

Algoritma Apriori untuk Pencarian *Frequent itemset* dalam *Association Rule Mining*

Lusa Indah Prahartiwi ^{1,*}, Wulan Dari ¹

¹ Sistem Informasi; STMIK Nusa Mandiri; Jl. Damai No. 8 Warung Jati Barat Jakarta Selatan DKI Jakarta, 021-78839502; e-mail: lusaindah@gmail.com,
e-mail: wulan.wld@nusamandiri.ac.id

* Korespondensi: e-mail: lusaindah@gmail.com

Diterima: 1 Agustus 2019; Review: 22 Agustus 2019; Disetujui: 12 September 2019

Abstract

Over decades, retail chains and department stores have been selling their products without using the transactional data generated by their sales as a source of knowledge. Abundant data availability, the need for information (or knowledge) as a support for decision making to create business solutions, and infrastructure support in the field of information technology are the embryos of the birth of data mining technology. Association rule mining is a data mining method used to extract useful patterns between data items. In this research, the Apriori algorithm was applied to find frequent itemset in association rule mining. Data processing using Tanagra tools. The dataset used was the Supermarket dataset consisting of 12 attributes and 108.131 transaction. The experimental results obtained by association rules or rules from the combination of item-sets beer wine spirit-frozen foods and snack foods as a *Frequent itemset* with a support value of 15.489% and a confidence value of 83.719%. Lift ratio value obtained was 2.47766 which means that there were some benefits from the association rule or rules.

Keywords: *Apriori, Association Rule Mining.*

Abstrak

Selama beberapa dekade rantai ritel dan department store telah menjual produk mereka tanpa menggunakan data transaksional yang dihasilkan oleh penjualan mereka sebagai sumber pengetahuan. Ketersediaan data yang melimpah, kebutuhan akan informasi (atau pengetahuan) sebagai pendukung pengambilan keputusan untuk membuat solusi bisnis, dan dukungan infrastruktur di bidang teknologi informasi merupakan cikal-bakal dari lahirnya teknologi *data mining*. *Data mining* menemukan pola yang menarik dari *database* seperti *association rule*, *correlations*, *sequences*, *classifier* dan masih banyak lagi yang mana *association rule* adalah salah satu masalah yang paling populer. *Association rule mining* merupakan metode *data mining* yang digunakan untuk mengekstraksi pola yang bermanfaat di antara data barang. Pada penelitian ini diterapkan algoritma Apriori untuk pencarian *frequent itemset* dalam *association rule mining*. Pengolahan data menggunakan tools Tanagra. Dataset yang digunakan adalah dataset Supermarket yang terdiri dari 12 atribut dan 108.131 transaksi. Hasil eksperimen diperoleh aturan asosiasi atau *rules* dari kombinasi *itemsets* beer wine spirit-frozen foods dan snack foods sebagai *Frequent itemset* dengan nilai *support* sebesar 15,489% dan nilai *confidence* sebesar 83,719%. Nilai *Lift ratio* yang diperoleh sebesar 2,47766 yang artinya terdapat manfaat dari aturan asosiasi atau *rules* tersebut.

