

Application of Pizza Sales Data Mining Using Apriori Method

Rusdiansyah
Bina Sarana Informatika University
West Jakarta Indonesia
Rusdiansyah.rds@bsi.ac.id

Nining Suharyanti
Bina Sarana Informatika University
West Jakarta City, Indonesia
Nining.nni@bsi.ac.id

Triningsih
Bina Sarana Informatika University
West Jakarta Indonesia
triningsih.tnh@bsi.ac.id

Murniyati
Bina Sarana Informatika University
West Jakarta Indonesia
murni.ndk@gmail.com

Submitted: Feb 19, 2020

Accepted: Mar 27, 2020

Published: Apr 1, 2020

Abstract— Pizza is a processed food originating from Italy and has been spread in various other countries including one of them in Indonesia. Pizza is a processed food that is currently sought after by various groups of people so as to make the pizza business opportunity very profitable, if it is run in a food business. Currently the pizza business has very favorable prospects when compared to other businesses. Moreover, the targeted target can be from all walks of life from children to adults. Pizza sales transactions that produce sales data every day, have not been able to maximize the use of sales data. Sales data is only stored as an archive, so it becomes a pile of data. Therefore the use of data mining is used to solve this problem. A priori algorithm is a data mining method by using minimum support parameters, minimum confidence and will analyze in the period of every month of sales transactions. This study produces data on the results of the process of association rules from the data collection of sales transactions. From the association rules it can be concluded that the pattern of pizza sales, where consumers more often buy Meatzza and Cheese Mania, as evidenced by the results of calculations using Apriori Algorithm and Rapidminer 5.3, with support of 30% and 60% confidence.

Keywords— Pizza Sales, Apriori Algorithms, Association Rule

I. INTRODUCTION

Pizza is a processed food that is currently sought after by various groups of people, thus making the pizza business opportunity very profitable, if run in the food business. To start a business selling Pizza, including one selling fast food. (Adibah & Novita, 2019). The Pizza sales business has quite a number of competitors, so skills are needed to attract the attention of consumers by refining, perfecting, and providing various pizza variants. Business activities are marked by intensifying competition between one business actor and another business actor.

Pizza is quite popular among the people and quite popular from children to adults, research to examine which pizza is most often bought by consumers. To find out the buyer's interest, one of the information that can be obtained is from the Sales Data, from the Sales Data it will show which pizza variants are often bought by consumers.

Problems that arise can be identified as follows:

1. A lot of data that has not been neatly organized and only stored without being processed into information that is useful for increasing sales.

2. Choosing a pizza variant that is too much for new consumers to have difficulty in choosing, causing the queue to be too long and take longer and occasionally errors in consumer orders.

II. LITERATURE REVIEW

Simply put Data Mining is mining or discovering new information by looking for certain patterns or rules from a very large amount of data (Rismayanti, Damayanti, & Khairunnisa, 2019), Data mining is a method for finding knowledge in a large pile of data. Data mining is the process of digging and analyzing a very large amount of data to obtain something that is true, new and useful and finally can be found a pattern in the data (Rodiyansyah, 2015). A priori algorithms include types of association rules in Data Mining. Rules that state the association between several attributes are often called affinity analysis or market basket analysis. Association analysis or association rule mining is a Data Mining technique to find the rules for a combination of items. One of the stages of association analysis that attracts many researchers to produce efficient algorithms is analysis of high frequency patterns (frequent pattern mining) (Badrul, 2016). FP-Growth is a development of a priori algorithm. A priori algorithm requires generating candidates to get frequent itemsets (Meilani & Azinar, 2015). However, in the FP-Growth algorithm, generating candidates is not done because FP-Growth uses the concept of tree development called Frequent Pattern-Tree (FP-Tree) in the search for frequent itemsets. That causes FP-Growth algorithm to be faster than Apriori algorithm. So the shortcomings of the a priori algorithm are corrected by the FP-Growth algorithm. The FP-growth algorithm can be implemented in a short time and produces high accuracy when used to analyze the association rule (Muliono, 2017).

III. PROPOSED METHOD

The research instrument is a tool or facility used by researchers in gathering data to make work easier and the results better, in the sense of being more accurate, complete, and systematic so that it is easier to process (Nasution, 2019).

