Hasan - cek.04

by Firman Noor Hasan, Achmad Sufyan Aziz, Yos Nofendri

Submission date: 29-Dec-2022 06:20PM (UTC+0700)
Submission ID: 1987273803
File name: Hasan,_Aziz_-_cek.04.docx (417.76K)
Word count: 3646
Character count: 18762

Utilization of Data Mining on MSMEs Using FP-Growth Algorithm for Menu Recommendations

Firman Noor Hasan[1], Achmad Sufyan Aziz[2], Yos Nofendri[3]
[1,2,3]Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

Article Info

Article history:

Received mm dd, yyyy
Revised mm dd, yyyy
Accepted mm dd, yyyy

Keywords:

Data Mining
FP-Growth
Menu Recommendation
MSME

ABSTRACT

Existing transaction data is only recorded and stored as a sales transaction memorandum, so it has not been utilized optimally. The data is only stored and used as transaction history. The availability of a lot of data and having a pattern of sales transactions that are similar to MSME Cafe Over Limit will be utilized by using data mining science. This research uses the association rules method. Implementation of fp-growth to get item combinations. The goal is to make it easier for MSMEs to determine menu recommendations for customers. The fp-growth algorithm is used to process as many as 2038 transaction data with a minimum support value of 10%, while for a minimum confidence value of 50%. So that there are 3 rules, namely "if you order Mariam chocolate cheese milk then the customer will order Kopsus Overlimit", from this rule it will form a support value of 10.79%, using a confidence value of 54.19% and a lift ratio of 0.93. Furthermore "if you order Kopsus Overlimit then you will order tofu at grandma's house", from the rule it will produce a support value of 34.69%, with a specified confidence value of 59.76%, so the lift ratio value is 1.15. The last rule "if you order tofu at grandma's house, the customer orders Kopsus Overlimit", from the rule that occurs, the support value is 34.69%, with a confidence value of 66.7% and a lift ratio of 1.15. The results of the study found the two best rules, namely "if the customer orders over-limit Kopsus, he will order tofu at grandma's house" and "if he orders tofu at grandma's house, the customer orders over-limit Kopsus". Based on the results of the rules formed, it can be concluded that only two rules can be categorized as valid and can be used as a reference in food and beverage menu recommendations at MSME Cafe Over Limit. So the results of this study can be useful to be applied to MSMEs, especially in terms of menu recommendations.

Corresponding Author:

Firman Noor Hasan, 081311406664,
Faculty of Industrial Technology and Informatics,
Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia.
Email: firman.noorhasan@uhamka.ac.id .

1 INTRODUCTION

Current technological developments have a major impact on human life. One of the factors is how humans use data [1]. Data becomes a very important factor in all things, such as how to increase sales results, consumer spending patterns, and their desires, and being able to see current market trends [2]. Along with data storage in the form of softcopy and cloud, it makes data easier to do computerized processing [3]. The data stored on the server or cloud is increasingly piling up and increasing in size and becoming big data [4]. Current technological developments make big data technology quite significant in providing results because it has been integrated with social media [5].

The Cafe is identical to a micro restaurant that sells a variety of snacks and drinks served with the concept of an interesting place to relax and spend time just chatting about personal and non-personal matters. In general, cafes only provide various kinds of coffee and non-coffee drinks to food dishes according to the concept of the cafe theme itself which is served dine-in or eat on the spot, however several cafes provide takeaway and delivery services [6]. Over Limit Café is a fairly popular cafe in the area. At first, Over Limit Café was a small cafe with few employees, however Over Limit Café already has approximately 4 employees. Cafe Over Limit has problems in determining menu recommendations. namely, the optimal presentation concept has not been created in determining the menu recommendations to be given to customers. For this reason, it is necessary to determine the right menu recommendations, so that customers are expected not to waste too much time when ordering the available menu items.

The transaction that occurs at Over Limit Café, is that the customer comes to the cashier and the cashier records what the customer wants to buy. The recorded data will be stored as a sales transaction memorandum and is not used optimally so that the existing data is only stored or used as transaction history. The availability of quite a lot of data at Over Limit Cafe, of course, has a sales transaction pattern that resembles each transaction that can be utilized and used as a consideration through the understanding of data mining science. Data mining is useful in exploratory analysis scenarios where there is no predetermined idea of what an interesting outcome concept would be. Data mining is needed in the new quest to determine the concept of valuable results and non-trivial information in the volume of datasets to draw conclusions on data that has been formed which is achieved in a balance from human knowledge to visualizing pictures of problems and specific goals assisted by computer search capabilities [7].

