

## IMPLEMENTASI ALGORITMA FP-GROWTH UNTUK MENEMUKAN POLA PEMBELIAN KONSUMEN PADA ANALISIS KERANJANG PASAR

### *IMPLEMENTATION OF THE FP-GROWTH ALGORITHM TO FIND CONSUMER BUYING PATTERNS IN MARKET BASKET ANALYSIS*

Lusa Indah Prahartiwi

Fakultas Teknologi Informasi, Program Studi Sistem Informasi

Universitas Nusa Mandiri

Email: lusa.lip@nusamandiri.ac.id

#### **Abstrak**

Analisis keranjang pasar adalah teknik dalam data mining yang biasanya digunakan untuk memprediksi perilaku pembelian pelanggan dan memprediksi produk apa yang diminati pelanggan. Tujuan utama analisis keranjang pasar yaitu memberikan informasi kepada pengecer untuk memahami perilaku pembelian dari pembeli, yang dapat membantu pengecer dalam pengambilan keputusan yang benar. Analisis keranjang pasar dapat dilakukan melalui pengetahuan Association Rule Mining. Association Rule Mining merupakan proses untuk menemukan hubungan antara item atau item set. Selama ini banyak perusahaan retail yang dalam menata produk-produknya dengan mengelompokkan berdasarkan jenisnya saja. Misalnya, produk roti diletakkan pada etalase khusus roti dan berdekatan dengan kue-kue lainnya. Padahal, kemungkinan besar jika konsumen membeli roti, konsumen tersebut juga akan membeli susu. Penelitian ini menggunakan *dataset* transaksi Online Retail II. *Dataset* terdiri dari 216 atribut yang diambil berdasarkan customer yang berasal dari negara Jepang. Pengolahan data dengan menerapkan algoritma FP-Growth dan menggunakan tools Rapid Miner Studio. Hasil eksperimen diperoleh aturan asosiasi sebanyak 324 rules dan *frequent itemset* yang dihasilkan berupa kombinasi produk Lunch Bag Dolly Girl Design-Red Spotty Biscuit Tin. Kekuatan korelasi atau *lift ratio* yang didapat oleh kombinasi Lunch Bag Dolly Girl Design-Red Spotty Biscuit Tin adalah sebesar 3,429

**Kata Kunci:** Analisis Keranjang Pasar; Association Rule Mining; FP-Growth

#### **Abstract**

*Market basket analysis is a technique in data mining that is usually used to predict customers' buying behavior and what products the customers are interested in. The primary purpose of market basket analysis is to provide retailers with information to understand the buying behavior of buyers, which can assist retailers in making the right decisions. Market basket analysis can be done through the*

*knowledge of Association Rule Mining. Association Rule Mining is a process of finding the relationship between items or itemset. So far, many retail companies manage their products by grouping them by type only. For example, bakery products are placed on a particular display case for bread and adjacent to other cakes. Most likely, if consumers buy bread, consumers will also buy milk. This study uses the Online Retail II transaction dataset. The dataset consists of 216 attributes taken based on customers from Japan. Data processing by applying the FP-Growth algorithm and using Rapid Miner Studio tools. The experimental results obtained as many as 324 association rules, and the resulting frequent itemset was a combination of Lunch Bag Dolly Girl Design-Red Spotty Biscuit Tin products. the strength or lift ratio obtained by the combination of Lunch Bag Dolly Girl Design-Red Spotty Biscuit Tin correlation is 3,429*

**Keywords:** Market Basket Analysis; Association Rule Mining; FP-Growth

## PENDAHULUAN

Analisis keranjang pasar adalah teknik penambangan data yang dapat digunakan di berbagai bidang, seperti pemasaran, bioinformatika, bidang pendidikan, ilmu nuklir, dan lain-lain [1]. Analisis keranjang pasar adalah teknik dalam data mining yang biasanya digunakan untuk memprediksi perilaku pembelian pelanggan dan memprediksi produk apa yang diminati pelanggan [2]. Analisis keranjang pasar dapat menemukan pola transaksi yang tidak disengaja dimana pembelian beberapa produk akan mempengaruhi pembelian produk lain [2]. Tujuan utama analisis keranjang pasar yaitu memberikan informasi kepada pengecer untuk memahami perilaku pembelian dari pembeli, yang dapat membantu pengecer dalam pengambilan keputusan yang benar [1]. Analisis keranjang pasar memiliki tiga parameter yaitu *support*, *confidence*, dan *lift*. *Support* adalah proporsi kejadian B karena kejadian A.

