

Analisa Data Transaksi Penjualan Barang Menggunakan Algoritma Apriori dan FP-Growth

Hariato¹, Aris Sudianto², Hariman Bahtiar²

^{1,2,3}Program Studi Teknik Informatika Fakultas Teknik Universitas Hamzanwadi
harianto.27@hamzanwadi.ac.id¹, a.sudianto@hamzanwadi.ac.id², harimaob@gmail.com³

Diterima : 25 Februari 2023

Disetujui : 29 Maret 2023

Abstract—Transaksi penjualan barang pada sebuah perusahaan terjadi setiap hari mengakibatkan semakin bertambah banyaknya catatan transaksi penjualan. Tidak banyak dari perusahaan menjadikan transaksi penjualan itu hanya sebagai arsip belaka, akibatnya perusahaan mengalami kekurangan stok barang. Tentunya hal demikian terjadi karena transaksi penjualan barang tidak dianalisa dan dipelajari polanya. Pada penelitian ini, bertujuan untuk menemukan pola pembelian barang secara bersamaan berdasarkan data bulan selama satu tahun. Ada banyak metode data mining yang digunakan untuk menganalisa pola pembelian barang secara bersamaan oleh pembeli. Diantara metode yang sering sekali digunakan adalah Apriori dan FP-Growth. Untuk menemukan pola pembelian tersebut digunakan algoritma Apriori dan FP-Growth kemudian membandingkan hasil dalam menemukan pola kombinasi yang dihasilkan. Algoritma mana yang lebih banyak menghasilkan pola dengan waktu yang sedikit. Analisis perilaku pelanggan untuk transaksi pembelian selama satu tahun, Algoritma FP-Growth lebih banyak menampilkan pola yang dihasilkan dibandingkan dengan algoritma Apriori baik pada Rules Support ≥ 0.5 paling banyak Apriori menghasilkan 5 Rules dan FP-Growth 338 Rules sedangkan pada Rules Lift Apriori 14 Rules 0.0051 detik dan FP-Growth 137 Rules 0.0068 detik.

Keywords —Data Mining, Association Rule, Apriori, FP-Growth

I. PENDAHULUAN

Market Basket Analysis (MBA), dikenal juga sebagai *Association Rule Mining* (ARM), adalah sebuah metode yang digunakan pada data mining untuk menemukan barang yang dibeli secara bersamaan [1]. Ada beberapa metode yang sering dipakai untuk melakukan analisis keranjang belanja pada *Market Basket Analysis* (MBA) diantaranya Algoritma Apriori [2][3], K-Apriori [4], ARM (Association Rule Miner)-*Predictor* [5], FP-Growth [6] dan Neural Networks [7] yang pernah dipakai oleh para peneliti dalam menganalisis *Market basket Analysis* (MBA) untuk mengetahui perilaku pelanggan dalam pembelian barang secara bersamaan, kemudian digunakan untuk membuat dan memberikan rekomendasi keputusan untuk tata letak barang, supaya mudah untuk dicari, memprediksi

kebutuhan inventaris barang pada tahun berikutnya [7] dan rekomendasi barang kepada pelanggan secara *online* [8].

Algoritma Apriori adalah algoritma yang paling lama dan cukup penting dalam *Frequent Itemsets Mining* (FIM). Walaupun sudah banyak dikembangkan algoritma serupa yang lebih efisien, seperti *FP-Growth*, LCM, dan sebagainya, Algoritma Apriori masih tetap paling banyak digunakan dan diimplementasikan dalam produk komersial untuk data mining karena dianggap sebagai algoritma yang lebih mapan [9]. Kunci utama proses Algoritma Apriori membuat beberapa tahap iterasi di dalam basis data [10]. Iterasi menghasilkan pola frekuensi yang dihitung dengan cara melakukan *scan* database untuk mendapatkan nilai *support* dari setiap item. Setelah *support* dari setiap item didapatkan, item

yang memiliki *support* di atas *minimum support* dipilih menjadi pola frekuensi tinggi dengan panjang satu atau sering disebut *1-itemset*.