This study uses association rules as a data mining method that will be applied in obtaining results. Association rules in the data mining method are used to determine the directed association of each item in the dataset to characterize the correlation or relationship between various items and other items. In short, the suitability of the characterization that is formed on each item from the dataset will be combined through association rules [8]. In determining the results in the suitability of the characterization formed for each item, the dataset will be processed using one of several association rules algorithms, namely fp-growth. The purpose of this study is to use the fp-growth method to process the sales transaction data so that it can make it easier for the cafe to determine menu item recommendations to customers [9].

The latest among other studies that are of relevance to previous research is that this study provides a solution using the association rules algorithm, namely fp-growth, crisp-DM, and association rules to recommend menus to customers by utilizing sales transaction update data at Cafe Over UMKM. Limits, which can later be applied to other types of MSMEs in making menu recommendations based on historical transaction data. Previous research stated that implementing a web-based system that uses the crisp-DM method to build a system that can recommend products [10]. Researchers have also made web-based applications with data processing using the fp-growth algorithm, resulting in an analysis of consumer buying patterns in motorcycle spare parts sales transactions [11]. The application of search association rules to sales transaction data using fp-growth produces a more significant level of accuracy than the a priori algorithm [12]. Furthermore, research conducted by Setyo to determine products that are often sold results that the fp-growth algorithm generating retail item data on CV.Cahaya Setya [13]. In contrast to the research conducted by Gunadi, it results that the fp-growth algorithm has lower or less significant results than the a priori algorithm on sales transactions [14]. The latest among other studies that are of relevance to previous research is that this research provides a solution using the association rules algorithm, namely fp-growth, crisp-dm, and association rules to recommend menus to customers by utilizing sales transaction update data at Over Limit Cafe.

2 RESEARCH METHOD

This study used a combined method of fp-growth, association rule, data mining, and crips-dm. Figure 1 below shows the flow of the stages of the research carried out. Frequent pattern growth, better known as the fp-growth algorithm, is an a priori

algorithm that was developed by definition as an approach to finding and determining the dominant data series in the itemset (frequent itemset) domain. The fp-growth algorithm does not require candidate generation activities, because it puts forward an approach from the concept of building a tree (fp-tree) at the stage of obtaining dominant data in the itemset domain. As a result, this algorithm becomes more efficient in obtaining dominant data in the itemset domain (frequent itemset) compared to the a priori algorithm [16]. Association rules are part of data mining techniques with the main procedure for finding patterns of relationships that are formed between an item contained in the dataset, in that context the items contained in the dataset must have a relationship between one item and another item, if the item is related to the item otherwise, the association rule is useful for discovering the rules governing how or why these items are often purchased together [17]. Data mining is useful in exploratory analysis scenarios where there is no predetermined idea of what an interesting outcome concept would be. Data mining is needed in the new quest to determine the concept of valuable results and non-trivial information in the volume of datasets to draw conclusions on data that has been formed which is achieved in a balance from human knowledge to visualizing pictures of problems and specific goals assisted by computer search capabilities [18]. CRISP-DM is an acronym that comes from the term Cross-Industry Standard Process for Data Mining, which is a model that provides an overview of the life cycle of a data mining project consisting of 6 stages [19].

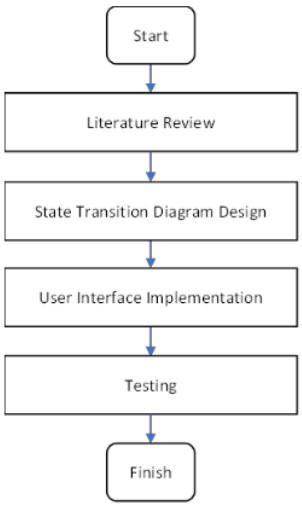

Figure 1. Research Flowchart

2.1. Literature Review

At this stage, the researcher conducts a literature review of journals conducted by previous researchers to get the latest from this research.

2.2. State Transition Diagram

State Transition Diagram with the Harel model is used in designing applications in this study. State Transition Diagrams are made as a design in illustrating ongoing processing on a system with states that will be connected between the states that are formed [15].

2.3. Implementation

At this stage, the focus is on the description of the results of the design implementation and significant testing of the application. At the design implementation stage using the Perl Hypertext Processor programming language.