*Confidence* adalah probabilitas kejadian B terjadi karena kejadian A secara dependen. *Lift* adalah peluang kejadian B terjadi karena kejadian A secara bebas [2].

Analisis keranjang pasar dapat dilakukan melalui pengetahuan *Association Rule Mining*. *Association Rule Mining* merupakan proses untuk menemukan hubungan antara *item* atau *itemset* [3]. Aturan asosiasi dibentuk dari adanya *frequent itemset*, maka langkah pertama dalam *Association Rule Mining* yakni penemuan *frequent itemset*.

Beberapa studi telah dilakukan oleh peneliti dalam proses analisis keranjang pasar menggunakan metode komputasi, antara lain algoritma Apriori [4] dan algoritma FP-Growth [5].

Algoritma Apriori adalah salah satu algoritma terkenal yang digunakan untuk mengekstraksi

*frequent itemset* dari *database* besar dan mendapatkan aturan asosiasi untuk menemukan pengetahuan [6]. Namun, algoritma Apriori ini membutuhkan banyak *scan database*, yang tidak efisien di era Big Data saat ini [3] sehingga menyebabkan aturan asosiasi yang terbentuk memiliki kekuatan korelasi yang rendah [7].

Algoritma FP-growth adalah algoritma terkenal lainnya untuk menemukan *frequent itemset*. FP-Growth dapat meningkatkan kinerja dalam pencarian *frequent itemset*. Tidak seperti algoritma Apriori, algoritma FP-Growth menambang *frequent itemset* melalui *frequent pattern tree* [8] dan dapat mengurangi ukuran *dataset* yang akan dicari [7].

Selama ini banyak perusahaan retail yang dalam menata produk-produknya dengan mengelompokkan berdasarkan jenisnya saja. Misalnya, produk roti diletakan pada etalase khusus roti dan berdekatan dengan kue-kue lainnya. Padahal, kemungkinan besar jika konsumen membeli roti, konsumen tersebut juga akan membeli susu. Oleh sebab itu, diperlukan sebuah algoritma untuk menemukan produk apa saja yang paling sering dibeli oleh konsumen dan memprediksi produk-produk apa saja yang akan dibeli oleh konsumen secara bersamaan dalam satu kali transaksi. Penelitian ini menggunakan algoritma FP-Growth dalam pencarian *frequent itemset* dan aturan asosiasi.

Penelitian ini bertujuan untuk menemukan *frequent itemset* dan aturan asosiasi menggunakan algoritma FP-Growth berdasarkan transaksi yang terjadi pada perusahaan retail

## LANDASAN TEORI

Penambangan asosiasi pertama kali diperkenalkan oleh Aggarwal [9]. Dalam aturan asosiasi, himpunan dengan elemen yang berbeda didefinisikan sebagai himpunan item [10]. Secara umum, penambangan aturan asosiasi terdiri dari dua tahap [11], yaitu:

1. Temukan semua *frequent itemset*: masing-masing *itemset* ini akan muncul setidaknya sesering jumlah minimum *support* yang telah ditentukan.
2. Hasilkan aturan asosiasi yang kuat dari *frequent itemset* yang ditemukan: aturan-aturan asosiasi ini harus memenuhi *minimum support* dan *minimum confidence*.

Aturan asosiasi dapat didefinisikan sebagai berikut: Misal,  $I = \{i_1, i_2, i_3, \dots\}$  menjadi satu set item, sedangkan Database menjadi satu set transaksi di mana setiap transaksi  $T$  adalah subset dari  $I$ . Aturan asosiasi adalah inferensi dari bentuk  $X \Rightarrow Y$ , dimana  $X$  dan  $Y$  merupakan himpunan bagian dari  $I$  dan  $X \cap Y = \emptyset$ . Himpunan item  $X$  disebut *antecedent* dan himpunan item  $Y$  disebut *consequent*. *Support* dan *Confidence* merupakan dua hal

yang harus dipertimbangkan dalam penambangan aturan asosiasi [9]. *Support* digunakan untuk menentukan berapa banyak aturan yang dapat diterapkan pada *dataset*, sedangkan *confidence* digunakan untuk menentukan berapa sering item Y muncul dalam transaksi yang berisi X [12]. Rumus dalam menemukan nilai *support* dan *confidence* [13] yaitu:

$$S(X \Rightarrow Y) = \frac{XUY}{N}$$

$$C(X \Rightarrow Y) = \frac{S(XUY)}{S(X)}$$

*Itemset* yang *support*-nya lebih besar atau sama dengan nilai *minimum support* dianggap sebagai *frequent itemset* [14].