K-itemset adalah istilah untuk satu set yang terdiri dari k item. Sementara iterasi kedua akan menghasilkan 2-itemset yang tiap setnya memiliki dua item [11]. Dalam penggunaannya, digunakan pemangkasan sebagai cara Algoritma Apriori dapat mengurangi jumlah kandidat yang harus dihitung *supportnya*. Pemangkasan inilah yang membuat Algoritma Apriori memiliki performa yang baik [12]. Di samping memiliki performa yang baik, Algoritma Apriori juga memiliki kelemahan [13]. Beberapa peneliti menyimpulkan pada proses scanning merupakan kelemahan yang ada di Algoritma Apriori. Hal ini wajib dilakukan pada setiap kali iterasi sehingga akan memerlukan waktu yang cukup lama dan kemampuan komputasi yang besar. Kelemahan dan kekurangan Algoritma Apriori ini sudah tidak lagi ditemukan dalam algoritma-algoritma baru yang serupa, seperti contohnya FP-Growth. Para peneliti juga mengatakan bahwa Algoritma Apriori masih perlu diteliti dan dikembangkan lagi dalam kaitannya dengan bidang data mining.

FP-Growth merupakan salah satu alternatif algoritma yang cukup baik dan efektif untuk mencari himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data yang besar [14]. FP-Growth memiliki kecepatan yang lebih baik dalam menampilkan hasil dibandingkan dengan Apriori, namun gagal dalam menghasilkan nilai *confidence* yang tinggi [15]. FP-Growth memiliki kelebihan mengenali suatu objek secara non-linier, mempermudah pemetaan masukan menjadi suatu hasil tanpa mengetahui proses sebenarnya, kuat di *parallel processing* dan kemampuan untuk mentoleransi [16]. Pada penelitian ini akan membandingkan algoritma Apriori dan algoritma FP-Growth dalam menampilkan himpunan data yang paling sering muncul (*frequent itemset*) per bulan dalam sebuah kumpulan data yang besar dan waktu yang dibutuhkan dalam menampilkan himpunan data tersebut.

II. METODOLOGI PENELITIAN

Para peneliti Rusia berhasil dalam penyatuan dua sumber data yang sangat berbeda. Penelitian itu juga dapat mengelola penyaringan dan pemilihan data yang sangat berguna untuk adaptasi data K-train dataset. Penemuan itu tentunya sangat berguna untuk menganalisis MBA dengan sumber data yang berbagai macam untuk perbaikan sistem tata letak barang. [17]. Sebagai contoh, para peneliti lebih banyak berputar studi mereka di sekitar aturan asosiasi, hubungan antar barang yang dibeli bersama oleh pelanggan [5][8].

Para peneliti mencoba mengembangkan algoritma apriori yang merupakan algoritma klasik dan cukup penting dalam FIM. Contoh dari beberapa algoritma apriori yang dikembangkan adalah ARM Predictor Algorithm. Algoritma tersebut secara otomatis melacak perubahan fakta dari data sebelumnya. Hasilnya sangat dipengaruhi oleh nilai ambang batas manual untuk skor, sehingga diperlukan untuk mengotomatiskan nilai ambang batas menjadi lebih baik dalam mencari outlier [5].

Dalam analisis keranjang pasar terhadap perilaku pelanggan dalam pembelian barang secara bersama beberapa peneliti yang dilakukan dengan menggunakan Apriori menyimpulkan bahwa Apriori memiliki kelemahan dalam kecepatan menampilkan hasil. Hal ini bisa diatasi dengan menggunakan algoritma sejenis dengan Apriori yaitu FP-Growth yang memiliki kecepatan yang lebih baik dalam menampilkan hasil dibandingkan dengan Apriori, namun gagal dalam menghasilkan nilai *confidence* yang tinggi [15].