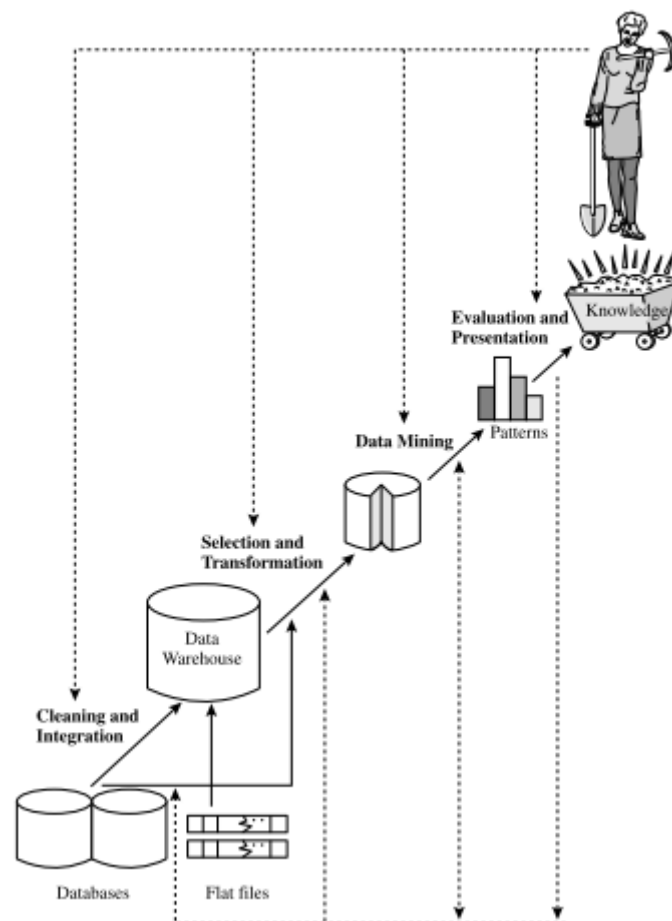
Kata kunci: *Apriori, Association rule mining*

1. Pendahuluan

Banyaknya persaingan bisnis khususnya Supermarket menuntut pemilik supermarket membuat strategi bisnis untuk meningkatkan penjualan. Salah satu strategi bisnis untuk

meningkatkan penjualan ialah melalui promosi. Untuk menentukan promosi yang tepat sasaran, pemilik supermarket harus mengetahui selera konsumen dan kondisi pasar. Pemilik supermarket dapat mengetahui apa saja barang yang dibeli oleh konsumen dengan menganalisis kebiasaan membeli konsumen. Selama beberapa dekade rantai ritel dan department store telah menjual produk mereka tanpa menggunakan data transaksional yang dihasilkan oleh penjualan mereka sebagai sumber pengetahuan [Videla-Cavieres and Ríos, 2014]. Ketersediaan data yang melimpah, kebutuhan akan informasi (atau pengetahuan) sebagai pendukung pengambilan keputusan untuk membuat solusi bisnis, dan dukungan infrastruktur di bidang teknologi informasi merupakan cikal-bakal dari lahirnya teknologi *data mining* [Moertini, 2009]. *Data mining* diprediksi menjadi “salah satu perkembangan paling revolusioner pada dekade berikutnya,” menurut majalah teknologi online ZDNET News (8 Februari 2001). Bahkan, MIT Technology Review memilih *data mining* sebagai salah satu dari 10 teknologi baru yang akan mengubah dunia [Larose, 2006].

Data mining merupakan teknologi baru yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka [Moertini, 2009]. Informasi dan pengetahuan yang diperoleh dapat digunakan untuk aplikasi mulai dari analisis pasar, deteksi penipuan, dan retensi pelanggan hingga kontrol produksi dan eksplorasi sains [Han and Kamber, 2014]. *Data mining* menemukan pola yang menarik dari database seperti *association rule*, *correlations*, *sequences*, *classifier* dan masih banyak lagi yang mana *association rule* adalah salah satu masalah yang paling populer [Annie and Kumar, 2012]. *Data mining* sering dikenal dengan istilah populer *Knowledge Discovery from Data* atau KDD [Han and Kamber, 2014]. Proses *Knowledge Discovery from Data* ditunjukkan pada Gambar 1.



Sumber: [Han and Kamber, 2014]

Gambar 1. Proses Knowledge Discovery from Data atau KDD

Langkah-langkah proses KDD [Han and Kamber, 2014] yaitu: **1) Data Cleaning** (untuk menghilangkan noise dan data yang tidak konsisten, **2) Data Integration** (di mana banyak sumber data dapat digabungkan), **3) Data Selection** (di mana data yang relevan dengan tugas analisis diambil dari database), **4) Data Transformation** (di mana data ditransformasikan atau dikonsolidasikan ke dalam bentuk yang sesuai untuk penambangan dengan melakukan operasi ringkasan atau agregasi, **5) Data Mining** (proses penting di mana metode cerdas diterapkan untuk mengekstraksi pola data, **6) Pattern Evaluation** (untuk mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan pada beberapa tindakan menarik, **7) Knowledge Presentation** (di mana Teknik visualisasi dan representasi pengetahuan digunakan untuk menyajikan pengetahuan yang ditambah kepada pengguna.