2.4. Testing

Obtaining the test results will describe in detail the results of the data testing that has been completed and processed by the system. The test results are the results of data calculations using the fp-growth algorithm or method to form a pattern of sales transactions at the MSME Over Limit Café and the search for recommendations for food and drink menus at the Over Limit Cafe.

3 RESULTS AND ANALYSIS

The stages in application design begin with the design of a state transition diagram, then proceed to the database design stage with the final output arrangement in the form of an introduction to the user interface or application interface display.

3.1. State Transition Diagram Design

The STD design of the FP-Growth page describes the calculation of a dataset used with the FP-Growth algorithm approach. Figure 2 shows where the FP-Growth page will input a calculation sheet in the context of selecting an existing transaction date and input the obtained result value from minimum support and the obtained result value from minimum confidence.

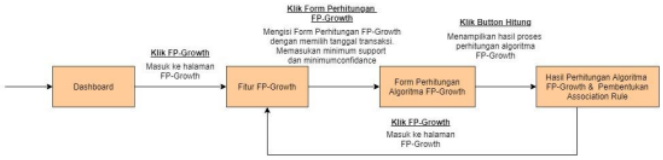

Figure 2. State Transition Diagram Design for the FP-Growth page

The State Transition Diagram of the Calculation Results page has a mechanism as a means of forming association rules with the FP-Growth algorithm approach. Assuming that each rule is formed can function as a recommended solution for food and beverage menus, as shown in Figure 3 below.

Figure 3. The design of the State Transition Diagram for the Calculation Results page

The State Transition Diagram of the Recommendation Results page has an output in the form of a recommendation for a food and beverage menu along with an explanation resulting from the output, as shown in the following figure.4.

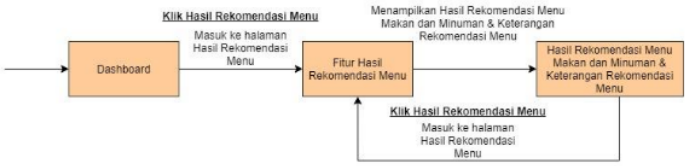

Figure 4. Design of the State Transition Diagram for the Menu Recommendation Results page

3.2. User Interface Implementation

The Add Data feature page contains a form that functions to store certain additional data for Over Limit Cave. In this form, you are required to input Transaction IDs, goods, or items as well as the time and date specified in the available columns. Figure 5 shows the web view for the added data page.

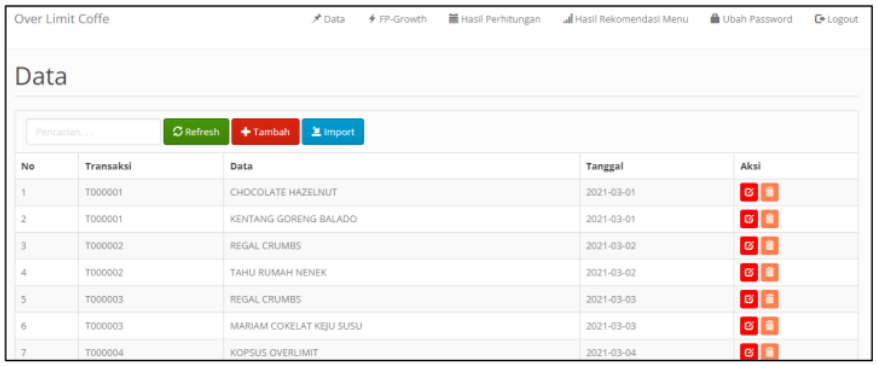

Figure 5. Web view for the Add Data page

The data import feature web page has a choose file button as a means of using the application to input data that will be processed with the fp-growth algorithm. It should be noted in the data input process if the excel file format is required to be in .csv format and have a writing format in the form of id_data, id_transaksi, item, and date fields. If the excel file is not in .csv format, then indirectly the file will be rejected by the system that will import the data into the database, so it will not appear on the application page. As shown in Figure 6 below.

Figure 6. Web view for the Import Data page

Figure 7 below shows a page that will present the configuration and results obtained from the association rule in the form of a structured table.