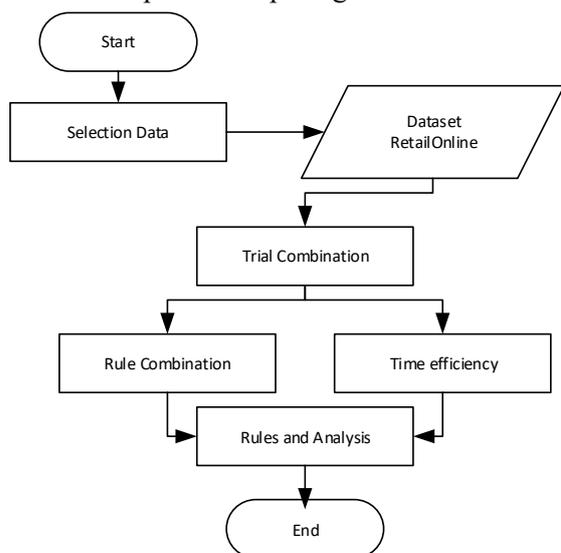
1.1. Retail Online

Ritel online atau belanja online adalah suatu bentuk perdagangan elektronik yang memungkinkan konsumen untuk langsung membeli barang atau jasa dari seorang penjual melalui internet tanpa jasa perantara. Sebuah toko online, e-toko, e-toko, toko internet, web-toko, web-store, toko online, atau toko virtual membangkitkan analogi fisik membeli produk atau jasa pada batu bata-dan-mortir pengecer atau pusat perbelanjaan. Proses ini disebut business-

to-consumer (B2C) belanja online. Dataset diambil pada retail online transaksi pada tahun 2011.

1.2. Tahapan Penelitian

Penelitian dimulai dengan mengambil dataset pada retail online kemudian memilih berdasarkan bulan dengan pemodelan percobaan menjadi 7 percobaan. Kemudian Algoritma Apriori dan Algoritma FP-Growth dibandingkan berdasarkan jumlah kombinasi itemset yang dihasilkan dan perbandingan efisiensi waktu yang digunakan dalam membuat aturan. Hasil percobaan tersebut membuktikan perbandingan pola yang dihasilkan dari setiap algoritma dan kemudian dapat digunakan sebagai bahan analisis dalam meningkatkan strategi penjualan bagi pihak lain yang ingin melakukan analisis menggunakan salah satu atau kedua dari algoritma tersebut. Alur penelitian dapat dilihat pada gambar 1.



Gambar 1 Alur Penelitian

1.3. Algoritma Apriori

Dalam penelitian ini rumus yang digunakan menghitung kecenderungan kemunculan itemset dalam jumlah transaksi yang mengacu pada *frequency*, *support count* atau *itemset*[18][14][19]. Dengan menggunakan aturan asosiasi melalui dua langkah yaitu :

1. Menemukan semua *frequent itemset*. Dari hasil proses ini sejumlah itemset dengan nilai support lebih besar atau sama dengan nilai minimum support yang diberikan. Nilai

support dari itemset {A,B} dihitung menggunakan persamaan 1.

$$s = \frac{frequency\{A,B\}}{\Sigma T} \quad (1)$$

2. Mementukan strong rule dari frequent itemset. Hasil dari proses ini adalah semua rule yang memenuhi minimum support dan minimum confidence. Nilai confidence dari suatu kandidat rule $A \rightarrow B$ dihitung dengan menggunakan persamaan 2.

$$c = \frac{frequency\{A,B\}}{frequency\{A\}} \quad (2)$$

1.4. Algoritma Fp Growth

Pada algoritma Fp Growth yang merupakan salah satu algoritma dari Teknik association rule yang digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data. Pada tahap ini penelitian kami menggunakan rumus dari analisis pola frekuensi tinggi dan pembentukan aturan asosiatif[20].

