
Analysis of Customer Transaction Data Associations Based on The Apriori Algorithm

Tri Astuti ^{a,*}, Bella Puspita ^b

^a Informatics Engineering Study Program, Faculty of Computer Science, Universitas Amikom Purwokerto

^b System Information Study Program, Faculty of Computer Science, Universitas Amikom Purwokerto

* corresponding author

Abstract

UD Dian Pertiwi is one of the small and medium enterprises engaged in materials with the main product is building materials. This business experiences large amounts of transactions every day, the data obtained becomes increasingly large, and it will only be limited to a pile of useless data or commonly called junk. By utilizing a data mining approach apriori algorithm technique, the data can be utilized to support the sales process and achieve a target of UD Dian Pertiwi. Based on research and data mining that has been done using association analysis and apriori algorithms by applying a minimum of support = 1% and a minimum of confidence = 70% resulted in the ten strongest association rules can be used by UD Dian Pertiwi in the process of applying a sales strategy including determining interrelationships, in short, the product has the potential to be purchased at the same time, increasing the amount of product stock and conducting promotions.

Keywords: Data Mining; Association Rules; Apriori Algorithm; Minimal Support; Confidence.

1. Introduction

UD Dian Pertiwi is one of the small-medium enterprises engaged in the material with the main product that is building materials. This business every day is experiencing large amounts of transactions so that the data obtained longer becomes more and more, when not used then it will only be the limit of useless data stacks, or commonly called data Garbage, while the data can be used to achieve the target of UD Dian Pertiwi which until now still has not achieved, with the utilization of data mining approach, especially the technique of using apriori algorithm that has Excellence of mining data quickly, then it can be utilized the data to support the sales process to achieve a target of UD Dian Pertiwi, through the research is expected to be found a rule of association with the value of confidence, in order to support the sales process, so that the target that has been set can be achieved.

2. Literature Review

2.1. Data Mining

Data mining as a process for obtaining useful information from a large database warehouse[1]. Data mining can also be interpreted as extracting new information extracted from large chunks of data that aid in decision making[2]. The term data mining is sometimes called knowledge discovery [3][4].

Several data mining phases include:

- a. Cleaning and Integration
- b. Selection and Transformation
- c. Evaluation and Presentation

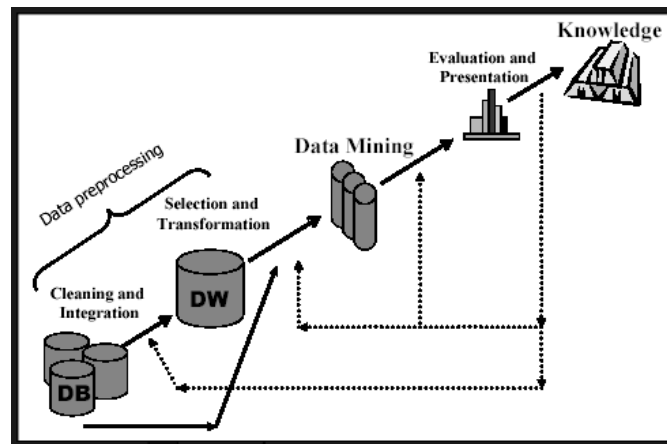


Fig 1. Data Mining Process

Description:

a. Cleaning and Integration

Data sanitization will also affect the formation of data mining systems because the data handled will be reduced in number and complexity.

Data integration is done on attributes that identify unique entities such as name attribute, product type, customer number, etc. Data integration needs to be done carefully because errors in data integration can produce deviant results and even misleading action later[5].

b. Selection and Transformation

Transformation and selection of this data also determine the quality of data mining results later because there are several characteristics of specific data mining techniques that depend on this stage[6].

c. Data Mining Process

The application of the data mining technique is only one part of the data mining process. Several data mining techniques are already commonly used[7]. We will go into more information on the techniques that are in the next section. Note that there are times when general data mining techniques available in the market are insufficient to carry out data mining in certain areas or for specific data.

d. Evaluation and Presentation

The last phase of the data mining process is how to formulate decisions or actions from the results of the analysis. Sometimes it should involve people who do not understand data mining. Therefore, the presentation of data mining results in the form of knowledge that everyone can understand is a necessary step in the process of data mining.