Association rule mining merupakan metode *data mining* yang digunakan untuk mengekstraksi pola yang bermanfaat di antara data barang [Cakir and Aras, 2012]. *Association rule mining* menemukan hubungan asosiasi atau korelasi yang menarik di antara sekumpulan besar item data [Annie and Kumar, 2012]. Hasil dari hubungan asosiasi atau korelasi tersebut dapat membantu pemilik supermarket untuk mengembangkan strategi penjualan dengan memperhatikan barang yang sering dibeli bersamaan oleh konsumen.

Association rule merupakan hubungan asosiasi dari bentuk $X \Rightarrow Y$, dimana $X \subset I$, $Y \subset I$, dan $X \cap Y = \emptyset$. *Support* dari rule X dan Y disebut *antecedent* (LHS: *left hand side*) dan *consequent* (RHS: *right hand side*) dari rule [Wisaeng, 2014]. Kekuatan *association rule* dapat diukur dengan *Support* dan *Confidence*. *Support* menentukan seberapa sering rule berlaku dalam dataset, sedangkan *confidence* menentukan seberapa sering item Y muncul dalam transaksi yang mengandung item X [Wisaeng, 2014]. Definisi formal dari *Support*, *Confidence* dan *Lift ratio* dijabarkan pada persamaan (1), persamaan (2) dan persamaan (3) [Wisaeng, 2014].

$$\text{Support}(X \rightarrow Y) = \frac{\text{Probability}(X \cup Y)}{\text{Total Number of Transaction}} \dots\dots\dots (1)$$

$$\text{Confidence}(X \rightarrow Y) = \frac{\text{Probability}(X \cap Y)}{\text{Number of Transaction}(X)} \dots\dots\dots (2)$$

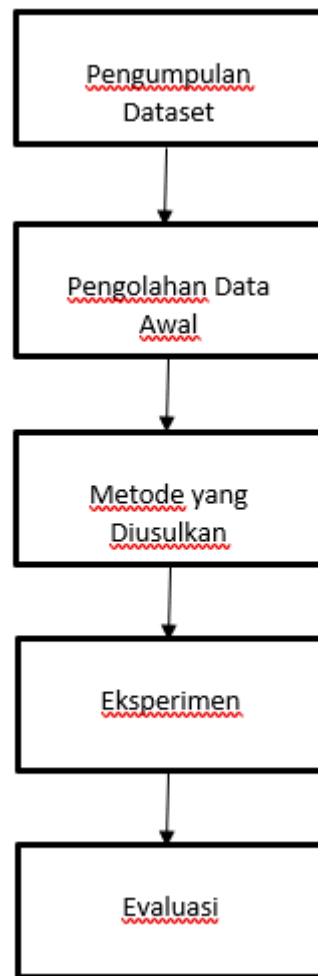
$$\text{Lift}(X \rightarrow Y) = \frac{\text{Probability}(X \cap Y)}{\text{Probability}(X) \text{Probability}(Y)} \dots\dots\dots (3)$$

Algoritma Apriori adalah salah satu algoritma paling populer dalam *data mining* untuk mempelajari konsep *association rule* yang digunakan oleh begitu banyak orang khususnya untuk operasi transaksi dan dapat juga digunakan dalam aplikasi real time dengan mengumpulkan barang yang dibeli oleh konsumen dari waktu ke waktu sehingga *frequent itemset* dapat dihasilkan [Bhandari et al., 2015].