The author uses quantitative data analysis in research taken by the author, this research which is an analysis prioritizing data in the form of numbers and also calculations using formulas related to writing analysis, in this case Apriori Algorithm analysis will be used as follows:

- A. Analysis of sales problems Problems will be analyzed using the Apriori Algorithm Method.
- B. Data processing with a priori algorithm calculation

The following are some of the stages that will be performed in calculations with the Apriori Algorithm (Satie, Suparni, & Pohan, 2020):

1. Look for the biggest value that has sold the most
2. The initial step is to find the highest sales value in a month's transaction data with the following steps:
 - a. determine the pizza list
 - b. determine pizza sales data
3. Determine pizza sales data
4. Grouping some of the most frequently purchased pizza.
5. Making a Tabular Format
6. If the biggest sales value from the sale of pizza that is often bought every month is known, it will be made tabular format so that it can be analyzed with Algorithm Apriori.

High Frequency Pattern Analysis in this stage to find a combination of items that meet the minimum requirements of the support value in the database, which can be formulated as follows (Wahyuningtias & Rusdiansyah, 2019):

$$Support(A) = \frac{\text{Number of Transactions } A \times 100\%}{\text{Transaction Total}}$$

Meanwhile, the formula 2 itemset with the formula:

$$Support(A, B) = \frac{\text{Number of Transactions } A \text{ and } B \times 100\%}{\text{Transaction Total}}$$

The search for high frequency patterns will be stopped if the combination does not meet the specified Support value.

IV. RESULT AND DISCUSSION

Transaction data that occurs every day to make it easier with a large amount of data because of sales activities every day the data is getting more and more time. The place has not been said to be neatly arranged in the processing of these data only stored and archived not utilized and processed into useful information to increase sales.

TABLE I. LIST OF PIZZA VARIAN PRODUCTS

PIZZA DATA TABLE		
Number	PIZZA CATEGORY (MEDIUM, THIN CRUST)	
1	CM	CHEESE MANIA
2	PF	PEPPER FEAST
3	CL	CHK LOVERS
4	ACCB	AMR CLS CHEESEBURGER
5	MT	MEATZZA
6	EX	EXTRAVAGANZZA
7	AS	AMR ALL STAR
8	GBS	GRILLED BEEF SUPREME
9	DBB	DBL BEEF BRGR
10	CHS	CHEESY SAUSAGE
11	VM	MEAT n MEAT
12	TNDE	TUNA DELIGHT
13	CDO	CHK DOMINATOR
14	BDL	BEEF DELIGHT
15	CPF	CHK PEP FEAST
16	MAR	MARGHERITA
17	CHD	CHICKEN DELIGHT
18	SBB	SAMBAL BEEF

Item set formation

Settlement based on Table 1 provided by the process of forming C1 or referred to as Itemset, Itemset with a Minimum Amount of 30%.

From the process of forming an itemset with a minimum support of 30%, it can be seen that those who meet the minimum support standards are: PF, CM, AS, TNDE, MT, ACCB, EX, VM, VEM, CL, MAR, GBS, CDO, DBB, CPF, BDL , CHD, SBB, CHS.Kombinasi 2 Itemset

The process of forming C2 or can be called 2 Itemset with a minimum amount of support of 30%. Can be solved by the following formula:

From the process of forming C2 or 2 Itemset in Table.II with a minimum support of 30%, it can be seen that those who meet the minimum support standards are:

TABLE II. 2 ITEM SET

Support Table of 2 Item set minimum of 30%		
Itemset	amount	Support
S (PF , CM)	12	38.7%
S (CM , MT)	12	38.7%
S (CM , AS)	11	35.5%
S (PF , MT)	11	35.5%

Formation of Association Rules

After the High frequency pattern is found, then look for the Association Rules that meet the Minimum Requirements for Confidence, by calculating the Confidence Rules Association A → B with a Minimum Confidence of 60%. This is the value of Rules A → B obtained.

$$Confidence = (A|B) = \frac{\text{Number of Transactions A and B X100}}{\text{Transaction Total A}}$$

