

# Implementation of Data Mining Using Association Rules for Transactional Data Analysis

Muhamad Brilliant<sup>1</sup>, DwiHandoko<sup>2</sup>, Sriyanto<sup>3</sup>

Department of Informatics Engineering.

Informatics and Business Institute Darmajaya

Email: [muhamad.1721211017@mail.darmajaya.ac.id](mailto:muhamad.1721211017@mail.darmajaya.ac.id), [dwihandoko2401@gmail.com](mailto:dwihandoko2401@gmail.com), [sriyanto@darmajaya.ac.id](mailto:sriyanto@darmajaya.ac.id)

## ABSTRACT

Data is an important property for everyone. Large amount of data is available in the world. There are various repositories to store the data into data warehouses, databases, information repository, etc.. This large amount of data needs to process so that we can get useful information. Data mining is a technique to get information that hidden from collections of data. There are several major functions in data mining such as estimation, prediction, classification, clustering and association. This research use association rule to find the interconnections of the association between the data items in data transaction. The technique used to find the rule is the FP-Growth. FP-Growth is one of the algorithms used to find frequent item sets in the set of transaction data.

This study aims to create a simulation using a data mining association rule with the FP - growth algorithm as a reference to determine a list of product packages that offered to consumers. Testing that has been done based on the results of functional testing with black box method, it can be concluded that by implementing data mining with association rule method can help the company in finding consumer pattern. It is expected that the company can create a list of entertainment service, product packages that can be offered to consumers based on the rules generated at competitive prices

**Keywords:** Association Rule, Data Mining, FP-Growth

## 1. Introduction

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. Data mining is an analytical tool for analyzing data [3]. It allows users to analyze data, categorize it, and summarize the relationships among data [1]. Many companies have problems on their business and need to solve using data mining approach. One of the companies is Barcelona Professional Sound System (BPSS). It is located at Jl. Gotong Royong No.164 at Pringsewu Lampung. This company engaged in event organizer services that rent out entertainment products and services at various events such as weddings, product promotion from the company. Services provided such as rents, lighting equipment, and sound system.

Based on interviews with the owners, obtained some products and services those are less in demands by consumers. There are many competitors who offer products and services with a cheaper price. With the onset of the matter, came the idea of the owners to run a sales strategy in the form of packets. The company expects this strategy can be offered to consumers with a price that can compete with other competitors. The availability of large data and has a pattern of similar data on each transaction that occurs then data mining can be considered for this research.

The method used in this research is the Association Rule. Association Rule is a method of data mining that can find the interconnections of the association between the data items in a data transaction. The technique used to find the

rules associative one is the FP Growth. Getting a combination of items, the data will be dimmed using FP-Growth algorithm.

## 2. Result Method

### 2.1 Data Collection Methods

Data collection methods used in this research to obtain data are interview, observation, and documentation

### 2.2 Data Extraction

Data extraction is performed in this stage is making the transaction data from database. Extracted data is stored in a database and then followed by the data processing to the next stage.

### 2.3 Attributes Selection

Based on the information you want to get by users regarding entertainment products and services anything purchased simultaneously by consumers, so at this stage the attributes that will be used is id-trx, item-name which aims to select what data will be used during the process of mining and the necessary database for purposes of analysis, since not all of the attributes used during the mining process.

### 2.4 Selecting Modeling Techniques

By using Data Mining, transaction data will be processed by association rules to find patterns of selection of entertainment products and services often hired by consumers. In this stage the method used is Association Rule.

**Table 1. Sample transaction data**

TID	ITEM
001	D
002	B,G
003	D,E
004	B,D
005	D,E
006	B,D
007	B,C,D
008	D,E,C,G
009	B,E
010	B,G
011	DCG
012	E,C
013	B
014	D,E
015	C
016	D,E
017	B,E,C
018	E
019	E,C
020	B,C,G
021	B,C,G
022	B,C,F
023	B,C
024	B,D,G
025	D
026	E,C
027	D,C,G
028	B,E,G
029	D
030	D,C,G
031	B,D,E
032	B,E,G
033	C,G
034	D
035	B,G
036	D,C
037	B,E
038	D,E,G
039	B,D
040	D,E
041	G
042	B,D
043	E,C
044	B,G
045	E,C
046	B,D,G
047	B,E
048	B,E
049	B,D,E
050	E,F

This study aims to find patterns of consumer purchasing associations. The modeling techniques used in this research is the method of Association Rules while the algorithm used in this study is the algorithm FP-Growth. The first phase of the database search is used to calculate the support value of each item and select the item that meets the minimum support. The results of the database search known the frequency of occurrence of each item in the database and sort by the frequency of occurrence of the largest item.

