

Analisis Algoritma Apriori dan FP-Growth Dalam Menemukan Pola Frequent Item Data Association Rule Pada Supermarket

Vita Rahayu¹, Bambang Soedijo², Sudarmawan³
Universitas Amikom Yogyakarta^{1,2,3}

vitarahayu1997@gmail.com, bambang.s@amikom.ac.id, sudarmawan@amikom.ac.id

Abstrak – Teknologi informasi yang setiap tahun nya semakin berkembang sehingga pada saat ini membuat seseorang tidak pernah lepas dari teknologi tersebut, teknologi digunakan diberbagai bidang yang salah satunya seperti usaha supermarket, pemanfaatan teknologi informasi pada supermarket dalam pengolahan data yang besar bisa menggunakan data mining dengan cara menganalisa data transaksi, dengan menganalisa data transaksi penjualan produk pada supermarket maka akan menghasilkan pola-pola kombinasi keterhubungan antar produk atau pola asosiasi. Penelitian ini penulis melakukan komparasi algoritma *apriori* dan *FP-Growth* yang dapat mengetahui pola pembelian atau pola asosiasi dengan membandingkan algoritma mana yang menghasilkan tingkat kepercayaan tertinggi dengan menggunakan data besar maka didapatkanlah *FP-Growth* yang menghasilkan tingkat ke percaya tertinggi dalam menemukan kombinasi antar item berdasarkan teknik *association rules* pada proses nya yang berdasarkan nilai *support* (penunjuang) dan *confidence* (kepercayaan) dari masing-masing proses mining.

Kata kunci: *Association rules; apriori; FP-Growth; supermarket;*

Abstract - *Information technology that is developing very rapidly at this time makes a person never separated from this technology, technology is used in various fields, one of which is the supermarket business, the use of information technology in supermarkets in large data processing can use data mining by analyzing transaction data, with Analyzing product sales transaction data at supermarkets will result in combination patterns of product connection or association patterns. In this study, the authors conducted comparisons of a priori algorithms and FP-Growth which can determine purchase patterns or association patterns by comparing which algorithms produce the highest level of trust by using big data, so FP-Growth is obtained which results in the highest level of trust in finding combinations between items based on techniques. association rules based on the value of support and confidence from each mining process.*

Keyword : *Association rules; apriori; FP-Growth; supermarket;*

1. Latar Belakang

Supermarket yang setiap harinya melakukan transaksi penjualan, dari data transaksi setiap hari tersebut, akan menyebabkan data transaksi yang diperoleh akan menjadi sangat banyak sehingga jumlah data yang sangat banyak dapat dimanfaatkan untuk menggali informasi baru guna untuk mengembangkan strategi pasar dalam meningkatkan penjualan. Pemanfaatan data pada supermarket dapat dijadikan suatu informasi yang berguna bagi pihak supermarket. Pada supermarket system yang digunakan untuk menemukan keteraturan produk dalam transaksi skala besar yaitu sistem *Point Of Scale* (POS) [1]. Pengetahuan dalam banyaknya data ini disebut dengan data mining. Data mining pada saat kemajuan teknologi saat ini sngat diperlukan terutama dalam hal mengelola data dalam jumlah yang besar sehingga dengan memanfaatkan pengolahan data menggunakan data mining dapat memudahkan suatu transaksi serta untuk proses data warehousing. Data *mining* diperlukan untuk mengelola data dalam jumlah yang besar atau dari banyak data yang ada

untuk memudahkan transaksi dan juga sebagai proses data yang disebut dengan *warehousing* sehingga dapat memberikan informasi yang akurat bagi pengguna data [2]. Dengan menganalisa data transaksi ini dengan menggunakan pola asosiasi. Pola asosiasi ini bisa didapatkan dengan menggunakan teknik *association rule* algoritma yang digunakan *apriori* dan *FP-Growth*. Apriori merupakan algoritma yang mencari pola hubungan antar satu atau lebih item dalam suatu dataset. *FP-Growth* struktur data yang digunakan dalam pohon fp yang sangat padat memberi manfaat dengan kinerja yang lebih baik dari pada algoritma lainnya seperti apriori [3]. Algoritma apriori bekerja dengan baik dalam database yang besar [4].