### FP-Growth

Algoritma FP-Growth diusulkan oleh Jiawei Han pada tahun 2000. Sebagai algoritma penambangan asosiasi klasik, algoritma ini efisien dan skalabel untuk menambang rangkaian *frequent pattern* dengan *pattern fragmen growth*, menggunakan struktur prefix-tree yang diperluas untuk menyimpan data terkompresi dan penting. Informasi mengenai *frequent pattern* disebut *frequent-pattern tree* (FP-tree). Dibandingkan dengan algoritma lain seperti Apriori Algorithm dan TreeProjection, FP-growth memiliki kinerja yang lebih baik

terutama untuk volume data yang besar [15].

### METODE PENELITIAN

Beberapa tahapan dalam penelitian ini disajikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Gambar 1 menunjukkan tahapan-tahapan dalam penelitian. Penelitian dimulai dari mencari *dataset* yang diperlukan, lalu dilakukan pengolahan data awal untuk menyeleksi atribut-atribut yang diperlukan dalam pengolahan data. Kemudian *dataset* yang telah diseleksi tersebut diolah dengan menerapkan algoritma FP-Growth untuk menemukan *frequent itemset* dan aturan asosiasi. *Frequent itemset* dan aturan asosiasi yang telah dihasilkan selanjutnya di evaluasi untuk mengetahui kekuatan korelasinya (*lift ratio*).

### Analisis Data

Penelitian ini menggunakan data sekunder yang diambil dari UCI Machine Learning Repository. *Dataset* yang digunakan yaitu data

transaksi Online Retail II. *Dataset* terdiri dari 216 atribut yang diambil berdasarkan customer yang berasal dari negara Jepang.

Pada pengolahan data awal, satu atribut perlu dihapus, yakni atribut invoice yang mana atribut tersebut tidak diperlukan setelah dilakukan seleksi fitur pada *dataset* Online Retail II.

*Minimum support* ditetapkan sebesar 20% sedangkan *minimum confidence* ditetapkan sebesar 90%.

## IMPLEMENTASI/HASIL/PEMBAHASAN

### Perhitungan Support:

Menghitung nilai *support* dari kombinasi *item*.

$$S(X \Rightarrow Y) = \frac{X \cup Y}{N}$$

$$\begin{aligned} \text{Support}(P72 \Rightarrow P141) &= \frac{P72 \cup P141}{24} \\ &= \frac{5}{24} = 0,208 \end{aligned}$$

$$\begin{aligned} \text{Support } P72 \Rightarrow P141(\%) &= 0,208 * 100\% \\ &= \mathbf{20,8\%} \end{aligned}$$

Kombinasi dari *item* P72 dan P141 menghasilkan nilai *support* sebesar 20,8%, hasil ini telah melebihi *minimum support* yang ditetapkan.

### Perhitungan Confidence:

Menghitung nilai *Confidence* dari kombinasi *item*.

$$C(X \Rightarrow Y) = \frac{S(X \cup Y)}{S(X)}$$

$$\text{Confidence } (P72 \Rightarrow P141) = \frac{S(P72 \cup P141)}{S(P72)}$$

$$= \frac{5}{5} = 1$$

$$\begin{aligned} \text{Confidence } P72 \Rightarrow P141(\%) &= 1 * 100\% \\ &= \mathbf{100\%} \end{aligned}$$

Kombinasi dari *item* P72 dan P141 menghasilkan nilai *confidence* sebesar 100%, hasil ini telah melebihi *minimum confidence* yang ditetapkan.

### Perhitungan Lift Ratio:

Untuk mengetahui kekuatan korelasi yang terbentuk, selanjutnya dilakukan perhitungan *lift ratio*. Sebelum menghitung *lift ratio*, terlebih dahulu hitung *confidence benchmark* sebagai berikut:

*Confidence benchmark:*

P72-P141

=

$$\frac{\text{jumlah transaksi dengan item dalam consequent}}{\text{jumlah transaksi dalam database}}$$

$$= 7 / 24 = 0,2916$$

Setelah didapat nilai *confidence benchmark* dari masing-masing aturan, selanjutnya dapat dihitung nilai *ratio* nya.