Tujuan dari algoritma Apriori adalah untuk menemukan hubungan antara *itemset* yang berbeda [Mangla et al., 2013]. Proses utama algoritma Apriori [Han, Kamber, & Pei, 2012] yaitu: 1) Tahap Penggabungan (*The Join Step*). Pada tahap ini setiap *item* dikombinasikan dengan *item* yang lainnya sampai tidak terbentuk kombinasi lagi. 2) Tahap Pemangkasan (*The Prune Step*). Pada tahap ini, hasil dari *item* yang telah dikombinasikan tadi dipangkas dengan menggunakan *minimum support* yang ditentukan oleh pengguna.

2. Metode Penelitian

Metode penelitian yang digunakan pada penelitian ini adalah metode penelitian eksperimen. Penelitian eksperimen melibatkan perlakuan parameter/variabel yang tergantung pada penelitiannya dan menggunakan tes yang dikendalikan oleh si peneliti itu sendiri. Metode penelitian ini disajikan pada Gambar 2.

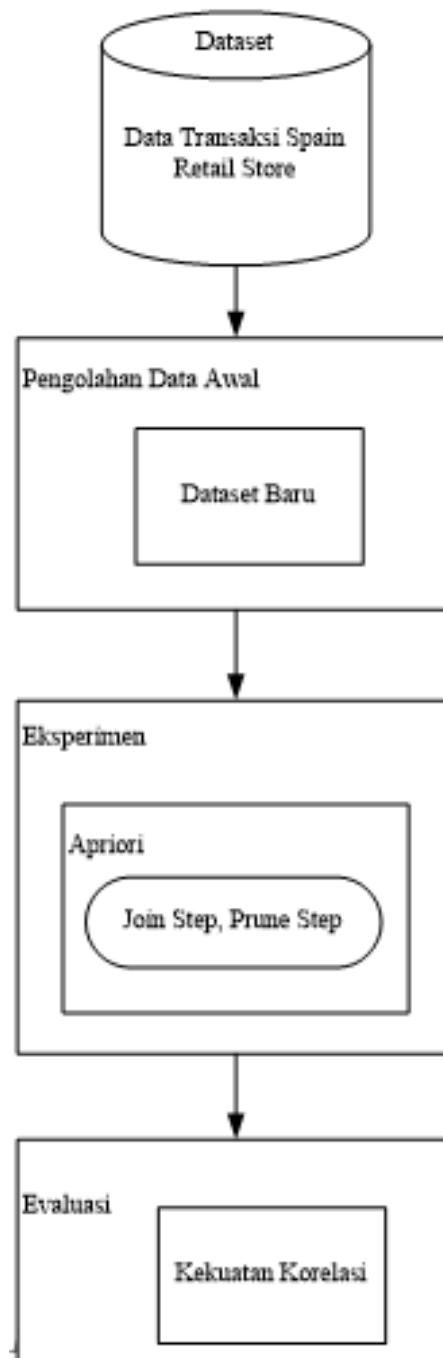


Sumber: Hasil Penelitian (2019)

