Engineering and Technology Journal e-ISSN: 2456-3358

Volume 09 Issue 07 July-2024, Page No.- 4520-4525

DOI: 10.47191/etj/v9i07.22, I.F. – 8.227

© 2024, ETJ

Implementation of Data Mining to Find Association Patterns of Tracer Study Data Using Apriori Algorithm

Alwis Nazir[1], Iwan Iskandar[2], Teddie Darmizal[3]

[1,2,3] Faculty of Science and Technology, Sultan Syarif Kasim Riau State Islamic University, Pekanbaru, Indonesia

ABSTRACT: Knowing the distribution of alumni from a university is very useful as an evaluation material and a benchmark for teaching and learning activities at related universities. One way to get the distribution of alumni is to conduct a tracer study. In this research, a tracer study will be conducted and then the data will be processed with data mining techniques using the apriori algorithm. The data results in the form of relationship patterns between attributes will make it easier for decision makers in higher education to gain new knowledge about graduates and can be used to improve and ensure the quality of higher education. The tracer study conducted focuses on Informatics Engineering students of UIN Suska Riau who graduated in 2019 and 2020. The results of this study obtained new knowledge such as the first job of graduates with a waiting period of less than 6 months is as a contract or honorary employee with a salary between 3-5 million and has an ipk between 3-3.5.

KEYWORDS: Data Mining, Association Patterns, Tracking Study, Apriori

I. INTRODUCTION

Tracer study is an important activity for a university to track graduates who have been produced in terms of measuring the relationship between educational goals and processes with the current condition of graduates [1],[2] . This is evidenced by the existence of a graduate tracing component (tracer study) in the Higher Education Institution Accreditation (AIPT) assessment [3]. The tracer study can provide a variety of information that is useful for the purposes of evaluating the success of learning that has been implemented in higher education and can then be used to improve and ensure the quality of the university concerned. Tracer studies are also useful in providing important information about the alignment of the relationship between the curriculum and the needs and expectations of the world of work, assessing the relevance of higher education, information for stakeholders in a university [4].

Sultan Syarif Kasim State Islamic University Riau (UIN Suska Riau) is one of the universities located in Riau Province and has produced many graduates since its establishment in 1970 (then IAIN and changed to UIN in 2005). Based on interviews conducted with the Vice Rector 3 of UIN Suska Riau, Mr. Edi Erwan, Ph.D, at UIN Suska Riau itself, a tracer study has been conducted, but the implementation has not been optimal and has not been carried out comprehensively. The tracer study that has been carried out is by each study program (prodi) and is carried out only to meet the accreditation needs of the study program concerned, there is no clear follow-up and output from the results of the tracer study. Therefore, this research will conduct a tracer study at UIN Suska Riau by focusing on the Informatics Engineering study program. The results of the tracer study will then be processed using data mining techniques to get an association pattern or relationship that has meaning.

Data mining is a series of processes using statistical, mathematical, artificial intelligence, and database research techniques to extract valuable information that is not known manually from the database. The resulting information is obtained by extracting and recognizing important or interesting patterns [5]. One of the data mining analysis techniques is association. Association analysis or association rule mining is a data mining technique for finding rules for a combination of items. The algorithm of association that can be used to analyze this tracer study data is apriori.

By using the apriori algorithm, the large amount of tracer study data is then extracted into information in the form of relationship patterns between attributes, this can make it easier for decision makers in higher education to gain deeper insight into alumni who have graduated, which in turn can be important information and can be used to improve and ensure the quality of the higher education institution.

The advantage of this apriori algorithm is that it is simpler and can explore association rules that often appear in data that has a large collection of items (itemsets) quickly. While other association algorithms have weaknesses in memory usage when the amount of data owned is large so that it affects the number of items that can be processed [6]. In addition, the apriori algorithm has also been successfully applied to find association rules in various types of data, such as: (1) sales transactions [7][8][9]; (2) smoking-related diseases [10]; and (3) library book borrowing [11]. Based on the description above, the author raises the title for this research, namely

"Implementation of Data Mining to Find Association Patterns of Tracer Study Data Using the Apriori Algorithm".