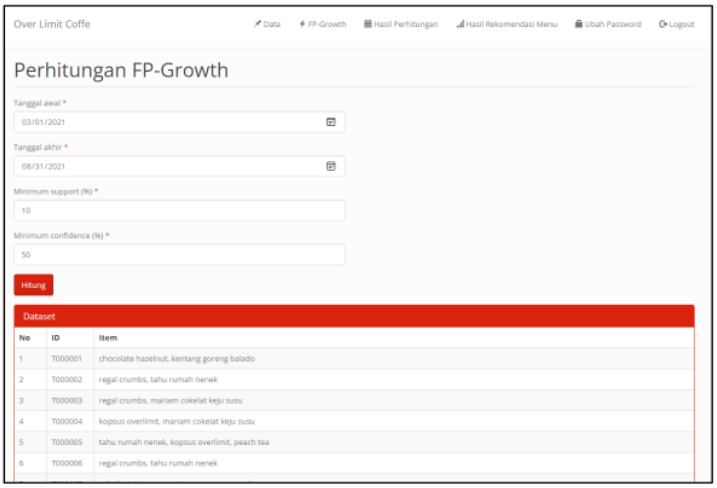

Figure 7. Web view for the Calculation Results page

The menu recommendation results page will present recommendations for food and beverage menus and descriptions of recommendations from Cafe Over Limit in graphical form, which is formed using calculations from the fp-growth algorithm. Figure 8 below shows the web view for the menu recommendation results page.

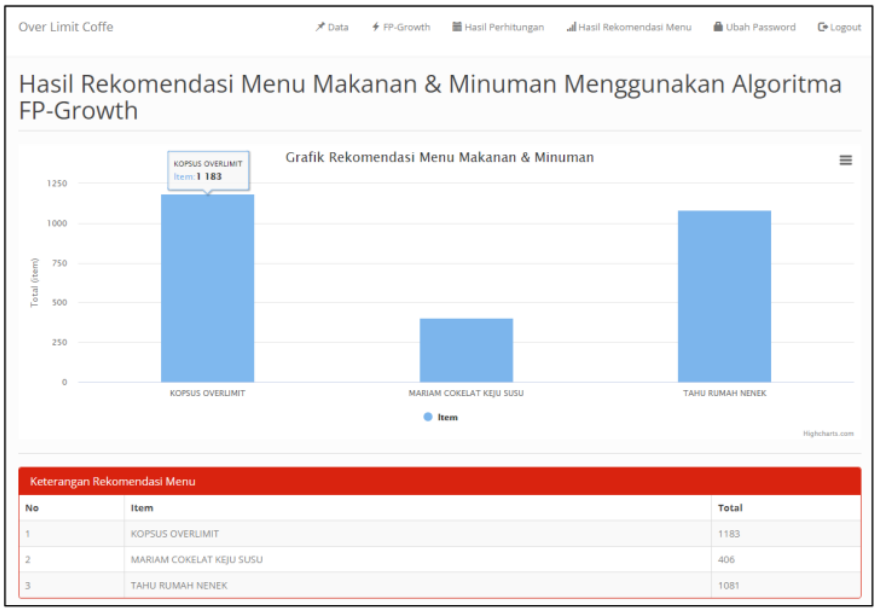

Figure 8. Web display for the Menu Recommendations Results page

3.3. Testing

The results of this test have the main focus on knowing the accuracy of the results on each data transaction formed or generated from the association rule. The lift ratio is a reference assumption in obtaining the validity of the transaction patterns that take place in this study. The lift ratio has a decision-making assumption, that is, if the lift ratio is less than one (<1), then it can be assumed that the resulting item is categorized as negatively correlated with item B so that the item does not have a significant relationship with other items. Meanwhile, if the lift ratio results are greater or more than number one (> 1), then it can be assumed that the relationship produced by item A with item B is categorized as positively correlated. The decision assumption if the resulting value is equal to one (= 1), has the assumption that item A and item B are categorized as an independent.

Table 1. Rule results obtained using the FP-Growth method

No	Rule	Support	Confidence	Lift Ratio
1	If customer orders milk chocolate cheese mariam, then the customer will order Kopsus Overlimit	220/2038 = 10.79%	220/406 = 54.19%	0.93
2	If customer order Kopsus Overlimit, then ths customer will order Tahu Rumah Nenek	707/2038 = 34.69%	707/1183 = 59.76%	1.15
3	If customer order Tahu Rumah Nenek, then the customer will order Kopsus Overlimit	707/2038 = 34.69%	707/1060 = 66.7%	1.15