TABLE III. CONFIDENCE

1 . Confidence [CM] --> [PF]	12/18*100 = 66.7
2 . Confidence [PF] --> [CM]	12/18*100 = 66.7
3 . Confidence [CM] --> [MT]	12/18*100 = 66.7
4 . Confidence [MT] --> [CM]	12/15*100 = 80
5 . Confidence [CM] --> [AS]	11/18*100 = 61.11
6 . Confidence [AS] --> [CM]	11/18*100 = 61.11
7 . Confidence [PF] --> [MT]	11/18*100 = 61.11
8 . Confidence [MT] --> [PF]	11/15*100 = 73.33

TABLE IV. ATURAN ASOSIASI

Table Calculation of Association Rules 60%		
The rules	Confidence	
If you buy CM, you will buy PF	12/18*100	66.7%
If you buy PF, you will buy CM	12/18*100	66.7%
If you buy CM, you will buy MT	12/18*100	66.7%
If you buy MT, you will buy CM	12/15*100	80.0%

If you buy CM, you will buy AS	11/18*100	61.1%
If you buy US, you will buy CM	11/18*100	61.1%
If you buy PF, you will buy MT	11/18*100	61.1%
If you buy MT, you will buy PF	11/15*100	73.3%

System Implementation

The a priori algorithm implementation in this study uses the Rapidminer application for testing.

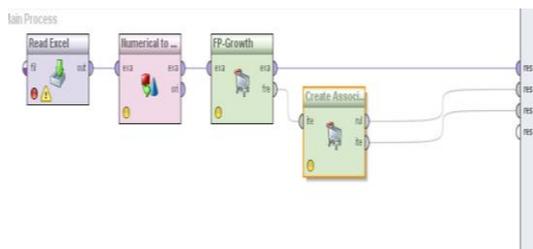


Fig. 1. Model Design

In Fig 1. The design model is formed through the process steps by connecting each operator, namely: Read Excel change to Numeric to Binomial, Numeric to Binomial transfer to Fp-Growth, Fp-Growth transform to create Association Rules and Create Association Rules change to the Results Process

Then the results of the Rule are shaped into 8 rules from the results of Rapidminer 5.3 as follows:

No.	Premises	Conclusion	Support	Confid.	LePl.	Gain	p-s	Lift	Conv.
1	PF	MT	0.355	0.611	0.657	-0.004	0.074	1.263	1.327
2	CM	AS	0.355	0.611	0.657	-0.004	0.055	1.164	1.244
3	PF	CM	0.387	0.667	0.878	-0.774	0.050	1.148	1.258
4	CM	PF	0.387	0.667	0.878	-0.774	0.050	1.148	1.258
5	CM	MT	0.387	0.667	0.878	-0.774	0.106	1.378	1.548
6	AS	CM	0.355	0.688	0.894	-0.671	0.055	1.164	1.342
7	MT	PF	0.355	0.733	0.913	-0.611	0.074	1.263	1.573
8	MT	CM	0.387	0.800	0.935	-0.581	0.106	1.378	2.097

Fig. 2. Final Association

In Fig 2. From the overall results of the a priori algorithm implementation process in manual calculations and calculations in Rapidminer for pizza sales transaction data as many as 224 data by providing a minimum limit of 30% Support and 60% Confidence that has been set, so as to get the results of Rule 8 as follows:

1. Pepper Feast, Meatzza with a Support Value of 0.355 and a Confidence Value of 0.6112.
2. Cheese mania, Amr All Star with a Support Value of 0.355 and a Confidence Value of 0.611

3. Pepper Feast, Chesee mania with Support Value of 0.387 and Confidence Value of 0.6674.
4. Cheese mania, Pepper Feast with Support Value of 0.387 and Confidence Value of 0.667
5. Cheese mania, Meatzza with Support Value of 0.387 and Confidence Value of 0.667
6. Amr All Star, Chesee mania with a Support Value of 0.355 and a Confidence Value of 0.688
7. Meatzza, Pepper Feast with Support Value of 0.355 and Confidence Value of 0.733
8. Meatzza, Chesee mania with Support Value of 0.387 and Confidence Value of 0.800



Fig. 3. Graph Display

In Fig 3. It can be concluded from the value of the most superior rules with 38.7% Support and 80.0% Confidence is pizza (If you buy Meatzza then you will buy Chesee mania).