Gambar 2. Metode Penelitian

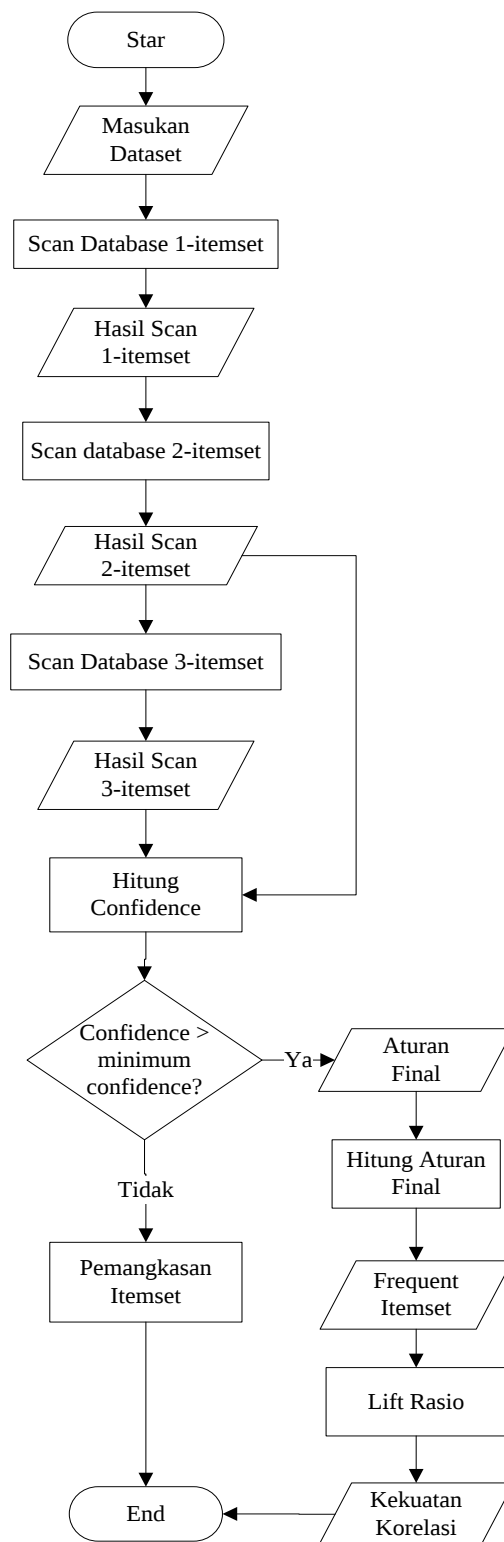
Berdasarkan Gambar 2, tahapan-tahapan dalam metode penelitian ini yaitu: **1) Pengumpulan Dataset.** Data yang digunakan pada penelitian ini merupakan data sekunder. Data sekunder yang digunakan adalah dataset Supermarket seperti yang telah digunakan oleh [Wisaeng, 2014]. Dataset Supermarket terdiri dari 12 atribut ((Receipt_id, desserts, meats, paper_goods, juices, frozen_foods, snack_foods, canned_goods, beer_wine_spirits, breads, dairys, produces) dan terdiri dari 108.131 transaksi. Nilai *minimum support* ditetapkan sebesar 10% dan nilai *minimum confidence* ditetapkan sebesar 70%. **2) Pengolahan Data Awal.** Pada dataset Supermarket terdapat 1 atribut yang perlu dihapus yaitu atribut Receipt_id dikarenakan atribut tersebut tidak diperlukan setelah dilakukan seleksi fitur. Sehingga atribut yang akan digunakan pada pengolahan dataset Supermarket yaitu 11 atribut. **3) Metode yang Di Usulkan.** Metode yang diusulkan pada penelitian ini merujuk pada penelitian terdahulu yang telah dilakukan oleh [Venkatachari, 2016] yakni menggunakan algoritma Apriori untuk menemukan korelasi *itemset* dari transaksi Supermarket. Tahap awal algoritma Apriori adalah melakukan *scanning* data 1-*itemsets* untuk mendapatkan kandidat 1-*itemsets*. Kemudian menghitung nilai *support* dari masing-masing kandidat 1-*itemset* menggunakan rumus (1). Nilai *minimum support* ditetapkan sebesar 10%. *Itemset* yang memiliki nilai *support* kurang dari 10% akan dipangkas (*The Prune Step*). Selanjutnya hasil dari pemangkasan akan dikombinasikan (*The Join Step*) menjadi 2-*itemset*. Setelah dikombinasikan akan dihitung kembali nilai *supportnya* dan 2-*itemset* yang nilai *supportnya* kurang dari 10% akan dipangkas. Hasil pemangkasan 2-*itemset* akan dikombinasikan menjadi 3-*itemset*. Langkah tersebut akan terus dilakukan hingga tidak ada lagi *itemset* yang dapat dikombinasikan. Kemudian menghitung nilai *confidence* dari masing-masing *itemset* yang terbentuk dengan menggunakan rumus (2). *Itemset* yang memiliki nilai *confidence* kurang dari 70% akan dipangkas. Hasil dari