2. Kajian Pustaka

Banyak algoritme telah diusulkan untuk menemukan pengetahuan dari database besar ini. Aturan asosiasi pertambahan adalah salah satu ukuran terpenting. Aturan asosiasi dalam bentuk $X \Rightarrow Y$, yang di mana X disebut anteseden dan Y sebagai konsekuensinya,

dan aturan tersebut mewakili pelanggan yang membeli X lebih cenderung membeli Y [5].

Algoritma apriori dan fp growth dalam menemukan hubungan set data transaksional mendapatkan hubungan antar item, data minimal support dan minimum confidence digunakan untuk mengukur akurasi [6]. Tujuan analisis keranjang pasar adalah untuk menentukan barang mana yang sering dibeli bersama oleh pelanggan. Istilah frequent items berarti kumpulan item yang memenuhi nilai persentase yang ditentukan sebelumnya oleh pengguna. Misalnya, jika pelanggan telah membeli susu di super market, maka seberapa besar kemungkinan untuk membeli roti secara bersamaan dengan susu [7]. Evaluasi kinerja algoritma Apriori dan FP-growth. Perbedaan antara kedua algoritme tersebut adalah bahwa algoritme apriori menghasilkan kumpulan item yang sering kandidat dan juga pertumbuhan FP algoritma menghindari pembuatan kandidat dan mengembangkan pohon dengan strategi memecahkan dan menaklukkan masalah yang ekonomis dan efisien [8].

Teknik *association rule* ini diukur berdasarkan nilai *support* dan *confidence* dimana *support* merupakan sebuah nilai yang menunjukkan sebuah *itemset* akan digunakan secara bersamaan atau hanya sebagai nilai pendukung. Kemudian untuk mengukur nilai kepercayaan bahwa item tersebut sudah pasti dibeli secara bersamaan dapat menggunakan nilai *confidence* yang merupakan sebuah nilai yang menunjukkan ukuran sebuah kepastian sebuah *itemset* yang diambil atau digunakan item tersebut dilambangkan dengan X atau Y. *Association rule* digunakan untuk mencari aturan atau hubungan ketergantungan antara suatu item dengan item yang lain. Hasil yang diinginkan dari *association rule* adalah bentuk relasi item:

$$X \rightarrow Y$$

Dimana X dan Y merupakan item disjoint dan $X \cap Y = \{ \}$. Yang berarti jika item X muncul, maka terdapat kemungkinan besar bahwa item Y juga akan muncul. Akan tetapi aturan yang diberikan oleh aturan asosiasi ini tidak disebabkan oleh hubungan sebab akibat (Prasetyo, 2012) [9].

Association rule ini nantinya akan menghasilkan rules yang menentukan seberapa besar hubungan antara item X dan item Y, dan diperlukan dua ukuran untuk rules ini, yakni *support* dan *confidence*. *Support* yaitu dimana kemungkinan X dan Y muncul bersamaan yang dapat dinotasikan sebagai berikut:

$$\text{Support}(X \rightarrow Y) = \frac{\text{Jumlah transaksi yang mengandung X dan Y}}{\text{jumlah transaksi}} \quad (1)$$

Sedangkan *confidence* yaitu dimana kemungkinan munculnya item Y ketika item X juga muncul yang dapat dinotasikan sebagai berikut:

$$\text{Confidence}(X \rightarrow Y) = \frac{\text{Jumlah transaksi yang mengandung X dan Y}}{\text{jumlah transaksi yang mengandung X}} \quad (2)$$