**Table 2. Tabel Frequent List**

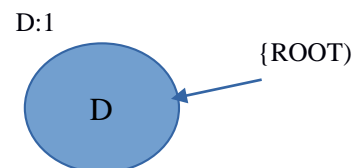
B	25
D	24
E	23
C	18
G	18
F	14
A	9

After getting the Frequent List, then change the item according to Frequent List by eliminating items A, F and in sort of the largest on each item-transaction. Making FP-Tree based Transaction ID will be given initials sequence number to simplify the process formation of FP-Tree according to which is listed in the table below.

**Table 3. Table Frequent List**

TID	ITEM
001	A,D
002	B,E,I
003	D,E
004	A,B,I
005	D,E
006	B,D,I
007	B,C,I
008	C,D,I
009	B,E,I
010	B,G
011	C,D,I
012	C,E
013	A,B
014	D,E,I
015	C
016	D,E,I
017	B,C,I
018	E
019	C,E
020	A,B,I
021	A,B,I
022	A,B,I
023	B,C
024	B,D,I
025	D,F
026	C,E
027	C,D,I
028	B,E,I
029	A,D
030	C,D,I
031	B,D,I
032	B,E,I
033	C,F,I
034	A,D
035	B,G
036	C,D
037	B,E,I
038	D,E,I
039	B,D,I
040	D,E
041	G
042	A,B,I
043	C,E
044	B,F,I
045	C,E,I
046	B,D,I
047	B,E
048	B,E
049	B,D,I
050	E,F

Making FP-Tree starts from TID 1 is {D}



**Figure 2. FP-tree TID 1**

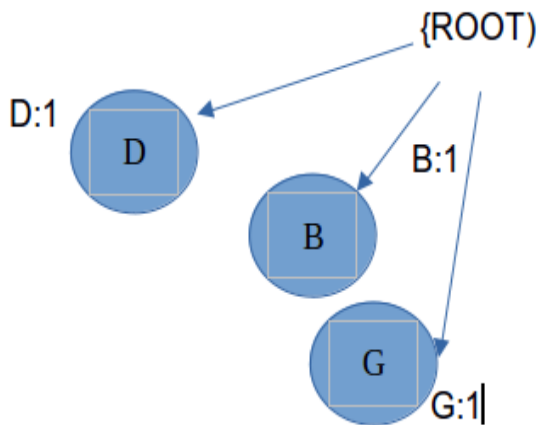


Figure 3. FP-tree TID 2

After reading TID 1, then TID 2 is done {B,G} to form a path {} → B → G with support count value for beginning is 1

After FP-Tree finished, FP-Growth algorithm searching all possible subsets by generating conditionals FP-Tree and look for frequent itemsets, by applying the method of Divide and Conquer in the order of Frequent List from the smallest number frequencies its appearance.

From calculation confidence against the pattern formed above, then Association Rule that meet the requirement of confidence  $\geq 0.75$

### 3. Results And Discussion

In this study we using rapidminer as simulation tool to perform data mining simulation using FP-Growth algorithm and association rules method on BPSS transaction data.

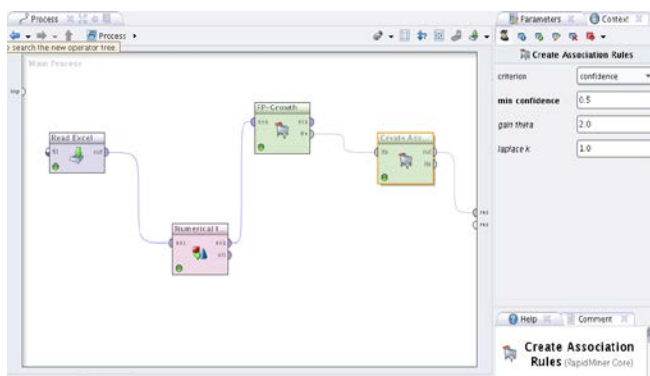


Figure 4 . Association rules in rapid miner

On the picture above shows the minimum value of confidence given is 0.5 or 50% then the results of the first simulation generate many rules, the highest value of confidence given the less rules will be generated. this is the result of his experiment

Premises	Conclusions	Support	Confidence
decoration, fashion makeup	party Tents	0.047619048	0.75
Regular Tents, Regular Sound System	decoration	0.047619048	0.75
party Tents, decoration, Regular Tents	Regular Sound System	0.047619048	0.75
Regular Tents, Regular Sound System	party Tents, decoration	0.047619048	0.75
party Tents, Regular Tents, Regular Sound System	decoration	0.047619048	0.75
decoration, Regular Sound System	party Tents	0.047619048	1
Regular Tents, Regular Sound System	party Tents	0.063492063	1
fashion makeup, Regular Sound System	party Tents	0.047619048	1

Picture 5 . Rules 1

for the results of rules that get better the value of confidence should we increase, in simulation 2 value of confidence increased to 80% or 0.8