pemangkasan tersebut akan ditetapkan sebagai aturan asosiasi atau *rules*. *Itemset* yang memiliki nilai *support* dan nilai *confidence* tertinggi ditetapkan sebagai *Frequent itemset*. Untuk mengetahui adanya manfaat dari *rules* tersebut, maka dilakukan evaluasi dengan menghitung *Lift ratio* dari masing-masing *rules* yang terbentuk dengan menggunakan rumus (3). Gambar 3 merupakan alur metode yang diusulkan sedangkan algoritma dari metode yang diusulkan digambarkan dengan *flowchart* pada Gambar 4. **4) Eksperimen.** Langkah-langkah eksperimen yang dilakukan adalah: (1) Menyiapkan dataset dan melakukan pengolahan data awal guna menghapus atribut yang tidak diperlukan, (2) menerapkan algoritma Apriori untuk menemukan *frequent itemset*, (3) melakukan evaluasi hasil dengan menghitung nilai *Lift Ratio*. **5) Evaluasi.** Metode yang diusulkan diuji untuk mengetahui kekuatan korelasinya melalui perhitungan *Lift Ratio*.



Sumber: [Venkatachari, 2016]

Gambar 3. Metode yang Diusulkan

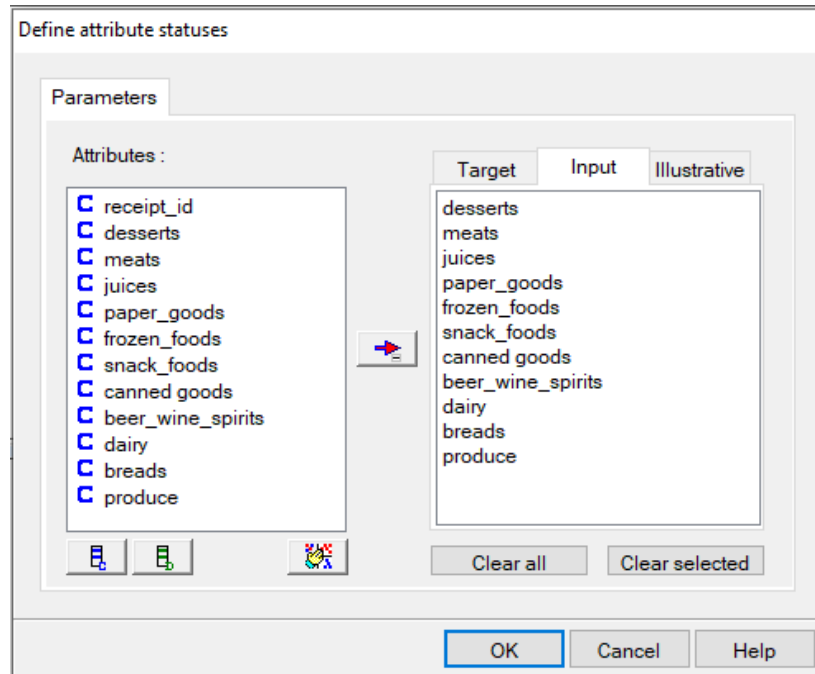


Sumber: [Venkatachari, 2016]

Gambar 4. Flowchart Algoritma Apriori

3. Hasil dan Pembahasan

Eksperimen algoritma Apriori pada dataset Supermarket dilakukan dengan menggunakan *tools* Tanagra. *Minimum support* ditetapkan sebesar 10% sedangkan *minimum confidence* ditetapkan sebesar 70%.



Sumber: Hasil Penelitian (2019)

Gambar 5. Proses Input Atribut yang Diperlukan

Gambar 5 menunjukkan proses *input* atribut ke dalam *tools* Tanagra. Ada 11 atribut yang dimasukan yaitu dessert, meats, juices, paper_goods, frozen_foods, snack_foods, canned goods, beer_wine_spirits, dairy, breads dan produce. Sedangkan atribut receipt_id tidak dimasukan karena atribut tersebut tidak diperlukan pada proses pengolahan data.