II. LITERATURE REVIEW

A. *Data Mining*

Data mining or data mining is an analysis step of a large set of data to get the relationship between the data and summarize it into a form that is easy to understand and can be useful. The main purpose of this data mining is to find new knowledge hidden from the database by extracting and identifying information using statistical, mathematical, artificial intelligence, and machine learning techniques [12], [13].

Data mining has several methods that can be used to explore and discover new knowledge, namely [14]:

1) *Description:* to provide a concise description of a number of large-scale and diverse data.
2) *Classification:* finding models or functions that can distinguish concepts or data classes, with the aim of estimating the class of an object whose label is unknown.
3) *Estimation:* a method of guessing an unknown value with some known related information.
4) *Prediction:* used in estimating a value in the future, for example predicting product sales.
5) *Clustering:* used for grouping in identifying data with certain characteristics. Association,used to find sets of items that appear frequently among large data sets.

B. *Apriori Algorithm*

The a priori algorithm is included in the type of association rules in data mining, namely rules that state the relationship between several attributes or what is often called affinity analysis. The a priori analysis method is divided into 2 stages, namely:

1) *Analysis of high frequency patterns (support):* this stage is done by looking for combinations of items that meet the minimum support value requirements in the database. The formula for calculating the support value of an item is [15]:

$$Support\ (A) = \frac{The\ number\ of\ transactions\ contains\ A}{Total\ Transactions}$$

The support value of 2 items is obtained using the following formula [15]:

$$Support(A,B) = P(A \cap B)$$

$$Support(A,B) = \frac{\sum Transaction\ contains\ A\ and\ B}{\sum Transaction}$$

2) *Establishment of associative rule:* After all the high-frequency patterns are found, then look for association rules that meet the minimum requirements for confidence. The formula for calculating the confidence value of the two items is as follows [15]:

$$Confidence\ P(A|B) = \frac{\sum Transaction\ contains\ (A)\ and\ (B)}{\sum Transaction(A)}$$

III. RESEARCH METHODS

The research method contains a structured framework from the initial stages of research to produce the desired achievement. The stages of this research can be seen in Figure 1.

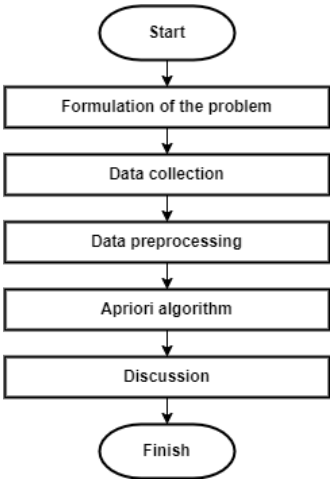


Figure 1: Stages of Research Method

A. *Problem Formulation*

Problem formulation is done by identifying problems that exist in the UIN Suska Riau environment.

B. *Data Collection*

Data collection, data collection in this study was carried out in two stages, namely literature study of several books, journals, and articles. Then, data collection is carried out using a questionnaire of alumni who have graduated in 2019 and 2020.

C. *Data Preprocessing*

Data preprocessing consists of three stages: data selection, data cleaning, and data transformation.

1) *Data selection:* at this stage the selection of attributes that will be used in this study is carried out. The attributes used are class, ipk, year of graduation, gender, type of first job, first job status, suitability of job relationship with major, waiting period to get first job.
2) *Data cleaning:* the stage of cleaning up data that has missing values, inconsistent data, and data discrepancies (outliers).
3) *Data transformation:* data that has gone through the next stage of the data cleaning process is transformed and then stored into a form that can be applied to the algorithm that will be used later.

D. Application of Algorithms

At this stage the tracer study data that has passed the data preprocessing stage will be applied to the apriori algorithm using the google colab application and the mlxtend library.