Table.1 shows the results of several rules, including "if customer orders mariam chocolate cheese milk, then the customer will order Kopsus Overlimit" with a support value of 10.79%, a confidence value of 54.19%, and a lift ratio value of 0.93, it can be categorized as invalid rules or negatively correlated because the lift ratio value is <1. While the next rule "If customers order Kopsus Overlimit, then the customer will order Tahu Rumah Nenek" with a support value of 34.69%, a confidence value of 59.76%, and a lift ratio value of 1.15, it can be categorized as a valid rule or positively correlated because the lift ratio value is > 1. Furthermore, the last rule "If customers order Tahu Rumah Nenek, then the customer will order Kopsus Overlimit" with a support value of 34.69%, a confidence value of 66.7%, and a lift ratio value of 1.15, can be categorized as a valid rule or positively correlated because the lift ratio value is > 1. From the results of the rules that have been made, it can be seen that only two rules can be categorized as valid and can be used as a reference in recommendations for food and drink menus at MSME Over Limit Cafe.

4 CONCLUSION

Implementation using the fp-growth algorithm approach can obtain significant output in knowing the pattern of transactions resulting from purchase transactions on menus or items from food and beverages. This implementation is also able to produce an exact determination of menu recommendations or items of food and beverages based on the pattern of continuity of purchases or transactions from UMKM Cafe Over Limit. The results of the three purchase transaction patterns or rules at UMKM Cafe Overlimit produce two significant purchase transaction patterns or rules, namely: "If a customer orders Overlimit Kopsus then he will order tofu from grandma's house" and "If a customer will order or buy tofu house grandma, indirectly the customer will order Kopsus Overlimit". Based on the results of the rules formed, it can be concluded that only two rules can be categorized as valid and can be used as a reference in food and beverage menu recommendations at UMKM Cafe Over Limit. The decision to obtain from the two patterns of purchase transactions or significant rules is based on the assumption that the higher or larger a support value is used, it will have an impact on the confidence value and also the value of the lift ratio for the rule that is formed, the better the rule will be. Suggestions for further research can use other methods such as a priori or market basket analysis.

The display on the web page of the menu recommendations presents recommendations for food and beverage menus and a description of recommendations from Cafe Over Limit in graphical form, which is formed using calculations from the fp-growth algorithm. So that it can make it easier for MSME Café Over Limit managers to determine the recommended menu for customers. So the results of this study can be useful to be applied to MSMEs, especially in terms of menu recommendations.

5 ACKNOWLEDGEMENTS

This research was funded by the Research and Development Institute of Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia. Number 257/F.03.07/2022.

REFERENCES

[1] F. N. Hasan and A. Febriandirza, "Perancangan Data Warehouse Untuk Data Penelitian di Perguruan Tinggi Menggunakan Pendekatan Nine Steps Methodologhy," *Pseudocode*, vol. VIII, no. 1, pp. 49–57, 2021.

[2] V. Plotnikova, M. Dumas, and F. Milani, "Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements," *Data Knowl. Eng.*, 2022.

[3] M. Shawkat, M. Badawi, S. El-ghamrawy, R. Arnous, and A. El-desoky, "An optimized FP-growth algorithm for discovery of association rules," *J. Supercomput.*, vol. 78, pp. 5479–5506, 2022.

[4] F. N. Hasan, "Implementasi Sistem Business Intelligence Untuk Data Penelitian di Perguruan Tinggi," in *Prosiding Seminar Nasional TEKNOKA 4*, 2019, vol. 4, no. 2502, pp. I1–I10.

[5] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, "Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 76–89, 2022.

[6] J. Ferreira, "Café nation? Exploring the growth of the UK café industry," *AREA Ethics In/Of Geoghraphical Res.*, vol. 49, no. 1, pp. 69–76, 2016.

[7] A. Bartschat, M. Reischl, and R. Mikut, "Data Mining Tools," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 9, no. 4, pp. 1–14, 2019.

[8] P. S. Nishtala, T. yuan Chyou, F. Held, D. G. Le Couteur, and D. Gnjidic, "Association rules method and big data: Evaluating frequent medication combinations associated with fractures in older adults," *Pharmacoepidemiol. Drug Saf.*, vol. 27, no. 10, pp. 1123–1130, 2018.

[9] J. R. Chang, Y. S. Chen, C. K. Lin, and M. F. Cheng, "Advanced Data Mining of SSD Quality Based on FP-Growth Data Analysis," *Appl. Sci.*, vol. 11, no. 4, pp. 1–15, 2021.