1. Analisis Pola Frekuensi Tinggi digunakan untuk mencari kombinasi item yang memenuhi syarat minimum dari nilai support dalam data. Nilai support sebuah item diperoleh dengan persamaan (1). Sedangkan nilai support dari 2 item diperoleh dari persamaan (2).

$$\begin{aligned} Support(A) &= \frac{\Sigma \text{Transaksi Mengandung } A}{\Sigma \text{Transaksi}} \times 100\% \\ & \quad (1) \end{aligned}$$

$$\begin{aligned} Support(A, B) &= P(A \cap B) \\ &= \frac{\Sigma \text{Transaksi Mengandung } A \text{ dan } B}{\Sigma \text{Transaksi}} \times 100\% \\ & \quad (2) \end{aligned}$$

2. Pembentukan aturan asosiatif digunakan setelah semua pola frekuensi tertinggi ditemukan, baru dicari aturan asosiatif yang memenuhi syarat minimum untuk confidence dengan menghitung confidence aturan asosiatif $A > B$ nilai confidence dari aturan $A > B$ diperoleh dari persamaan (3)

$$\begin{aligned} Confidence &= P\left(\frac{B}{A}\right) \\ &= \frac{\Sigma \text{Transaksi Mengandung } A \text{ dan } B}{\Sigma \text{Transaksi mengandung } A} \\ & \quad (3) \end{aligned}$$

III. HASIL DAN PEMBAHASAN

Berdasarkan metodologi penelitian, beberapa tahap penting yang akan dilakukan untuk menyelesaikan penelitian ini. Adapun tahapan tersebut terdiri dari perencanaan, pengumpulan dan pengolahan data, perhitungan Association Rule menggunakan algoritma priori dan algoritma FP-Growth, dan membandingkan hasil pola yang didapatkan oleh kedua algoritma. Data yang digunakan telah dilakukan preprocessing, dan melakukan pemilahan data berdasarkan bulan, setelah itu baru kemudian melakukan kombinasi eksperimen. Eksperimen dilakukan tujuh kali dalam data: Januari, Februari, Maret, April, Mei, Juni dan Juli.

Tabel 1 Retail Online Dataset

| No. | Bulan | Transaksi |
|-----|----------|-----------|
| 1. | Januari | 3.427 |
| 2. | Februari | 2.339 |
| 3. | Maret | 3.265 |
| 4. | April | 1.878 |
| 5. | Mei | 3.358 |
| 6. | Juni | 3.701 |
| 7. | Juli | 3.554 |

Percobaan dilakukan sebanyak 7 kali dengan menentukan nilai minimum support 0.05, nilai minimum confidence 0.08 dan nilai lift 6. Proses percobaan dilakukan dengan menggunakan Bahasa pemrograman Python 3 dengan platform Jupyter. Adapun tahap yang dilakukan adalah :

1. Import library dan dataset yang dibutuhkan bisa dilihat.
2. Cleaning Data
3. Memisahkan data sesuai dengan wilayah transaksi
4. Pengkodean data pandas
5. Buliding model dan menganalisis hasilnya

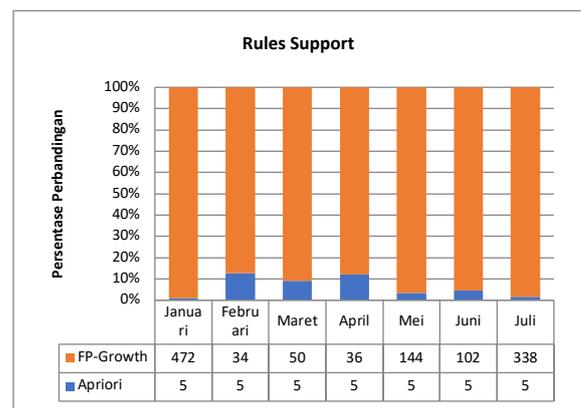
Tabel 2 Rules Lift dan efisiensi waktu

| Bulan | Jumlah Rules Lift ≥ 6 Confidence ≥ 0.08 dan Waktu Proses | |
|----------|---|-----------------------|
| | Apriori | FP-Growth |
| Januari | 6 Rules 0.0076 Detik. | 49 Rules 0.023 Detik. |
| Februari | 4 Rules 0.0050 Detik. | 8 Rules 0.0069 Detik. |