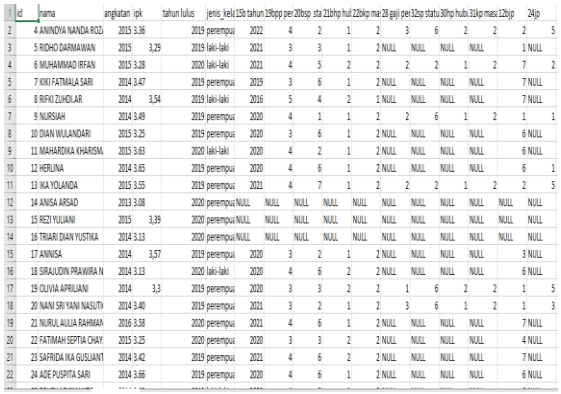
E. Data Interpretation

Figures must be numbered using Arabic numerals. Figure captions must be in 8 pt Regular font. Captions of a single line must be centered whereas multi-line captions must be justified. Captions with figure numbers must be placed after their associated figures.

IV. RESULT

A. Result of Data Collection

The data used in this study is data obtained from the results of the tracer study of Uin Suska Riau Informatics Engineering students (https://alumni.tif.uin-suska.ac.id/) who have graduated in 2019 and 2020 with a total of 161 data. Can be seen Figure 2.



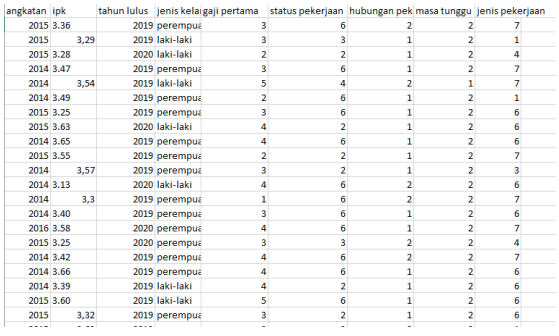
	id	nama	angkatan	ipk	tahun lulus	jenis_kela	15bln	tahun	19bpp	per	20bsp	sta	21bhp	hul	22bkp	ma	28 gaji	per	32cp	statu	30hp	hubi	31kp	mas	129jp	24jp	
2	4	ANINDYA NANDA ROZI	2015	3.36	2019	perempua	2022	4	2	1	2	3	6	2	2	2	5										
3	5	RIDHO DARMAWAN	2015	3,29	2019	laki-laki	2021	3	3	1	2	NULL	NULL	NULL	NULL	1	NULL										
4	6	MUHAMMAD IRFAN	2015	3.28	2020	laki-laki	2021	4	5	2	2	2	2	1	2	7	2										
5	7	KIKI FATIMALA SARI	2014	3.47	2019	perempua	2019	3	6	1	2	NULL	NULL	NULL	NULL	7	NULL										
6	8	RIFKI ZUHDILAR	2014	3,54	2019	laki-laki	2016	5	4	2	1	NULL	NULL	NULL	NULL	7	NULL										
7	9	NURSIAH	2014	3.49	2019	perempua	2020	4	1	1	2	2	6	1	2	1	1										
8	10	DIAN WULANDARI	2015	3.25	2019	perempua	2020	3	6	1	2	NULL	NULL	NULL	NULL	6	NULL										
9	11	MAHARDIKA KHARISM	2015	3.63	2020	laki-laki	2020	4	2	1	2	NULL	NULL	NULL	NULL	6	NULL										
10	12	HERLINA	2014	3.65	2019	perempua	2020	4	6	1	2	NULL	NULL	NULL	NULL	6	1										
11	13	IKA YOLANDA	2015	3.55	2019	perempua	2021	4	7	1	2	2	2	1	2	2	5										
12	14	ANISA ARSAD	2013	3.08	2020	perempua	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL										
13	15	REZI YULIANI	2015	3,39	2020	perempua	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL										
14	16	TRIARI DIAN YUSTIKA	2014	3.13	2020	perempua	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL										
15	17	ANNISA	2014	3,57	2019	perempua	2020	3	2	1	2	NULL	NULL	NULL	NULL	3	NULL										
16	18	SIRAJUDIN PRAWIRA N	2014	3.13	2020	laki-laki	2020	4	6	2	2	NULL	NULL	NULL	NULL	6	NULL										
17	19	OLIVIA APRILIANI	2014	3,3	2019	perempua	2020	3	3	2	2	1	6	2	2	1	5										
18	20	NANI SRI YANI NASUTI	2014	3.40	2019	perempua	2021	3	2	1	2	3	6	1	2	1	3										
19	21	NURUL AULIA RAHMAN	2016	3.58	2020	perempua	2021	4	6	1	2	NULL	NULL	NULL	NULL	7	NULL										
20	22	FATIMAH SEPTIA CHAY	2015	3.25	2020	perempua	2020	3	3	2	2	NULL	NULL	NULL	NULL	4	NULL										
21	23	SAFRIDA IKA GUSLIANT	2014	3.42	2019	perempua	2021	4	6	2	2	NULL	NULL	NULL	NULL	7	NULL										
22	24	ADE PUSPITA SARI	2014	3.66	2019	perempua	2020	4	6	1	2	NULL	NULL	NULL	NULL	6	NULL										