Algoritma apriori adalah algoritma penambangan klasik yang dapat menambang aturan asosiasi dan pola sekuensial. Namun, ketika algoritma Apriori diterapkan pada penambangan pola sekuensial yang berdekatan, itu tidak efisien. Dalam penambangan log web, pola sekuensial yang berdekatan dapat lebih mewakili informasi semantik dari akses pengguna ke situs karena kontinuitas kunjungan pengguna ke halaman situs. Pola sekuensial yang berdekatan dapat digunakan tidak hanya untuk memprediksi permintaan akses pengguna berikutnya, tetapi juga untuk menyempurnakan struktur topologi situs dan menyetel halaman iklan. algoritma apriori digunakan untuk menghasilkan sejumlah besar kandidat saat menambang pola urutan yang berdekatan yang sering memindai database transaksi. algoritme peningkatan apriori dapat mengurangi waktu pemindaian database transaksi sambil mempertahankan efek penambangan penuh, yang mengurangi runtime dan meningkatkan efisiensi penambangan dibandingkan dengan algoritma apriori (Jin & Xiaohui 2017) [10].

Pendekatan MapReduce adalah algoritma paling efisien dalam kerangka data besar yang menangani data dalam jumlah besar dan memberikan hasil yang cepat. Algoritme Apriori adalah algoritma yang lebih kuat untuk menambang hubungan yang menarik antara kumpulan data di semua jenis database atau database yang sama (Pandey & Kamlesh Kuma)[11].

3. Metode Penelitian

a. Jenis, sifat dan pendekatan penelitian

- Jenis penelitian ini dilakukan untuk mengetahui metode yang menghasilkan tingkat ketepatan hasil tertinggi berdasarkan tingkat akurasi dari support, confidence dan lift rasio dalam menemukan keterhubungan antar produk dalam itemset.

- Sifat penelitian yang digunakan yaitu deskriptif yang menjelaskan tahapan-tahapan untuk

mengetahui metode yang akan menghasilkan tingkat tingkat akurasi tertinggi dalam menemukan pola asosiasi produk.

- Pendekatan penelitian yang digunakan yaitu kuantitatif yaitu hasil dari penerapan metode-metode berupa angka dan diagram yang menunjukkan tingkat akurasi dari penerapan metode-metode tersebut.

b. Sumber dan Teknik Pengumpulan Data

Pada tahap pengumpulan data, penulis melakukan wawancara dan observasi ke supermarket Al baik. Wawancara merupakan salah satu teknik yang digunakan dalam pengumpulan data, wawancara dilakukan dengan tanya jawab antara peneliti dan narasumber dengan tujuan untuk menggali data guna memperoleh informasi yang berhubungan dengan penelitian. Serta untuk mendapatkan data, data sekunder yang berupa data transaksi dari supermarket. Observasi dilakukan melalui pengamatan terhadap kejadian yang sedang berlangsung pada supermarket.

4. Implementasi Sistem dan Hasil

a. Gambaran Umum Objek Penelitian

supermarket al baik merupakan swalayan yang bergerak di bidang retail yang sudah memiliki cabang di kota tanjungpinang. Pada Swalayan al baik ini merupakan perbelanjaan modern untuk masyarakat berbelanja swalayan al baik menyediakan beragam kebutuhan rumah tangga, seperti makanan olahan, baju, kebutuhan rumah tangga seperti baju, sepatu dan lain sebagainya.

b. Tahapan Preprocessing

pada tahapan ini merupakan tahapan yang akan digunakan untuk implementasi algoritma yang berisi cara menggabungkan fungsi serta untuk mengelompokkan data yang akan mengurangi jumlah waktu yang digunakan pada saat pemindaian data. Preprocessing pada algoritma apriori akan menjadikan sebagai fungsi *agregat* untuk mengelompokkan beberapa item yang akan ditransaksikan kedalam database didalam satu transaksi. Sedangkan pada algoritma fp-growth algoritma ini mencari aturan yang sedikit, dan algoritma ini membutuhkan waktu dan memori yang lebih banyak untuk melakukan preprocessing data.

c. Hasil Rule

Algoritma Apriori

Rule Yang Dihasilkan	support	confidence
makanan ringan > mie	7,1%	16,7%

makanan ringan > telur	7,1 %	16,7%
makanan ringan > susu	21 %	50%
mie > telur	32%	69%

Dengan minimum support yang di inputkan sebesar 20% dan minimum confidence sebesar 50% maka didapatkan hasil sebagai berikut :

- 1) Jika Membeli Makanan Ringan maka akan membeli Susu dengan confidence sebesar 50%.
- 2) Jika Membeli Mie maka akan membeli Telur dengan confidence sebesar 69%.