V. CONCLUSION AND SUGGESTION

Based on the data and the results of the discussion, the following conclusions can be drawn: This study produces process results data from the association rules of the sales transaction data set. From the association's rules can be obtained by the pattern of pizza purchases, where consumers more often buy Meatzza and Cheese Mania, as evidenced by the results of calculations using Apriori Algorithms and Rapidminer 5.3, that with 30% support and 60% confidence and the final results of the rules 8. Support 38.7% and 80.0% Confidence are pizza. If you buy Meatzza, you will buy Cheese Mania.

From the description of the research the advice is as follows:

1. Data Mining with Apriori Algorithm has the disadvantage of having to scan the database every time, so for very large databases it takes a long time.

In implementing the pizza sales report not only can use a data mining method a priori algorithm but can use other methods for future comparison. Dalam mengimplementasikan laporan penjualan pizza bukan hanya bisa menggunakan data mining metode algoritma apriori tetapi bisa menggunakan metode lain sebagai bahan perbandingan kedepannya. Rodiyansyah, S. F. (2015). Algoritma Apriori untuk Analisis Keranjang Belanja pada Data Transaksi Penjualan. *Infotech*, 1(1), 36–39. Retrieved from <http://jurnal.unma.ac.id/index.php/infotech/article/view/42>

2.

3. The addition of research objects does not only cover one month but can be up to several months backward so that data is more accurate and in subsequent studies it is expected to develop a computerized system based on data mining processing using a priori algorithms to analyze transactions that occurred in the last period to be made business decision making process going forward. in an application that can be easily used for data processing.

V. REFERENCES

- Adibah, F., & Novita, D. (2019). STRATEGI UNTUK PERMINTAAN PENUH PADA RESTAURAN M2M INDONESIA FAST FOOD CABANG BANGIL. *Majalah Ekonomi*, 69-83.
- Badrul, M. (2016). ALGORITMA ASOSIASI DENGAN ALGORITMA APRIORI UNTUK ANALISA DATA PENJUALAN. *Jurnal Pilar Nusa Mandiri*, 121-129.
- Meilani, B. D., & Azinar, A. W. (2015). Penentuan Pola Yang Sering Muncul Untuk Penerima Kartu Jaminan Kesehatan Masyarakat (JAMKESMAS) Menggunakan Metode FP-Growth. *Seminar Nasional "Inovasi dalam Desain dan Teknologi" - IDEaTech 2015*, 424-431.
- Muliono, R. (2017). Analisis Efisiensi Algoritma Data Mining. *Semantika (Seminar Nasional Teknik Informatika)* (pp. 117-123). Medan: Politeknik Ganesha.
- Nasution, I. S. (2019). PENGARUH MODEL PEMBELAJARAN KOOPERATIF TIPE THINK PAIR SHARE TERHADAP HASIL BELAJAR MAHASISWA PADA MATA KULIAH PENGANTAR DASAR MATEMATIKA-FKIP UMSU. *MES: Journal of Mathematics Education and Science*, 160 - 166.
- Rismayanti, R., Damayanti, F., & Khairunnisa, K. (2019). Penerapan Data Mining Algoritma C4.5 dalam Menentukan Rekam Jejak Kinerja Dosen STT Harapan Medan. *Sinkron (Jurnal & Penelitian Teknik Informatika)*, 99-104.
- Rodiyansyah, S. F. (2015). Algoritma Apriori untuk Analisis Keranjang Belanja pada Data Transaksi Penjualan. *Infotech Journal*, 36-39.
- Satie, D. E., Suparni, S., & Pohan, A. B. (2020). Analisa Algoritma Apriori Pada Pola Peminjaman Buku di Perpustakaan ITB Ahmad Dahlan. *Media Informatika Budi Dharma*, 136-143.
- Wahyuningtias, Y., & Rusdiansyah, R. (2019). ANALISIS PENERAPAN ASOSIASI UNTUK MENENTUKAN TRANSAKSI PENJUALAN PADA WHAT'S UP CAFÉ DENGAN METODE ALGORITMA APRIORI. *Jurnal Riset informatika*, 181-186.