Figure 2: Sample Tracer Study Data

A. Results of Data Preprocessing

The tracer study data that has been collected is then preprocessed with three stages, namely attribute selection, data cleaning, and data transformation.

1) *Attribute Selection Results:* After selecting and adjusting to the research objectives, the data to be used is data on students who have worked and 9 attributes are selected. The 9 attributes are class, ipk, year of graduation, gender, type of first job, first job status, suitability of job relationship with major, and waiting period to get first job. After selecting students who have worked, the total data becomes 129 data.



angkatan	ipk	tahun lulus	jenis kelai	gaji pertama	status pekerjaan	hubungan pek	masa tunggu	jenis pekerjaan
2015	3.36	2019	perempua	3	6	2	2	7
2015	3,29	2019	laki-laki	3	3	1	2	1
2015	3.28	2020	laki-laki	2	2	1	2	4
2014	3.47	2019	perempua	3	6	1	2	7
2014	3,54	2019	laki-laki	5	4	2	1	7
2014	3.49	2019	perempua	2	6	1	2	1
2015	3.25	2019	perempua	3	6	1	2	6
2015	3.63	2020	laki-laki	4	2	1	2	6
2014	3.65	2019	perempua	4	6	1	2	6
2015	3.55	2019	perempua	2	2	1	2	7
2014	3,57	2019	perempua	3	2	1	2	3
2014	3.13	2020	laki-laki	4	6	2	2	6
2014	3,3	2019	perempua	1	6	2	2	7
2014	3.40	2019	perempua	3	6	1	2	6
2016	3.58	2020	perempua	4	6	1	2	7
2015	3.25	2020	perempua	3	3	2	2	4
2014	3.42	2019	perempua	4	6	2	2	7
2014	3.66	2019	perempua	4	6	1	2	6
2014	3.39	2019	laki-laki	4	2	1	2	6
2015	3.60	2019	laki-laki	5	6	1	2	6
2015	3,32	2019	perempua	3	2	1	2	6

Figure 3: Attribute Selection Result

In Figure 3 there are codes for the values of the attributes of first salary, employment status, first job relationship with major, first job waiting period, type of job. The explanation for these values can be seen in the Table IV-1.

Table IV-1: Explanation of Value for Each Attribute

Attributes	***Value***	***Description***
First Salary	1	Under Rp. 1.000.000
	2	Between Rp. 1.000.000, -- Rp. 1.500.000
	3	Between Rp. 1.500.000, -- Rp. 3.000.000
	4	Between Rp. 3.000.000, -- Rp. 5.000.000
	5	Above Rp. 5.000.000
First Employment Status	1	Civil Servant
	2	Contract employee
	3	Honorer
	4	Director
	5	Manager
	6	Staff
	7	Intern
Relationship between First Job and Major	1	Appropriate/Related
	2	Not suitable at all
Waiting Period for First Job	1	Before graduation
	2	< 6 months after graduation
	3	≥ 6 months after graduation
Type of Employment	1	Government Agency
	2	Multinational SOEs
	3	National SOE
	4	Non-profit Organization/Non-governmental Organization
	5	Multinational Private Company
	6	National Private Company
	7	Self-employed / Own company

B. Data Cleaning

Data cleaning is carried out on data that contains missing values, inconsistent data, and data discrepancies (outliers). Can be seen Figure 4.