Algoritma FP-Growth

Frequent Itemset

Rule	Jumlah transaksi	support	confidence
makanan ringan > mie	2	7,1%	16,7%
makanan ringan > telur	2	7,1%	16,7%
makanan ringan > susu	6	21%	50%
mie > telur	9	32%	69,2%

5. Kesimpulan

- a. Dari hasil analisis tahapan preprocessing akan menghasilkan pencarian itemset menjadi lebih cepat. Dan algoritma fp growth pada saat pencarian membutuhkan waktu lebih singkat, berbeda dengan algoritma apriori membutuhkan waktu lebih lama pada saat pemrosesan data
- b. Untuk rule yang dihasilkan antara algoritma apriori dan fp-growth, rule dan confidence tertinggi yang dihasilkan yaitu untuk apriori jika pelanggan membeli mie maka akan membeli telur dengan nilai confidence yang dihasilkan sebesar 69%. Algoritma fp-growth rule dan confidence tertinggi yang dihasilkan yaitu jika pelanggan membeli mie maka akan membeli telur dengan nilai confidence yang dihasilkan sebesar 69,2%.

- c. Dari hasil support dan confidence kedua algoritma ini menghasilkan nilai yang tidak terlalu jauh dengan jumlah data yang digunakan sebanyak 28 transaksi.

International Journal of Computer Applications, 79(10), 34–37.

6. Pustaka

- [1] Agarwal, R. C., Aggarwal, C. C., & Prasad, V. V. V. (2001). A Tree Projection Algorithm for Generation of Frequent Item Sets. *Journal of Parallel and Distributed Computing*, 61(3), 350–371.
- [2] Widiati, E., & Evita Dewi, K., 2014, Implementasi Association Rule Terhadap Penyusunan Layout Makanan Dan Penentuan Paket Makanan Hemat Di Rm Roso Echo Dengan Algoritma Apriori, *Jurnal Ilmiah Komputer Dan Informatika (KOMPUTA)*, 96(2), 2089–9033.
- [3] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *ACM sigmod record*, 2000, Vol. 29, No 2, pp. 1-12
- [4] Kavitha, M. M., & Tamil Selvi, S. T. (2013). Comparative Study on Apriori Algorithm and Fp Growth Algorithm with Pros and Cons. *International Journal of Computer Science Trends and Technology (IJCS T)*, 4(4), 161–164.
- [5] Dhanabhakym, M., & Punithavalli, M. (2011). A survey on Data mining algorithm for market basket analysis. *Global Journal of Computer ...*, 11(11), 1–7.
- [6] Hossain, M., Sattar, A. H. M. S., & Paul, M. K. (2019). Market basket analysis using apriori and FP growth algorithm. 2019 22nd International Conference on Computer and Information Technology, ICCIT 2019, 18–20.
- [7] Yongmei, L., & Yong, G. (2008). FP-growth algorithm for application in research of market basket analysis. *ICCC 2008 - IEEE 6th International Conference on Computational Cybernetics*, Proceedings, 269–272.
- [8] S.Mythili, M., & R. Mohamed Shanavas, A. (2013). Performance Evaluation of Apriori and FP-Growth Algorithms.
- [9] Prasetyo, Eko, 2012, "Data Mining : Konsep dan Aplikasi Menggunakan Matlab", Yogyakarta:ANDI.
- [10] Yang, J., Huang, H., & Jin, X. (2017). Mining web access sequence with improved apriori algorithm. *Proceedings - 2017 IEEE International Conference on Computational Science and Engineering and IEEE/IFIP International Conference on Embedded and Ubiquitous Computing, CSE and EUC 2017*, 1, 780–784.
- [11] Pandey, K. K., & Shukla, Di. (2018). Mining on Relationships in Big Data era using Improve Apriori Algorithm with MapReduce Approach. 2018 International Conference on Advanced Computation and Telecommunication, ICACAT 2018, 1–5.