Attribute	Target	Input	Illustrative
receipt_id	-	-	-
desserts	-	yes	-
meats	-	yes	-
juices	-	yes	-
paper_goods	-	yes	-
frozen_foods	-	yes	-
snack_foods	-	yes	-
canned goods	-	yes	-
beer_wine_spirits	-	yes	-
dairy	-	yes	-
breads	-	yes	-
produce	-	yes	-

Sumber: Hasil Penelitian (2019)

Gambar 6. Atribut yang Terdapat pada Dataset Supermarket

Gambar 6 menunjukkan atribut-atribut yang terdapat pada dataset supermarket setelah dimasukan ke dalam *tools* Tanagra

	receipt_id	desserts	meats	juices	paper_goods	frozen_food	snack_food	canned_goods	beer_wine	dairy	bread
1	1	0	1	1	0	1	0	0	0	0	0
2	2	1	0	1	1	0	0	0	0	1	0
3	3	1	1	1	1	1	0	1	1	1	1
4	4	1	1	0	1	1	0	0	0	0	0
5	5	0	0	0	0	0	1	0	1	0	0
6	6	1	0	1	0	0	0	0	0	0	0
7	7	1	0	0	1	0	0	0	0	0	0
8	8	0	0	0	0	0	1	0	0	1	0
9	9	1	0	0	1	0	0	0	0	0	0
10	10	0	1	0	0	0	0	0	0	0	0
11	11	0	0	0	0	1	0	0	0	0	0
12	12	1	0	0	1	1	0	0	0	1	0
13	13	1	0	0	0	0	0	0	0	0	1
14	14	0	1	0	1	1	1	0	1	0	0
15	15	1	0	0	1	0	0	0	0	0	0
16	16	0	1	0	0	0	0	0	0	0	0
17	17	0	0	1	0	1	0	0	1	1	0
18	18	0	1	0	0	1	1	0	1	0	0
19	19	1	0	0	0	1	0	1	1	0	0
20	20	1	1	1	0	1	1	0	1	0	1
21	21	1	1	1	0	0	0	0	0	0	0

Sumber: Hasil Penelitian (2019)

Gambar 7. Dataset Supermarket yang Telah Dimasukan ke dalam *Tools* Tanagra

Setelah atribut-atribut tersebut dimasukan, dataset yang terdiri dari 108.131 transaksi dapat dilihat pada Gambar 7

Sumber: Hasil Penelitian (2019)

Gambar 8. Proses Penetapan *Minimum Support* dan *Minimum Confidence*

Proses selanjutnya yaitu penetapan *minimum support* dan *minimum confidence* seperti yang terlihat pada Gambar 8. Hasil Eksperimen Algoritma Apriori pada dataset Supermarket menggunakan *tools* Tanagra terlihat pada Gambar 9.

A priori 1

Parameters

A-Priori parameters

Support min	0.10
Confidence min	0.70
Max rule length	4
Lift filtering	1.10

Results

ITEMS

Transactions	108131
Counting items	
All items	11
Filtered items	10
Counting itemsets	
card(itemset) = 2	14
card(itemset) = 3	1
Rules	
Number of rules	3

RULES

Number of rules : 3					
N°	Antecedent	Consequent	Lift	Support (%)	Confidence (%)
1	"beer_wine_spirits=true" - "frozen_foods=true"	"snack_foods=true"	2.47766	15.489	83.719
2	"beer_wine_spirits=true" - "snack_foods=true"	"frozen_foods=true"	2.46724	15.489	81.265
3	"snack_foods=true" - "frozen_foods=true"	"beer_wine_spirits=true"	2.43118	15.489	82.304

Sumber: Hasil Penelitian (2019)