1). Missing Values: after checking in Ms. Excel, no blank data was found.

2). Inconsistent Data: this stage is done to check for inconsistent data. It is known that the ipk value uses a comma sign (,) instead of a period (.). After checking there are 101 data that use a comma (,), then the ipk data is replaced with a period (.).

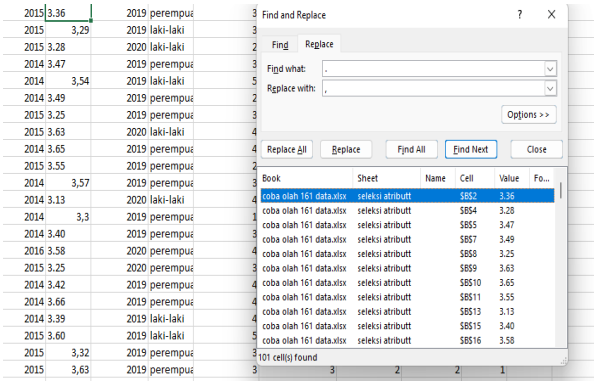


Figure 4: Inconsistent Data Check

3). Outliers: this stage is carried out to check data whose values are too far from other data. An example is ipk whose value cannot be greater than 4.0. After checking, there is no outlier data.

C. Data Transformation Results

Data transformation is done by changing the measurement scale into another form so that the data can meet the assumptions of the appropriate analysis for processing. There are 4 attributes that will be transformed, namely ipk, first salary, first job status, and first job type. Transformation of these 4 attributes is done to minimize the choice options so that the support value can later increase. Can be seen Table IV-2 and Table IV-3.

Table IV-2: *Grade Point Average Attributes*

Numerical Data	***Nominal Data***
< 3	IPK1
3 – 3.5	IPK2
> 3.5	IPK3

Table IV-3: Transformation of attributes of first salary, first employment status, and employment type

Attributes	***Initial Category***		***New Category***	
First Salary	Value	Description	Value	Description
	1	Below Rp. 1,000,000	1	Below Rp. 1,500,000
	2	Between Rp. 1,000,000, - Rp. 1,500,000	2	Between Rp. 1,500,000, - Rp. 3,000,000
.	3	Between Rp. 1,500,000 and Rp. 3,000,000	3	Between Rp. 3,000,000, - Rp. 5,000,000
	4	Between Rp. 3,000,000, - Rp. 5,000,000	4	Above Rp. 5,000,000
	5	Above Rp. 5,000,000		
First Job Status	Value	Information	Value	Information
	1	CIVIL SERVANTS	1	PNS
	2	Contract employee	2	Contractor Honorary employee
	3	Honorer	3	Director or manager
	4	Director	4	Staff or apprentice
	5	Manager		
	6	Staff		
	7	Intern		
Job Type	Value	Description	Value	Description
	1	Government Agency	1	Government Agency
	2	Multinational SOEs	2	National SOE or Multinational SOE
	3	National SOEs	3	National Private Company / Multinational Private Company
	4	Non-profit organization/ non-governmental organization	4	Self-employed or non-governmental organization
	5	Multinational Private Company		

	6	National Private Company		
	7	Self-employed/own company		

The next stage is that the data is transformed into one hot encoding, this is done so that the form of data is in accordance with what is needed by the apriori algorithm. The following data has been transformed and is ready to be applied to the apriori algorithm. Can be seen Figure 5.

