

Analisa data transaksi penjualan barang menggunakan algoritme Apriori dan FP-Growth

Hariato¹ dan Hadryan Eddy²

- 1 Program Magister Teknik Informatika Program Pascasarjana Universitas Amikom Yogyakarta
Jl. Ring Road Utara, Condong Catur, Sleman, Yogyakarta 55283
hariato.27@students.amikom.ac.id
- 2 Program Magister Teknik Informatika Program Pascasarjana Universitas Amikom Yogyakarta
Jl. Ring Road Utara, Condong Catur, Sleman, Yogyakarta 55283
hadrian@students.amikom.ac.id

Abstrak

Transaksi penjualan barang pada sebuah perusahaan terjadi setiap hari mengakibatkan semakin bertambah banyaknya catatan transaksi penjualan. Banyak dari perusahaan menjadikan transaksi penjualan itu hanya sebagai arsip belaka sehingga pada akhirnya mengakibatkan sebuah perusahaan mengalami kekurangan stok barang. Tentunya hal demikian terjadi karena transaksi penjualan barang tidak dianalisa dan dipelajari polanya. Terdapat beberapa metode data mining yang dapat digunakan untuk menganalisa pola pembelian barang secara bersamaan oleh pembeli. Diantara metode yang sering kali digunakan adalah Apriori dan FP-Growth. Pada penelitian ini, bertujuan untuk menemukan pola pembelian barang secara bersamaan berdasarkan bulan selama satu tahun. Untuk menemukan pola pembelian tersebut digunakan algoritme Apriori dan FP-Growth kemudian membandingkan hasil dalam menemukan pola kombinasi yang dihasilkan dalam dataset. Hasil dari penelitian ini algoritme Apriori membutuhkan waktu yang lebih cepat dalam memproses dan menampilkan hasil tapi rules yang didapatkan lebih sedikit dibandingkan algoritma FP-Growth. Sedangkan algoritme FP-Growth membutuhkan waktu yang lebih lama dibandingkan algoritme Apriori tapi menghasilkan rules yang lebih banyak dibandingkan algoritme Apriori.

Kata Kunci transaksi penjualan, data mining, association rule, apriori, fp-growth

1 Pendahuluan

Terdapat sebuah teknik dalam *data mining* disebut *Market Basket Analysis* (MBA) atau dikenal juga sebagai *Association Rule Mining* (ARM) yang merupakan teknik untuk menemukan barang yang dibeli bersama berdasarkan pada perilaku pembelian pelanggan [1]. Metode yang sering digunakan untuk menganalisis keranjang belanja pada MBA adalah algoritme Apriori [2; 3], K-Apriori [4], ARM-Predictor [5], FP-Growth [6] dan Neural Networks [7]. Metode tersebut pernah dipakai oleh para peneliti dalam menganalisis MBA untuk mengetahui perilaku pelanggan dalam pembelian barang bersama kemudian digunakan untuk membuat rekomendasi keputusan untuk tata letak barang, memprediksi kebutuhan inventaris barang pada tahun berikutnya [7] dan rekomendasi barang kepada pelanggan online [8].

Apriori adalah algoritme yang paling klasik dan cukup penting dalam *Frequent Itemsets Mining* (FIM). Walaupun banyak dikembangkan algoritme serupa yang lebih efisien, seperti FP-Growth, LCM, dan sebagainya, Apriori masih tetap paling banyak digunakan dan diimplementasikan dalam produk komersial untuk data mining karena dianggap sebagai algoritme yang lebih mapan [9]. Kunci utama dalam proses algoritme Apriori adalah



© Harianto dan Hadryan Eddy;
licensed under Creative Commons License CC-BY
Jurnal Open Access
Yayasan Lentera Dua Indonesia

Jnalanaka

membuat beberapa tahap iterasi di dalam database [10]. Dijelaskan juga bahwa tiap iterasi menghasilkan pola frekuensi yang dihitung dengan cara meng-scan *database* untuk mendapatkan *support* dari setiap item. Setelah *support* dari setiap item didapat, item yang memiliki *support* di atas minimum *support* dipilih menjadi pola frekuensi tinggi dengan panjang satu atau sering disebut $1 - \text{itemset}$.

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

FP-Growth merupakan salah satu alternatif algoritme yang cukup efektif untuk mencari himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data yang besar [14]. FP-Growth memiliki kecepatan 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 input menjadi suatu hasil tanpa mengetahui proses sebenarnya, kuat di *parallel processing* dan kemampuan untuk mentoleransi. Pada penelitian ini akan membandingkan algoritme Apriori dan algoritme 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.

2 Metodologi

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 [16]. Sebagai contoh, para peneliti lebih banyak berputar studi mereka di sekitar aturan asosiasi, hubungan antar barang yang dibeli bersama oleh pelanggan [5; 8].

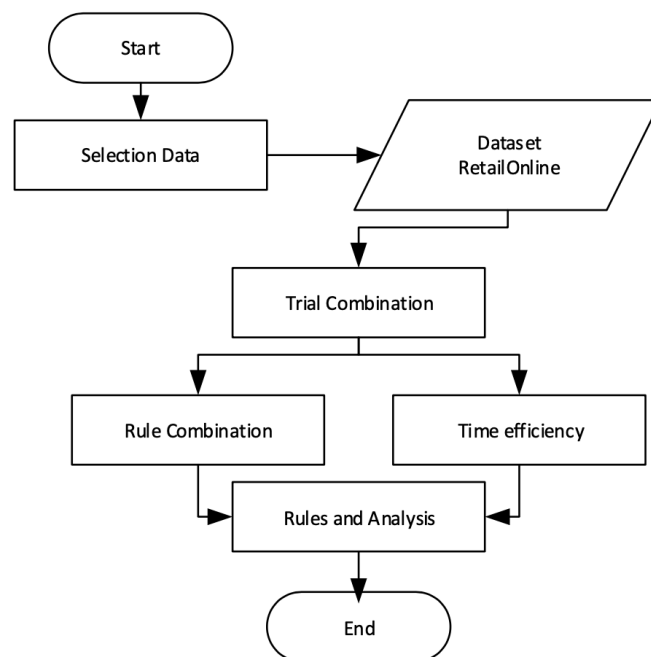
Para peneliti mencoba mengembangkan algoritme Apriori yang merupakan algoritme klasik dan cukup penting dalam FIM. Contoh dari beberapa algoritme apriori yang dikembangkan adalah *ARM Predictor Algorithm*. Algoritme tersebut secara otomatis melacak perubahan fakta dari data sebelumnya. Hasilnya sangat 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 algoritme 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].