Data mining contains the desired pattern search in an extensive database to help decision making in the future. This pattern is recognizable by specific devices that can provide a useful and insightful data analysis that can then be learned more thoroughly, which may use other decision-making devices [8].

2.2. Association Rule

The Analysis of association or association rule mining is a data mining technique to find associative rules between a combination of items. The analysis of the association is also known as one of the data mining techniques that become the basis of one of the other data mining techniques. In particular, one stage of the association analysis attracted the attention of many researchers to produce an efficient algorithm, namely, the analysis of the high-frequency pattern (a frequent pattern mining) [8].

2.3. Apriori Algorithm

Apriori algorithm is one of the algorithms that can be used in the implementation of the basket analysis market to find the rules of association that meet the limits of support and confidence [9]. During the process stage, the algorithm generates a systematic excavation without exploring all the candidates, while in the second stage is carried out

extraction against a strict rule. Frequent itemset one usually refers to a collection of items that often appear together in transactional data[10][15].

2.4. Weka

WEKA is a Java API that provides APIS for processing in data mining that is open source-based (GPL) and JAVA-engine. WEKA is developed by the University of Waikato in New Zealand and is also free software available under the GNU (General Public License)[11][14]. WEKA provides the use of classification techniques using a decision tree with the J48 algorithm. The technique used by WEKA is a classifier.

3. Method

3.1. Problem Identification

In this research, the first thing to do is to identify the problem. Identification of the problem is a vital research process as it can determine the quality of research conducted and to formulate the problems that will be the background in the research object that is done, the problem identified is how to analyze sales data on UD Dian Pertiwi based on an Apriori algorithm.

3.2. Data Collection

After completing the problem identification, then proceed to the next stage of data collection. At this stage, researchers collect the data needed to find association rules based on a priori algorithm[12][13]. Data needed in the form of data from interviews directly conducted by asking questions to the admin of UD Dian Pertiwi to know the problems that exist in order to be a reference in the research, from now on by the way of the library study collecting data and studying the data of customer spending transactions in UD Dian Pertiwi within 5 months, from January to May 2019, as for the documentation level to complete the data retrieval

3.3. Application Of The Method Of Association Using Apriori Algorithm

By implementing a priori algorithm, the authors will find the association rules with a good confidence value to support the sales process in UD Dian Pertiwi by looking for a combination of association rules items. The first step is to find a combination of items that meet the minimum support requirement. The support value is a value that shows the combination of items in the database taken from the number of transactions containing A divided by the number of transactions. Then searchable association rules that meet the minimum requirements for confidence.

4. Results and Discussion

In this study, customer spending transaction data at UD Dian Pertiwi within 5 months from January to May 2019.

Table 1. Purchased Materials

Transaction ID	Materials Purchased
2nd January 2019	
T1	Tidal Sand – Cement
T2	Indoboard ("Gypsum") – Hollo – Cable – Kompond – Concrete nail – grinder – Ceramic pieces – Screw
T3	Tora meter
T4	Glass – Ryben Glass
T5	Glass – Glass mirror – Glass Ryben
T6	Ryben Glass – Glass
T7	Nails – Plywood – thin hinges – Grendel
...	...
...	...
T1301	Etherna Brush – Sandpaper Roll
T1302	Paralon Clamp

4.1. Pre-processing Data

In this research, authors perform Pre-processing data, which aims to transform raw data into the appropriate format for analysis, to make it easier then it is done coding on each item of goods.