*Lift ratio* (P72-P141) =

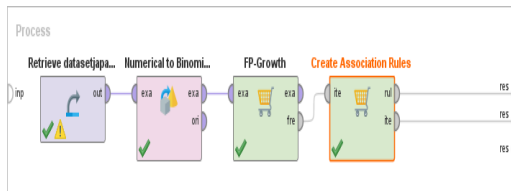
$$\frac{\text{confidence}}{\text{benchmark confidence}}$$

$$= \frac{1}{0,2916} = 3,429$$

*Lift ratio* yang dihasilkan dari kombinasi P72-P141 adalah sebesar **3,429**

Eksperimen algoritma FP-Growth pada *dataset* Online Retail II dilakukan dengan menggunakan

tools Rapid Miner Studio. Gambar 2 menunjukkan proses pengolahan *dataset* untuk menghasilkan *frequent itemset* dan aturan asosiasi.



Gambar 2. Proses Eksperimen Algoritma FP-Growth pada *Dataset* Online Retail II menggunakan *Tools* Rapid Miner Studio

Berdasarkan proses yang disajikan pada Gambar 2, diperoleh hasil *frequent itemset* dan aturan asosiasi seperti yang disajikan pada Gambar 3.

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Conv
1	P72	P141	0.208	1	1	-0.208	0.148	3.429	∞
2	P12	P141	0.167	1	1	-0.167	0.118	3.429	∞
3	P123	P141	0.167	1	1	-0.167	0.118	3.429	∞
4	P154	P141	0.167	1	1	-0.167	0.118	3.429	∞
5	P101	P141	0.125	1	1	-0.125	0.089	3.429	∞
6	P119	P141	0.125	1	1	-0.125	0.089	3.429	∞
7	P122	P141	0.125	1	1	-0.125	0.089	3.429	∞
8	P137	P141	0.125	1	1	-0.125	0.089	3.429	∞
9	P139	P141	0.125	1	1	-0.125	0.089	3.429	∞
10	P202	P141	0.125	1	1	-0.125	0.089	3.429	∞
11	P119	P22	0.125	1	1	-0.125	0.094	4	∞
12	P202	P22	0.125	1	1	-0.125	0.094	4	∞
13	P119	P72	0.125	1	1	-0.125	0.099	4.800	∞
14	P202	P72	0.125	1	1	-0.125	0.099	4.800	∞
15	P202	P115	0.125	1	1	-0.125	0.104	6	∞

Gambar 3. Hasil Eksperimen Algoritma FP-Growth menggunakan *Tools* Rapid Miner Studio.

Keterangan:

P72 : Lunch Bag Dolly Girl Design

P141 : Red Spotty Biscuit Tin

Pada Gambar 3, dapat dilihat aturan asosiasi yang terbentuk sebanyak 324 *rules*. *Frequent itemset* yang dihasilkan adalah kombinasi *itemset* Lunch Bag Dolly Girl Design-Red Spotty Biscuit Tin. Nilai *support* yang dihasilkan sebesar 0,208 atau 20,8% sedangkan nilai *confidence* yang dihasilkan sebesar 1 atau 100%. Kekuatan korelasi atau *lift ratio* yang didapat oleh kombinasi Lunch Bag Dolly Girl Design-Red Spotty Biscuit Tin adalah sebesar 3,429. Karena *lift ratio* yang dihasilkan lebih dari satu, maka menunjukkan adanya manfaat dari aturan asosiasi tersebut.

## KESIMPULAN

Analisis keranjang pasar dapat dilakukan dengan menerapkan *Association Rule Mining*. Eksperimen dilakukan menggunakan *dataset* Online Retail II yang diambil dari UCI Machine Learning Repository serta menggunakan algoritma FP-Growth. Hasil eksperimen diperoleh aturan asosiasi sebanyak 324 *rules* dan *frequent itemset* yang dihasilkan berupa kombinasi produk Lunch Bag Dolly Girl Design-Red Spotty Biscuit Tin. Nilai *support* yang dihasilkan sebesar 0,208 atau 20,8% sedangkan nilai *confidence* yang dihasilkan sebesar 1 atau 100%. Artinya, jika konsumen membeli produk Lunch Bag Dolly Girl

Design maka konsumen tersebut juga akan membeli Red Spotty Biscuit Tin. Sehingga pemilik perusahaan retail dapat meletakkan produk-produk yang dihasilkan dari aturan asosiasi tersebut saling berdekatan. Pemilik perusahaan retail juga dapat menjalankan strategi marketing dengan memberikan promosi apabila konsumen membeli beberapa produk Lunch Bag Dolly Girl Design maka akan mendapat bonus Red Spotty Biscuit Tin. Kekuatan korelasi yang didapat oleh kombinasi Lunch Bag Dolly Girl Design-Red Spotty Biscuit Tin adalah sebesar 3,429 yang menunjukkan adanya manfaat dari aturan asosiasi tersebut.

