilizing association rule mining is to uncover hidden relationships and correlations among various items within the database. This is achieved by identifying frequently occurring item sets in the database.(Karthikeyan & Ravikumar, 2014). There are measurements used to evaluate association rules, with the two most renowned types being support (s) and confidence (c). The rules must exhibit support and confidence values not less than the specified minimum. (Buathong & Jarupunphol, 2023).

Let $I = \{i_1, i_2, ..., i_n\}$ be as a collection of all different items in the database and D be a set of transactions, where each transaction T is a set of items so that T⊆I, the total number of transactions it contains is N.Therefore, an association rule is an implication of the form, $X \rightarrow Y$

Support is a Frequency of X and Y occurring together divided by total transactions:

$$Support = \frac{How\ \ many\ \ the\ \ items\ \ are\ \ bough\ \ togther}{Numer\ \ of\ \ \ Transaction} \quad (1)$$

Confidence is ratio of Frequency of X and Y occurring to the Frequency of X.

$$Confidence = \frac{How\ \ many\ \ times\ \ the\ \ items\ \ x, y\ \ are\ \ bought\ \ together}{How\ \ many\ \ times\ \ an\ \ itemx\ \ is\ \ brought} \quad (2)$$

The support and confidence measures alone are not sufficient for filtering out uninteresting association rules (Chen, 2019).. To address this limitation, a correlation measure known as lift can be employed to enhance the support-confidence framework for association rules. Lift is a straightforward correlation measure defined as follows:

$$Lift = \frac{\left(Confidence\ of\ \left(X \rightarrow Y\right)\right)}{\left(Frequency\ of\ \left(Y\right)\right)} \quad (3)$$

This study utilizes the FP-Growth Algorithm to discover frequent item sets without the necessity of candidate generation. The FP-Growth algorithm generates frequent item sets through only two scans of the database. Suppose there is a transaction database as shown in the table below. The first scan of the database is conducted to obtain a set of frequent items and their support count. Based on Table 1, the first scan produces the support count presented in Table 2. Infrequent items are then discarded using a support threshold of 50%. The items are arranged in descending order of support (I2:5. I1:4, I3:4, I4:4). In the second scan, the FP-Tree is constructed by traversing through the entire transaction database. During this step, the algorithm initializes the root node of the tree with the tag 'null.' Commencing from the root, transactions are progressively added one by one to the subtree using a prefix tree approach. When reading each transaction, the items are reordered based on their frequency, as illustrated in Table 3. For instance, in the first transaction, I2 is linked as a child to the root, followed by I1 linked to I2, and I3 linked to I1. Similarly, in the second transaction, I2 is linked to the root, I3 is linked to I2, and I4 is linked to I3. However, this branch shares the I2 node in common, as it is already used in T1. Therefore, the count of I2 is incremented by 1, and I3 is linked as a child to I2, while I4 is linked as a child to I3. The count is [10], {I3:1}, {I4:1}. After completing the second scan on all transactions, the FP-tree is generated as shown in Figure 1.

Table 1: Transaction database.

TID	Items
T1	I1,I2,I3
T2	I2,I3,I4
T3	I4,I5
T4	I1,I2,I4
T5	I1,I2,I3,I5
T6	I1,I2,I3,I4

Table 2: Count of each item.

Item	Count
I1	4
I2	5
I3	4
I4	4
I5	2

Table 3: Ordered and truncated Transactional database.

TID	Items
T1	I2,I1,I3
T2	I2,I3,I4
T3	I4
T4	I2,I1,I4
T5	I2,I1,I3
T6	I2,I1,I3,I4

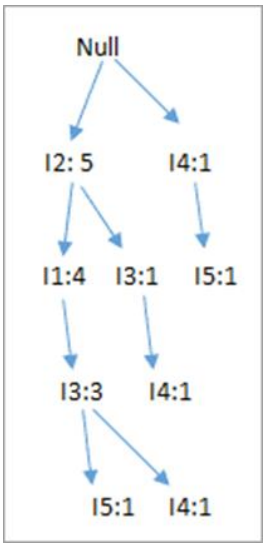

Figure 1: Generating FP-tree.