Table 2. Pre-processed Data

Code	Item Name
1	aluminum foil
2	Sandpaper
3	Sandpaper Roll
4	Asbestos
5	Steel
6	Mild Steel
7	Paint Tub
8	Sink
...	...
...	...
237	wood eco
238	wood plank

After doing the coding process, the tabular table is made to make it easier to know how many items were purchased in a transaction. Tabulation data is the creation of tables by containing data that has been coded initial goods according to the required analysis. Transactions are given by giving a "Y" (yes), meaning that the transaction contains such goods and a "?" sign, which means that the transaction does not contain the goods.

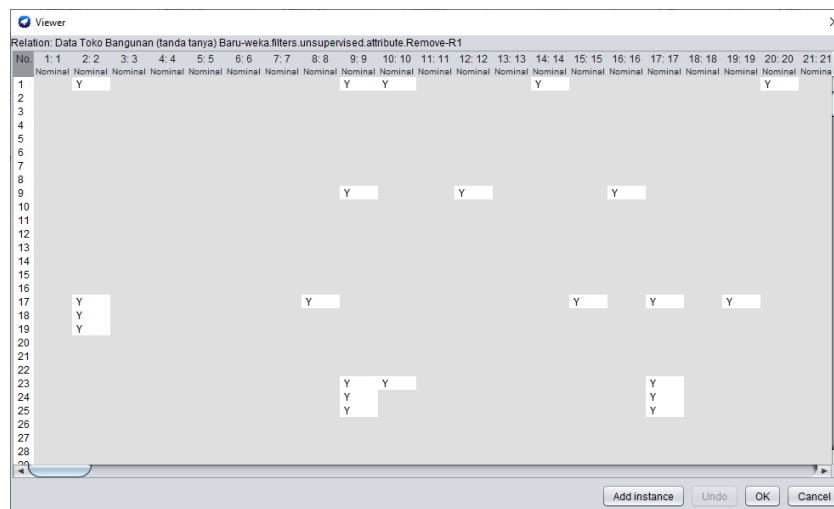


Fig 2. Weka Results

4.2. Apriori with Weka

At this stage, the process is done from a priori algorithm to acquire a combination pattern of goods in Weka. In this process is determined in advance a minimum of support and confidence, where researchers determine 1% for minimum support and 70% for minimum confidence. 1% stated that of all transactions of goods showing goods purchased simultaneously and 70% confidence stated that if a transaction occurs on 1 item, then there is a 70% possibility to buy 2nd item. With a minimum of 1% support and confidence 70% then the process on Weka becomes like this:

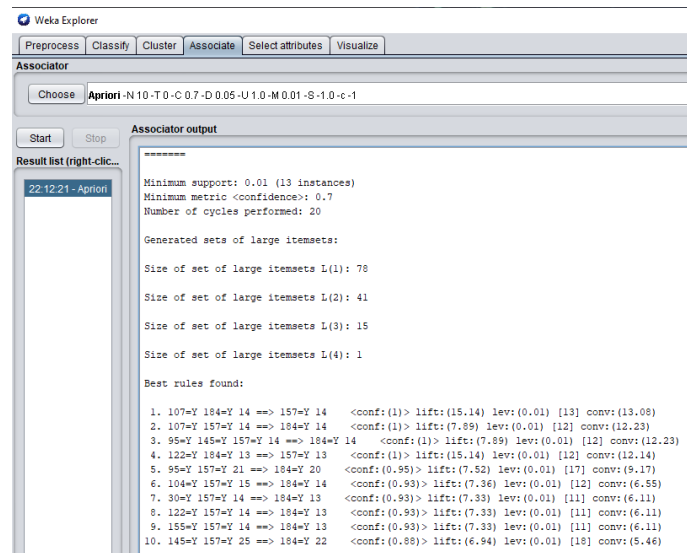


Fig 3. Weka Results

Then with a minimum of 1% support and confidence, 70% formed final Association rules as much as 10 final association rules.