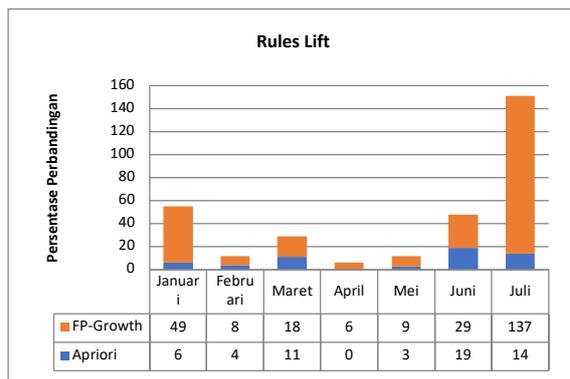
| | | |
|-------|------------------------|-------------------------|
| Maret | 11 Rules 0.0067 Detik. | 18 Rules 0.0068 Detik. |
| April | 0 Rules 0.0060 Detik. | 6 Rules 0.0050 Detik. |
| Mei | 3 Rules 0.00585 Detik. | 9 Rules 0.0089 Detik. |
| Juni | 19 Rules 0.0060 Detik. | 29 Rules 0.0069 Detik. |
| Juli | 14 Rules 0.0051 Detik. | 137 Rules 0.0068 Detik. |

Tabel 3 Rules Support dan efisiensi waktu

| Bulan | Jumlah Rules Support ≥ 0.05 dan Waktu Proses | |
|----------|---|-------------------------|
| | Apriori | FP-Growth |
| Januari | 5 Rules 0.0375 Detik. | 472 Rules 0.9009 Detik. |
| Februari | 5 Rules 0.0311 Detik. | 34 Rules 0.0338 Detik. |
| Maret | 5 Rules 0.0436 Detik. | 50 Rules 0.0363 Detik. |
| April | 5 Rules 0.0262 Detik. | 36 Rules 0.0292 Detik. |
| Mei | 5 Rules 0.0352 Detik. | 144 Rules 0.0463 Detik. |
| Juni | 5 Rules 0.1203 Detik. | 102 Rules 0.0803 Detik. |
| Juli | 5 Rules 0.0416 Detik. | 338 Rules 0.0837 Detik. |



Gambar 2 Perbandingan Rules Support



Gambar 2 Perbandingan Rules Lift

IV. KESIMPULAN

Hasil dari penelitian ini adalah menemukan perbedaan pola pembelian barang yang dilakukan oleh pembeli berdasarkan bulan dengan menggunakan algoritma Apriori dan algoritma FP-Growth. Dari percobaan yang dilakukan sebanyak 7 kali dengan menentukan nilai minimum support 0.05 Apriori mendapatkan total 57 rules sedangkan FP-Growth 256 rules, sedangkan pada percobaan dengan nilai minimum confidence 0.08 dan nilai lift 6 Apriori mendapatkan total 35 rules sedangkan FP-Growth 1176 rules. Algoritma FP-Growth pada tiap bulannya lebih banyak menampilkan pola yang dihasilkan dibandingkan dengan algoritma Apriori. Untuk meningkatkan hasil pola yang didapatkan dan untuk menyempurnakan hasil penelitian yang telah dibuat ini, maka peneliti memberikan saran sebagai berikut :

1. Menambah algoritma data mining yang digunakan untuk MBA untuk menemukan pola item set yang lebih bervariasi.
2. Diharapkan menambahkan jumlah bulan yang akan di analisis pola pembelian barang menggunakan beberapa algoritma.