Experiments and Results

The experiments were conducted using a dataset collected from three private universities. The dataset includes the following attributes: gender, place of birth, faculty, specialty, cumulative grade point average (CGPA), year of graduation, previous certificate, place of the previous certificate, year of the previous certificate, and grade point average (GPA) of the previous certificate. These attributes were obtained after cleaning and integration. The first university contributed 771 rows, the second university contributed 514 rows, and the third university contributed 1105 rows. All experiments were conducted using a minimum support of 0.1 and a minimum confidence of 0.95. According to **Table 4**, the results obtained using the dataset from the first university are as follows: there is a strong correlation between the 2015 graduates and the Faculty of Medical Sciences, particularly the Bachelor of Pharmacy program. This confidence is derived from students enrolled in the pharmacy diploma program. Similarly, the 2020 graduates exhibit the same confidence, with the majority of the enrolled students coming from high schools. All graduates achieving a very good score in the Bachelor of Pharmacy program have a full confidence coefficient (Confidence=1.0).

N0	Premises	Conclusion	Support	Confidence	Lift
1	specialty = 9, Graduate year = 15	previous certificate = 2	0.15	0.95	2.39
2	Graduate year = 20	previous certificate = 1	0.26	0.95	1.67
3	previous certificate = 2	Faculty = 5	0.38	0.95	1.17
4	GPA of previous certificate = 2, previous certificate = 2	Faculty = 5	0.14	0.96	1.17
5	previous certificate = 2, CGPA = 2	Faculty = 5	0.14	0.96	1.17
6	CGPA = 3, previous certificate = 2	Faculty = 5	0.17	0.96	1.17
7	GPA of previous certificate = 2, Graduate year = 20	previous certificate = 1	0.11	0.99	1.69
8	previous certificate = 1, specialty = 7	Faculty = 5	0.14	1.0	1.23
9	CGPA = 3, specialty = 9	Faculty = 5	0.22	1.0	1.23
10	GPA of previous certificate = 2, specialty = 9	Faculty = 5	0.18	1.0	1.23
11	specialty = 9, CGPA = 2	Faculty = 5	0.15	1.0	1.23
12	specialty = 9, GPA of previous certificate = 3	Faculty = 5	0.16	1.0	1.23
13	CGPA = 3, specialty = 9, previous certificate = 2	Faculty = 5	0.15	1.0	1.23
14	CGPA = 3, specialty = 9, GPA of previous certificate = 3	Faculty = 5	0.11	1.0	1.23
15	GPA of previous certificate = 2, specialty = 9, previous certificate = 2	Faculty = 5	0.10	1.0	1.23
16	specialty = 9, Graduate year = 15, previous certificate = 2	Faculty = 5	0.15	1.0	1.23
17	specialty = 9, previous certificate = 2, CGPA = 2	Faculty = 5	0.10	1.0	1.23
18	specialty = 9, previous certificate = 2, GPA of previous certificate = 3	Faculty = 5	0.12	1.0	1.23

Table 5, illustrates the results obtained from utilizing the second dataset. It is evident that students graduating with an overall assessment of 'good' and a Bachelor of Pharmacy degree share a close connection (Confidence=1.0). Additionally, there is a strong correlation between students who obtained 'good' and 'very good' grades in the diploma degree pharmacy and those enrolled in the bachelor's degree program in pharmacy.

N0	Premises	Conclusion	Support	Confidence	Lift
1	Graduate year = 20	previous certificate = 1	0.11	0.96	1.64
2	Faculty = 5, specialty = 7	previous certificate = 1	0.12	0.98	1.67
3	CGPA = 3, specialty = 9	Faculty = 5	0.15	1.0	1.67
4	GPA of previous certificate = 3, specialty = 9	Faculty = 5	0.13	1.0	1.67
5	GPA of previous certificate = 2, specialty = 9	Faculty = 5	0.13	1.0	1.67
6	specialty = 9, CGPA = 2	Faculty = 5	0.11	1.0	1.67
7	specialty = 9, previous certificate = 2	Faculty = 5	0.12	1.0	1.67

Table 6, presents the results obtained from the use of the second dataset. A strong correlation is observed with high confidence levels among students who achieved 'good' and 'very good' grades, majoring in jurisprudence and fundamentalism and information technology, respectively.

N0	Premises	Conclusion	Support	Confidence	Lift
1	GPA of previous certificate = 2, specialty = 1	Faculty = 1	0.13	1.0	2.75
2	GPA of previous certificate = 2, specialty = 3	Faculty = 3	0.14	1.0	4.36
3	CGPA = 3, specialty = 2	Faculty = 2	0.11	1.0	2.68
4	specialty = 1, CGPA = 2	Faculty = 1	0.12	1.0	2.75

Conclusion

This study employs the FP-Growth Algorithm to reveal intriguing patterns and correlations among various attributes of graduate students. These attributes encompass information about the graduate students themselves as well as details about their prior qualification from bachelor's programs. The experiments were conducted using datasets from three private universities. The results indicate the presence of correlations; however, no relationship was identified between a graduate's cumulative grade point average (CGPA) and their high school GPA.