- a. If you buy "faucets" and "cement," then it will buy "sand."
- b. If buying "faucets" and "sand," then it will buy "cement."
- c. If you buy "ceramic," "nail," and "sand," then it will buy "cement."
- d. If buying "glue isarplas" and "cement," then it will buy "sand."
- e. If you buy "ceramic" and "sand," then it will buy "cement."
- f. If you buy "Kloset" and "sand," then it will buy "cement."
- g. If you buy "cat aviat" and "sand," then it will buy "cement."
- h. If buying "glue isarplas" and "sand," then it will buy "cement."
- i. If you buy a "Lank lips" and "sand," then it will buy "cement."
- j. If you buy "nail" and "sand," then it will buy "cement."

5. Conclusion

The final result of the association pattern on the purchase of material goods UD Dian Pertiwi using a apriori algorithm method of transaction data for 5 months is from testing 1,302 transaction data for 5 months using Weka, generating the best rules with Minimum 1% support and confidence 70% formed final Association rules as many as 10 association rules.

The approach of data mining by implementing association technology using apriori algorithm on sales transactions in UD Dian Pertiwi can make a useful contribution, so UD Dian Pertiwi can find the rules of association by looking at the material sold The same, and more precise in providing stock for building materials that have the highest interest.

References

- [1] P. N. Tan, M. Stenbach, and V. Kumar, "Introduction to Data Mining." Boston: Pearson Education, 2006.
- [2] F. Goronescu, "Data Mining: Concepts, Models, and Techniques." Verlag Berlin Heidelberg: Springer. 2011.
- [3] I. H. Witten, E. Frank, and M. A. Hall. "Data Mining: Practical Mchine Learning Tools and Techniques, 3rd ed". Burlington, MA: Morgan Kaufmann, 2011.
- [4] Lahouar A, Ben Hadj Slama J. Day-ahead load forecast using random forest and expert input selection. Energy Convers Manage 2015;103:1040–51.

- [5] Mahmoud Thair S, Habibi Daryoush, Hassan Mohammed Y, et al. Modelling selfoptimised short term load forecasting for medium voltage loads using tuning fuzzy systems and Artificial Neural Networks. *Energy Convers Manage* 2015;106:1396–408.
- [6] Saboori H, Hemmati R, Abbasi V. Multistage distribution network expansion planning considering the emerging energy storage systems. *Energy Convers Manage* 2015;105:938–45.
- [7] Wolisz H, Punkenburg C, Streblov R, Müller D. Feasibility and potential of thermal demand side management in residential buildings considering different developments in the German energy market. *Energy Convers Manag* 2016;107:86–95.
- [8] Wang F, Xu H, Xu T, et al. The values of market-based demand response on improving power system reliability under extreme circumstances. *Appl Energy* 2017;193:220–31.
- [9] Wang F, Zhou L, Ren H, et al. Multi-objective optimization model of source-load-storage synergetic dispatch for building energy system based on TOU price demand response. *IEEE Trans Ind Appl* 2017;54:1017–28.
- [10] Chicco G. Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy* 2012;42:68–80.
- [11] Chicco G, Ilie I-S. Support vector clustering of electrical load pattern data. *IEEE Trans Power Syst* 2009;24:1619–28.
- [12] Chicco G, Ionel O, Porumb R. Electrical load pattern grouping based on centroid model with ant colony clustering. *IEEE Trans Power Syst* 2013;28:1706–15.
- [13] Chen Q, Wang F, Hodge BM, et al. Dynamic price vector formation model based automatic demand response strategy for PV-assisted EV charging station. *IEEE Trans Smart Grid* 2017;8:2903–15.
- [14] Granell R, Axon CJ, Wallom DCH. Clustering disaggregated load profiles using a Dirichlet process mixture model. *Energy Convers Manag* 2015;92:507–16.
- [15] Rhodes JD, Cole WJ, Upshaw CR, Edgar TF, Webber ME. Clustering analysis of residential electricity demand profiles. *Appl Energy* 2014;135:461–71.