DAFTAR PUSTAKA

- [1] N. Isa, N. A. Kamaruzzaman, M. A. Ramlan, N. Mohamed, and M. Puteh, "Market Basket Analysis of Customer Buying Patterns at Corm Café," vol. 7, pp. 119–123, 2018.
- [2] N. M. Kawale and D. S. Dahima, "Market Basket Analysis using Apriori Algorithm in R Language," *Int. J. Trend Sci. Res. Dev.*, vol. Volume-2, no. Issue-4, pp. 2628–2633, 2018, doi: 10.31142/ijtsrd15677.
- [3] I. Surjandari and A. C. Seruni, "Design of Product Placement Layout in Retail Shop Using Market Basket Analysis," *MAKARA Technol. Ser.*, vol. 9, no. 2, pp. 43–47, 2010, doi: 10.7454/mst.v9i2.379.
- [4] D. A. Valarmathi, "Market Basket Analysis for Mobile Showroom," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. V, no. X, pp. 1279–1284, 2017, doi: 10.22214/ijraset.2017.10185.
- [5] 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.
- [6] N. Maheshwari, "Market Basket Analysis using Association Rule Learning Find Frequent Itemsets using Minimum Process from Frequent Itemsets," *Int. J. Comput. Appl.*, pp. 20–24, 2016.
- [7] A. Mansur and T. Kuncoro, "Product Inventory Predictions at Small Medium Enterprise Using Market Basket Analysis Approach-Neural Networks," *Procedia Econ. Financ.*, vol. 4, no. Icsmed, pp. 312–320, 2012, doi: 10.1016/s2212-5671(12)00346-2.
- [8] I. A., O. U., O. T., F. T., and A. A., "Automated Market Basket Analysis System," *Int. J. Comput. Appl.*, vol. 180, no. 39, pp. 44–51, 2018, doi: 10.5120/ijca2018917043.
- [9] K. S. Kumar and R. M. Chezian, "A Survey on Association Rule Mining using Apriori Algorithm," *Int. J. Comput. Appl.*, vol. 45, no. 5, pp. 47–50, 2012, doi: 10.1109/ACCT.2015.69.
- [10] R. Gupta et al., "Sequencing and analysis of a South Asian-Indian personal genome," *BMC Genomics*, vol. 13, no. 1, p. 1, 2012, doi: 10.1186/1471-2164-13-440.
- [11] P. Devi et al., "Prevalence, risk factors and awareness of hypertension in India: A systematic review," *J. Hum. Hypertens.*, vol. 27, no. 5, pp. 281–287, 2013, doi: 10.1038/jhh.2012.33.
- [12] B. P. Patel, N. Gupta, R. K. Karn, and Y. . Rana, "Optimization of Association Rule Mining Apriori Algorithm Using ACO," *Int. J. Emerg. Technol.*, vol. 2, no. 1, pp. 87–92, 2011.
- [13] S. K. S, S. K. S, K. C. Akshay, A. Prabhu, and B. K. M, "Improved Aprori Algorithm Based on bottom up approach using Probability and Improved Aprori Algorithm Based on bottom up approach using Probability and Matrix," no. March 2012, 2016.
- [14] J. Matematika and F. Universitas, "ASOSIASI DATA MINING MENGGUNAKAN ALGORITMA FP-GROWTH UNTUK MARKET BASKET ANALYSIS Fathimah Fatihatul , Atje Setiawan , Rudi Rosadi," pp. 1–8.
- [15] A. N. Sagin and B. Ayvaz, "Determination of Association Rules with Market Basket Analysis: Application in the Retail Sector," *Southeast Eur. J. Soft Comput.*, vol. 7, no. 1, 2018, doi: 10.21533/scjournal.v7i1.149.
- [16] N. Network, P. Unit, T. Delay, P. Pembelajaran, and N. N. Proses, : "Single-Layer Neural Network : Multilayer Perceptron Neural Network : Recurrent Neural Networks : Supervised Learning Masukan Proses Pembelajaran," *Image (Rochester, N.Y.)*, pp. 1–5.
- [17] 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.
- [18] S. F. Rodiyansyah, “Algoritma Apriori untuk Analisis Keranjang Belanja pada Data Transaksi Penjualan,” *Infotech*, vol. 1, no. 1, pp. 36–39, 2015.
- [19] G. Karahan Adalı and M. E. Balaban, “Elektrik Malzemeleri Sektöründe R ve Shiny ile Pazar Sepet Analizine Yönelik Dinamik Bir Uygulama,” *Bilişim Teknol. Derg.*, pp. 93–102, 2019, doi: 10.17671/gazibtd.448245.
- [20] R. Fitria, W. Nengsih, and D. H. Qudsi, “Implementasi Algoritma FP-Growth Dalam Penentuan Pola Hubungan Kecelakaan Lalu Lintas,” *J. Sist. Inf.*, vol. 13, no. 2, p. 118, 2017, doi: 10.21609/jsi.v13i2.551.