#### DAFTAR PUSTAKA

- [1] M. Kaur and S. Kang, "Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining," *Procedia Comput. Sci.*, vol. 85, no. Cms, pp. 78–85, 2016, doi: 10.1016/j.procs.2016.05.180.
- [2] S. Halim, T. Octavia, and C. Alianto, "Designing facility layout of an amusement arcade using market basket analysis," *Procedia Comput. Sci.*, vol. 161, pp. 623–629, 2019, doi: 10.1016/j.procs.2019.11.165.
- [3] S. Bagui, K. Devulapalli, and J. Coffey, "A heuristic approach for load balancing the FP-growth algorithm on MapReduce," *Array*, vol. 7, no. May, p. 100035, 2020, doi: 10.1016/j.array.2020.100035.
- [4] K. Venkatachari, "Market Basket Analysis: Understanding Indian Consumer Buying Behavior of Spain Market," *BVIMSR's J. Manag. Res.*, vol. 8, no. 1, pp. 49–55, 2016, [Online]. Available: [https://search.proquest.com/docview/1789990237?accountid=10297%0Ahttp://resolver.ebscohost.com/openurl?ctx\\_ver=Z39.88-2004&ctx\\_enc=info:ofi/enc:UTF-8&rft\\_id=info:sid/ProQ%3Aabicomplete&rft\\_val\\_fmt=info:ofi/fmt:kev:mtx:journal&rft.genre=article&rft.jtitle=](https://search.proquest.com/docview/1789990237?accountid=10297%0Ahttp://resolver.ebscohost.com/openurl?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQ%3Aabicomplete&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&rft.genre=article&rft.jtitle=)
- [5] K. Wisaeng, "Association rule with frequent pattern growth algorithm for frequent item sets mining," *Appl. Math. Sci.*, vol. 8, no. 98, pp. 4877–4885, 2014, doi: 10.12988/ams.2014.46432.
- [6] A. Bhandari, A. Gupta, and D. Das, "Improved apriori algorithm using frequent pattern tree for real time applications in data mining," *Procedia Comput. Sci.*, vol. 46, no. Icict 2014, pp. 644–651, 2015, DOI: 10.1016/j.procs.2015.02.115.
- [7] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques Third Edition*, Third Edit. Morgan Kaufmann, 2012.
- [8] C. Zhang, *Research of big-data mining visualization*

- application, vol. 259 LNEE. 2014.
- [9] M. Narvekar and S. F. Syed, "An optimized algorithm for association rule mining using FP tree," *Procedia Comput. Sci.*, vol. 45, no. C, pp. 101–110, 2015, doi: 10.1016/j.procs.2015.03.097.
- [10] X. Chen, S. Xue, T. Yang, and Q. Yang, "Research on influencing factors of line loss rate of regional distribution network based on apriori-interpretative structural model," *Energy Reports*, vol. 8, pp. 53–64, 2022, doi: 10.1016/j.egyr.2022.03.048.
- [11] J. Han and M. Kamber, *Data mining: Data mining concepts and techniques*. 2014.
- [12] A. Bhandari, A. Gupta, and D. Das, "Improvised apriori algorithm using frequent pattern tree for real time applications in data mining," in *Procedia Computer Science*, 2015, vol. 46, pp. 644–651, doi: 10.1016/j.procs.2015.02.115.
- [13] T. Kutuzova and M. Melnik, "Market basket analysis of heterogeneous data sources for recommendation system improvement," *Procedia Comput. Sci.*, vol. 136, pp. 246–254, 2018, doi: 10.1016/j.procs.2018.08.263.
- [14] M. John and H. Shaiba, "Apriori-Based Algorithm for Dubai Road Accident Analysis," *Procedia Comput. Sci.*, vol. 163, pp. 218–227, 2019, doi: 10.1016/j.procs.2019.12.103.
- [15] J. Wang and Z. Cheng, "FP-Growth based Regular Behaviors Auditing in Electric Management Information System," *Procedia Comput. Sci.*, vol. 139, pp. 275–279, 2018, doi: 10.1016/j.procs.2018.10.268.