References

[1] Agrawal, R., Imielinski, T., & Swami, A. (1993). Database mining: A performance perspective. *IEEE transactions on knowledge and data engineering*, 5(6), 914-925.

[2] Agrawal, R., Imieliński, T., & Swami, A. (1993). *Mining association rules between sets of items in large databases*. Paper presented at the Proceedings of the 1993 ACM SIGMOD international conference on Management of data.

[3] Borgelt, C. (2005). *An Implementation of the FP-growth Algorithm*. Paper presented at the Proceedings of the 1st international workshop on open source data mining: frequent pattern mining implementations.

[4] Buathong, W., & Jarupunphol, P. (2023). ARM for Analyzing Factors Influencing Vaccinations During the COVID-19 Outbreak. *Journal of Data Science and Intelligent Systems.*

[5] Chen, Y. (2019). The Application of FP-Growth Algorithm on Data Mining of Association Rules From Law Lists. *Available at SSRN 3474162*.

[6] Jang, H.-J., Yang, Y., Park, J. S., & Kim, B. (2021). FP-growth algorithm for discovering region-based association rule in the IoT environment. *Electronics*, 10(24), 3091.

[7] Karthikeyan, T., & Ravikumar, N. (2014). A survey on association rule mining. *International Journal of Advanced Research in Computer and Communication Engineering*, 3(1), 2278-1021.

[8] Kotsiantis, S., & Kanellopoulos, D. (2006). Association rules mining: A recent overview. *GESTS International Transactions on Computer Science and Engineering*, 32(1), 71-82.

[9] Liu, Y., Wang, L., Miao, R., & Ren, H. (2022). A Data Mining Algorithm for Association Rules with Chronic Disease Constraints. *Computational Intelligence and Neuroscience*, 2022, 8526256. doi:10.1155/2022/8526256

[10] Mandave, P., Mane, M., & Patil, S. (2013). Data mining using Association rule based on APRIORI algorithm and improved approach with illustration. *International Journal of Latest Trends in Engineering and Technology (IJLTET)*, 3(2), 107-113.

مقالة بحثية

العلاقة بين خصائص المؤهل السابق وخريجي الجامعة باستخدام قواعد الارتباط

رويس محمد سالم الحتيش[1،*] ، انيس عبدالله شعفل علي[2] و خالد قائد شعفل علي[3]

[1] قسم تكنولوجيا المعلومات، كلية المجتمع عدن، عدن، الجمهورية اليمنية
[2] قسم علوم الحاسوب، كلية الحاسوب وتقنية المعلومات، جامعة ابين، ابين، الجمهورية اليمنية
[3] قسم الحاسوب، كلية التربية، جامعة عدن، عدن، الجمهورية اليمنية

* الباحث الممثل: رويس محمد سالم الحتيش؛ البريد الالكتروني: roiss2000@hotmail.com

استلم في: 16 سبتمبر 2023 / قبل في: 02 نوفمبر 2023 / نشر في: 31 ديسمبر 2023

المُلخّص

تبحث هذه الدراسة في عوامل الارتباط بين المؤهلات السابقة وخريجي الجامعات الخاصة باستخدام تعدين قواعد الارتباط. أجريت التجارب على مجموعة بيانات حقيقية تضم خريجين من ثلاث جامعات خاصة. تم استخدام FP-Growth لإنشاء قواعد الارتباط في التجارب. وكانت معايير تقييم الأداء هو الحد الأدنى من الدعم والذي كانت قيمته 0.10 والحد الأدنى من الثقة وكانت قيمته 0.95. تشير النتائج إلى وجود علاقة قوية بين خريجي الصيدلة المتفوقين والذين كان تعليمهم السابق عبارة عن دبلوم في الصيدلة.

الكلمات المفتاحية: تعدين قواعد الارتباط، الحد الأدنى من الدعم، الحد الأدنى من الثقة.

How to cite this article:

R. M. S. Alhutaish, A. A. S. Ali, K. K. S. Ali. (2023). STUDENTS' PERCEPTIONS OF FACTORS AFFECTING THE MA APPLIED LINGUISTICS PROGRAM AT ADEN FACULTY OF EDUCATION. *Electronic Journal of University of Aden for Humanity and Social Sciences*, 4(4), p673-678. https://doi.org/10.47372/ejua-hs.2023.4.316