[10] S. Kurniawan, W. Gata, and H. Wiyana, "Analisis Algoritma FP-Growth Untuk Rekomendasi Produk Pada Data Retail Penjualan Produk Kosmetik (Studi Kasus: MT Shop Kelapa Gading).," *Semin. Nas. Teknol. Inf. Dan Komun. 2018 (SENTIKA 2018)*, vol. 8, pp. 61–69, 2018.

[11] A. Maulana and A. A. Fajrin, "Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Transaksi Penjualan Spare Part Motor," *KLIK Kumpul. J. Ilmu Komput.*, vol. 5, no. 1, 2018.

[12] A. Salam, J. Zeniarja, W. Wicaksono, and L. Kharisma, "Pencarian Pola Asosiasi Untuk Penataan Barang Dengan Menggunakan Perbandingan Algoritma Apriori Dan Fp-Growth (Study Kasus Distro Epo Store Pemalang)," *DInamik*, vol. 23, no. 2, pp. 57–65, 2018.

[13] W. N. Setyo and S. Wardhana, "Implementasi Data Mining Pada Penjualan Produk Di Cv Cahaya Setya Menggunakan Algoritma Fp-Growth," *PETIR J. Pengkaj. dan Penerapan Tek. Inform.*, vol. 12, no. 1, pp. 54–63, 2019.

[14] G. Gunadi and D. I. Sensuse, "Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori dan Frequent Pattern Growth(FP-Growth): Studi Kasus Percetakan PT.Gramedia," *Telemat. MKOM*, vol. 4, no. 1, pp. 118–132, 2022.

[15] D. Venkatesan, S. K. Kannan, M. Arif, M. Atif, and A. Ganeshan, "Sentimental Analysis of Industry 4.0 Perspectives Using a Graph-Based Bi-LSTM CNN Model," *Mob. Inf. Syst.*, vol. 2022, no. Special Issues, pp. 1–14, 2022.

[16] F. Sidik, I. Suhada, A. H. Anwar, and F. N. Hasan, "Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier," *J. Linguist. Komputasional*, vol. 5, no. 1, pp. 34–43, 2022.

[17] F. N. Hasan and M. Dwijayanti, "Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan

Multinominal Naïve Bayes Classifier," *J. Linguist. Komputasional*, vol. 4, no. 2, pp. 52–58, 2021.

[18] L. Shabtay, P. Fournier-viger, R. Yaari, and I. Dattner, "A Guided FP-Growth algorithm for mining multitude-targeted item-sets and class association rules," *Inf. Sci. (Ny).*, vol. 553, pp. 353–375, 2020.

[19] V. Plotnikova, M. Dumas, and F. Milani, "Adapting the CRISP-DM Data Mining Process: A Case Study in the Financial Services Domain," *RCIS (International Conf. Res. Challenges Inf. Sci.*, vol. 415, 2021.

Hasan - cek.04

ORIGINALITY REPORT

7%	6%	4%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universiti Teknikal Malaysia Melaka Student Paper	1%
2	Submitted to University of College Cork Student Paper	1%
3	Aditya Rizky Fadillah, Intan Nurma Yulita, Aditya Pradana, Mira Suryani. "Data Mining Implementation Using Frequent Pattern Growth on Transaction Data for Determining Cross-selling and Up-selling (Case Study: Cascara Coffee)", 2021 International Conference on Artificial Intelligence and Big Data Analytics, 2021 Publication	1%
4	ijain.org Internet Source	1%
5	ijcis.net Internet Source	1%
6	Gufron Amirullah, Irdalisa Irdalisa. "Mastery of STEM-Based Research Approach of Science	<1%

Teachers In Jakarta", AL-ISHLAH: Jurnal Pendidikan, 2022
Publication

7	dergipark.org.tr Internet Source	<1 %
8	ejournal.ikmi.ac.id Internet Source	<1 %
9	ijair.id Internet Source	<1 %
10	jurnal.darmajaya.ac.id Internet Source	<1 %
11	simakip.uhamka.ac.id Internet Source	<1 %
12	www.coursehero.com Internet Source	<1 %
13	www.hindawi.com Internet Source	<1 %
14	2021.jnic.es Internet Source	<1 %
15	Amneh Shaban, Fadi Almasalha, Mahmoud H. Qutqut. "Hybrid user action prediction system for automated home using association rules and ontology", IET Wireless Sensor Systems, 2019 Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		