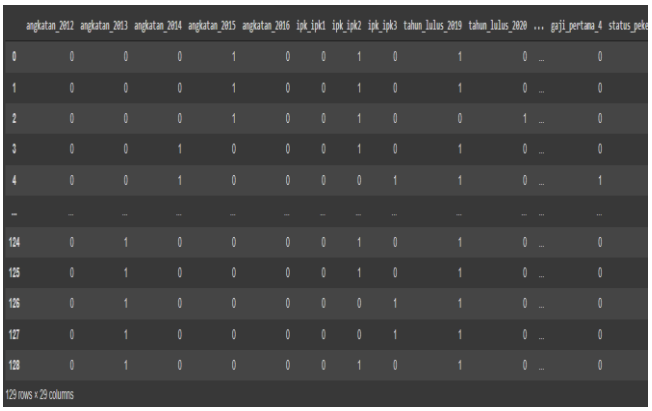


Figure 5: Data Transformation

D. Apriori Algorithm Application Results

Analysis using data mining is intended to find certain forms of rules that are still hidden, which are expected to be used as policy making by interested universities. The forms of rules found in this study are not all used, only some that are deemed relevant. Based on the results of the application of the apriori algorithm using google colab and the mlxtend library with a minimum support of 10% and a minimum confidence of 70%, the rules formed are as follows Table IV-4.

Table IV-4: Apriori Algorithm Results

Number	***If***	***Then***	***Support***	***Confidence***	***Lift***
1	salary_first_3, employment_status_2, ipk_ipk2	time_wait_2	13.17%	94%	1.25%
2	status_employment_2, gender_male, relationship_first_employment_with_degree_1, ipk_ipk2	time_waiting_2	17.05%	84.61%	1.12%
3	class_2014, year_graduate_2019, gender_female	time_wait_2, relationship_employment_1	12.40%	76.19%	1.32%
4	ipk_ipk3, type_gender_female	time_wait_2, relationship_employment_1	10.85%	73.68%	1.28%

E. Pattern Interpretation

Based on the rules generated from the application of the apriori algorithm with a minimum support of 10% and a minimum confidence of 70%, new knowledge is obtained, namely:

1). If a graduate has an ipk between 3-3.5 and the first job status is a contract or honorary employee and has a first salary between 3-5 million then the time to get the first job is less than 6 months. This rule has support 10.85%, confidence 100%, and lift ratio 1.32%.
2). If the male with ipk 3-3.5 and the employment status of contract or honorary employees and work in accordance with the major then the waiting period for the first job is less than 6 months. This rule has support 17.05%, confidence 84.61%, and lift ratio 1.12%.
3). If a female student of class 2014 and graduated in 2019 or graduated within 5 years then has a waiting period for the first job less than 6 months and has a job in accordance with the major. This rule has support 12.40%, confidence 76.19%, and lift ratio 1.32%.
4). For female graduates who have an ipk above 3.5, the waiting period for the first job is less than 6 months and has a job relationship that matches the major. This rule has support 10.85%, confidence 73.68%, and lift ratio 1.28%.

Thus, some information can be concluded, namely:

1). The first job of graduates with a waiting period of less than 6 months is as contract or honorary employees with salaries between 3-5 million and have ipk between 3-3.5.
2). On average, graduates get their first job quickly, which is less than 6 months after graduation with an ipk between 3-3.5.
3). For female students who want to get a job quickly and in accordance with their majors, the ipk must be above 3.5.

V. CONCLUSION

Based on the research that has been done, it can be concluded that the implementation of data mining using the apriori algorithm on tracer study data can be done to find patterns of relationships between attributes. With 10% support and 70% confidence, 4 rules are obtained which are considered relevant. The new knowledge obtained is that the first job of graduates with a waiting period of less than 6 months is as a contract or honorary employee with a salary between 3-5 million and has an ipk between 3-3.5.