Gambar 9. Hasil Eksperimen Algoritma Apriori Menggunakan Tools Tanagra

Pada Gambar 9 terlihat bahwa data transaksi Supermarket yang terdiri dari 108.131 transaksi dengan *minimum support* sebesar 10% dan *minimum confidence* sebesar 70% telah memperoleh hasil yang menjadi *frequent itemset* yaitu kombinasi dari *itemset* beer wine spirit-frozen foods dan snack foods. Persentase nilai *support* yang dihasilkan yaitu 15,489% dan persentase nilai *confidence* yang dihasilkan yaitu 83,719%. *Lift Ratio* yang diperoleh dari kombinasi *itemsets* beer wine spirit-frozen foods dan snack foods sebesar 2,47766. Dikarenakan hasil Lift Rasio yang diperoleh lebih dari 1, maka menunjukkan adanya manfaat dari *rules* tersebut.

4. Kesimpulan

Salah satu strategi bisnis untuk meningkatkan penjualan ialah melalui promosi. Untuk menentukan promosi yang tepat sasaran, pemilik supermarket harus mengetahui selera konsumen dan kondisi pasar. Pemilik supermarket dapat mengetahui apa saja barang yang dibeli oleh konsumen dengan menganalisis kebiasaan membeli konsumen. Untuk menganalisis kebiasaan membeli konsumen dapat dilakukan dengan *Association rule mining*. Algoritma Apriori digunakan untuk melakukan eksperimen terhadap dataset Supermarket. Hasil eksperimen menggunakan *tools* Tanagra diperoleh aturan asosiasi atau *rules* dari kombinasi

itemsets beer wine spirit-frozen foods dan snack foods sebagai *Frequent itemset* dengan nilai *support* sebesar 15,612% dan nilai *confidence* sebesar 83,847%. Nilai *Lift ratio* yang diperoleh sebesar 2,47675 yang artinya terdapat manfaat dari aturan asosiasi atau *rules* tersebut. Hasil kombinasi *itemsets* tersebut menunjukkan bahwa jika konsumen membeli beer wine spirit dan Frozen Foods maka konsumen tersebut juga akan membeli Snack Foods. Hal ini berarti, pemilik Supermarket dapat mengubah tata letak toko dengan menempatkan item-item dari kombinasi itemset tersebut secara berdekatan, selain itu pemilik juga dapat membuat strategi dengan memberikan diskon untuk itemset yang memperoleh nilai *support* dan nilai *confidence* rendah.

Daftar Pustaka

- Annie LC, Kumar. 2012. Market Basket Analysis for a Supermarket based on Frequent Itemset Mining. *Int. J. Comput. Sci. Issues* 9: 257–264.
- Bhandari A, Gupta A, Das D. 2015. Improved apriori algorithm using frequent pattern tree for real time applications in data mining. In: *Procedia Computer Science.*, p 644–651.
- Cakir O, Aras ME. 2012. A Recommendation Engine by Using Association Rules. *Procedia - Soc. Behav. Sci.* 62: 452–456.
- Han J, Kamber M. 2014. Data mining: Data mining concepts and techniques. 203–207 p.
- Han, J., Kamber, M., & Pei, J. 2012. Data Mining Concept and Techniques Third Edition. Elsevier
- Larose DT. 2006. Data Mining Methods and Models. 1–322 p.
- Mangla V, Sarda C, Nadu T. 2013. Improving the efficiency of Apriori Algorithm in Data Mining. 3: 393–396.
- Moertini VS. 2009. Predicting Students Drop Out. 2nd Int. Conf. Educ. Data Min. 7: 44–56.
- Venkatachari, K. 2016. Market Basket Analysis: Understanding Indian Consumer Buying Behavior of Spain Market. *BVIMSR's Journal of Management Research*.
- Videla-Cavieres IF, Ríos SA. 2014. Extending market basket analysis with graph mining techniques: A real case. *Expert Syst. Appl.* 41: 1928–1936.
- Wisaeng K. 2014. Association rule with frequent pattern growth algorithm for frequent item sets mining. *Appl. Math. Sci.* 8: 4877–4885.