No.	Premises	Conclusion	Support	Confidence
1	sound system rigging, fashion makeup	party Tents	0.063	0.500
2	fashion makeup	sound system rigging	0.127	0.533
3	party Tents, Regular Tents	decoration	0.063	0.571
4	party Tents, Regular Tents	Regular Sound System	0.063	0.571
5	party Tents, Regular Sound System	Regular Tents	0.063	0.571
6	fashion makeup	party Tents	0.143	0.600
7	decoration	party Tents	0.190	0.632
8	Regular Sound System	party Tents	0.111	0.700
9	sound system rigging, decoration	party Tents	0.079	0.714
10	decoration, fashion makeup	party Tents	0.048	0.750
11	Regular Tents, Regular Sound System	decoration	0.048	0.750
12	party Tents, decoration, Regular Tents	Regular Sound System	0.048	0.750
13	Regular Tents, Regular Sound System	party Tents, decoration	0.048	0.750
14	party Tents, Regular Tents, Regular Sound System	decoration	0.048	0.750
15	decoration, Regular Sound System	party Tents	0.048	1
16	Regular Tents, Regular Sound System	party Tents	0.063	1
17	fashion makeup, Regular Sound System	party Tents	0.048	1
18	decoration, Regular Sound System	Regular Tents	0.048	1
19	decoration, Regular Sound System	party Tents, Regular Tents	0.048	1
20	party Tents, decoration, Regular Sound System	Regular Tents	0.048	1

Figure 6 . Rules 2

based on the simulation results obtained rules are shown in table 4

Table 4. Result Rule being Package

No.	Premises	Conclusion	Support	Confidence
1	decoration, fashion makeup	party Tents	0.048	0.750
2	Regular Tents, Regular Sound System	decoration	0.048	0.750
3	party Tents, decoration, Regular Tents	Regular Sound System	0.048	0.750
4	Regular Tents, Regular Sound System	party Tents, decoration	0.048	0.750
5	party Tents, Regular Tents, Regular Sound System	decoration	0.048	0.750
6	decoration, Regular Sound System	party Tents	0.048	1
7	Regular Tents, Regular Sound System	party Tents	0.063	1
8	fashion makeup, Regular Sound System	party Tents	0.048	1

From the results it shows when the customer selects Regular Tents, Regular Sound System it will also choose Decoration with the value of confidence 75% and the value of support 0,4.

### 4. Conclusion

Based on the analysis and discussion of the problem, the conclusions of this study are as follows :

1. Simulation is done to provide information menu by the consumer selection patterns of entertainment products and services.
2. Applications are built to help provide information about services and products often selected by consumers so they can considered for the formation of entertainment products and service packages for consumers.

## Bibliographies

- [1] Larose, Daniel. T. 2005. "Discovering Knowledge In Data – An Introduction to Data Mining". John Wiley & Sons, Inc, New Jersey
- [2] Kumar,B.Santhosh , K.V.Rukmani .Implementation of Web Usage Mining Using APRIORI and FP Growth Algorithms, Int. J. of Advanced Networking and Applications Volume:01, Issue:06, Pages: 400-404 (2010)
- [3] Tan, P. N., M. Steinbach, V. Kumar, "Introduction to Data Mining", Addison-Wesley, 2005, 769pp.
- [4] Raorane A.A. Kulkarni R.V. and Jitkar B.D. Association Rule – Extracting Knowledge Using Market Basket Analysis. Research Journal of Recent Sciences ISSN 2277-2502 Vol. 1(2), 19-27, Feb. (2012)
- [5] Vijaylakshmi S., Mohan V., Suresh Raja S., Mining of users access behavior for frequent sequential pattern from web logs, International Journal of Database Management System (IJDM), 2, (2010)
- [6] Parvinder S. Sandhu Dalvinder, Dhaliwal S. and Panda S.N., Mining utility-oriented association rules: An efficient approach based on profit and quantity, International Journal of the Physical Sciences, 6(2), 301-307 (2011)
- [7] Narvekar ,Meera , ShafaqueFatmaSyed.An optimized algorithm for association rule mining using FP tree International Conference on Advanced Computing Technologies and Applications (ICACTA- 2015)
- [8] Yen-Liang Chen ,Kwei Tang , \*, Ren-Jie Shen a, Ya-Han Hu. Market basket analysis in a multiple store environment. Elsevier joournal Decision Support Systems 40 (2005) 339 – 354
- [9] Karabatak ,Murat , , M. CevdetInce An expert system for detection of breast cancer based on association rules and neural network Elsevier joournal Expert Systems with Applications 36 (2009) 3465–3469
- [10] Kaur ,Manpreet, ShivaniKang.Market Basket Analysis: Identify the changing trends of market data using association rule mining.International Conference on Computational Modeling and Security (CMS 2016) Procedia Computer Science 85 ( 2016 ) 78 – 85
- [11] Borgelt, T. 2005. "An Implementation of FP Growth Algorithm" ACM Digital Library, Chicago, USA