REFERENCES

1. N. R. Dewi, P. Listiaji, M. Taufiq, E. N. Savitri, A. Yanitama, and A. P. Herianti, "Development of a tracer study system for graduates of the Integrated Science Department, Universitas Negeri Semarang," in Journal of Physics: Conference Series, IOP Publishing Ltd, Jun. 2021. doi: 10.1088/1742-6596/1918/4/042010.
2. V. Berton and A. Rosadi Kardian, "Penerapan Algoritma K-Means Untuk Analisis Tracer Alumni Universitas Gunadarma Jurusan Sistem Informasi dan Sistem Komputer Angkatan 2013," Jurnal Ilmiah Komputasi, vol. 18, no. 3, Oct. 2019, doi: 10.32409/jikstik.18.3.2639.
3. Majelis Akreditasi, "Ban-PT," 2019.
4. Y. Nugraheni, "Analisis Tracer Study Lulusan Politeknik Dengan Exit Cohort Sebagai Pendekatan Target Responden," Seminar Nasional Sistem Informasi Indonesia, 2018.
5. D. Merawati, "Penerapan Data Mining Penentu Minat Dan Bakat Siswa Smk Dengan Metode C4.5," JURNAL ALGOR, vol. 1, no. 1, 2019, [Online]. Available: https://jurnal.buddhidharma.ac.id/index.php/algor/index
6. M. Fauzy, K. W. Rahmat Saleh, I. Asror, J. Telekomunikasi No, and T. Buah Batu Bandung, "Penerapan Metode Association Rule Menggunakan Algoritma Apriori Pada Simulasi Prediksi Hujan Wilayah Kota Bandung," 2016.
7. M. Sholik and A. Salam, "Implementasi Algoritma Apriori untuk Mencari Asosiasi Barang yang Dijual di E-commerce OrderMas," Techno.COM, vol. 17, no. 2, pp. 158–170, 2018.
8. I. Djamaludin and A. Nursikuwagus, "Analisis Pola Pembelian Konsumen Pada Transaksi Penjualan Menggunakan Algoritma Apriori," Jurnal SIMETRIS, vol. 8, 2017.
9. J. Lasmana Putra, M. Raharjo, T. Alfian Armawan Sandi, and R. Prasetyo, "Implementasi Algoritma Apriori Terhadap Data Penjualan Pada Perusahaan Retail," Jurnal PILAR Nusa Mandiri, vol. 15, no. 1, p. 85, 2019, [Online]. Available: https://www.kaggle.com.
10. F. Tinus Waruwu, E. Buulolo, and E. Ndruru, "Implementasi Algoritma Apriori Pada Analisa Pola Data Penyakit Manusia Yang Disebabkan Oleh Rokok," KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer) , vol. 1, 2017, [Online]. Available: http://ejurnal.stmik-budidarma.ac.id/index.php/komik
11. E. Srikanti, R. Fitri Yansi, I. Permana, and F. Nur Salisah, "Penerapan Algoritma Apriori Untuk Mencari Aturan Asosiasi Pada Data Peminjaman Buku Di Perpustakaan," Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi, vol. 4, no. 1, pp. 77–80, 2018.
12. A. Asroni, B. Masajeng Respati, and S. Riyadi, "Penerapan Algoritma C4.5 untuk Klasifikasi Jenis Pekerjaan Alumni di Universitas Muhammadiyah Yogyakarta," Semesta Teknika, vol. 21, no. 2, 2018, doi: 10.18196/st.212222.
13. B. Septia Pranata and D. Putro Utomo, "Bulletin of Information Technology (BIT) Penerapan Data Mining Algoritma FP-Growth Untuk Persediaan Sparepart Pada Bengkel Motor (Study Kasus Bengkel Sinar Service)," Bulletin of Information Technology (BIT), vol. 1, no. 2, pp. 83–91, 2020.
14. S. Sibagariang, A. Riyadi, A. Dzikri, F. Suandi, K. T. Sirait, and F. Setiawan, "Prediksi Prospek Kerja Alumni Dengan Algoritma Neural Network," CESS (Journal of Computer Engineering System and Science), vol. 6, no. 1, pp. 2502–714, 2021.
15. G. Cakra Sutradana,) M Didik, and R. Wahyudi, "Penerapan Data Mining Untuk Analisis Pengaruh Lama Studi Mahasiswa Teknik Informatika Uin Sunan Kalijaga Yogyakarta Menggunakan Metode Apriori," 2017.