Dataset yang digunakan dalam penelitian ini diambil pada retail online transaksi pada tahun 2011. 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.

2.1 Tahapan penelitian

Penelitian dimulai dengan mengambil dataset pada retail online kemudian memilih berdasarkan bulan dengan pemodelan percobaan menjadi 7 percobaan. Kemudian algoritme Apriori dan 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 algoritme dan kemudian dapat digunakan sebagai bahan analisis dalam meningkatkan strategi penjualan bagi pihak lain yang ingin melakukan analisis menggunakan salah satu atau kedua dari algoritme tersebut. Alur penelitian dapat dilihat pada Gambar 1 berikut.



■ **Gambar 1** Alur penelitian

Pada penelitian ini rumus dalam algoritme Apriori yang digunakan menghitung kecenderungan kemunculan *itemset* dalam jumlah transaksi yang mengacu pada *frequency*, *support count* atau *count itemset* [14; 17; 18]. Pada algoritme FP-Growth yang merupakan salah satu algoritme dari teknik *association rule* digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data. Pada bagian ini penelitian menggunakan rumus dari analisis pola frekuensi tinggi dan pembentukan aturan asosiatif [19].

3 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 algoritme Apriori dan FP-Growth serta membandingkan hasil pola yang didapatkan oleh kedua algoritme. 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

Bulan	Januari	Februari	Maret	April	Mei	Juni	Juli
Jumlah transaksi	3.427	2.339	3.265	1.878	3.358	3.701	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. Sebagai contoh pada tahap yang dilakukan akan dipaparkan langkahnya pada percobaan pertama yaitu pada bulan Januari. Adapun tahap yang dilakukan adalah :

1. mengimport library dan dataset,
2. membersihkan data,
3. memisahkan data sesuai dengan bulan transaksi,
4. pengkodean data, dan
5. membangun model dan menganalisis hasilnya.

Hasil dari analisis atau pembuatan *rules* pada item set menggunakan algoritme Apriori ataupun FP-Growth dalam menemukan pola penjualan barang dapat dilihat pada Gambar 2 dan Gambar 3.

Total Waktu Proses 0.007633686065673828 Detik.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
1	(CHILDRENS APRON SPACEBOY DESIGN)	(PLASTERS IN TIN SPACEBOY)	0.080537	0.154362	0.067114	0.833333	5.398551	0.054682	5.073826
6	(DOLLY GIRL LUNCH BOX)	(SPACEBOY LUNCH BOX)	0.100671	0.147651	0.080537	0.800000	5.418182	0.065673	4.261745
36	(SET OF 6 SPICE TINS PANTRY DESIGN)	(SET OF 3 CAKE TINS PANTRY DESIGN)	0.073826	0.140940	0.067114	0.909091	6.450216	0.056709	9.449664
39	(SET/6 RED SPOTTY PAPER CUPS)	(SET/6 RED SPOTTY PAPER PLATES)	0.080537	0.093960	0.073826	0.916667	9.755952	0.066258	10.872483
41	(DOLLY GIRL LUNCH BOX, PLASTERS IN TIN WOODLAND...)	(SPACEBOY LUNCH BOX)	0.067114	0.147651	0.060403	0.900000	6.095455	0.050493	8.523490
46	(PLASTERS IN TIN SPACEBOY, SPACEBOY LUNCH BOX)	(PLASTERS IN TIN WOODLAND ANIMALS)	0.067114	0.147651	0.060403	0.900000	6.095455	0.050493	8.523490

■ **Gambar 2** Hasil algoritme Apriori data bulan januari

Dari hasil percobaan pertama dengan data transaksi penjualan pada bulan Januari sampai percobaan ke tujuh dengan data pada masing-masing bulan dapat dilihat perbandingan hasil pola item set atau rules yang dihasilkan oleh masing-masing algoritme dan waktu efisiensi

Total Waktu Proses 0.005202054977416992 Detik.									
	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
68	(SET/6 RED SPOTTY PAPER CUPS)	(SET/6 RED SPOTTY PAPER PLATES)	0.104348	0.121739	0.095652	0.916667	7.529762	0.082949	10.539130
73	(SET/6 RED SPOTTY PAPER CUPS, SET/20 RED RETRO...	(SET/6 RED SPOTTY PAPER PLATES)	0.060870	0.121739	0.060870	1.000000	8.214286	0.053459	inf
74	(SET/6 RED SPOTTY PAPER PLATES, SET/20 RED RET...	(SET/6 RED SPOTTY PAPER CUPS)	0.069565	0.104348	0.060870	0.875000	8.385417	0.053611	7.165217
181	(WOODLAND PARTY BAG + STICKER SET)	(RETROSPOT PARTY BAG + STICKER SET)	0.060870	0.078261	0.052174	0.857143	10.952381	0.047410	6.452174
189	(BLUE POLKADOT CUP)	(RED RETROSPOT BOWL)	0.060870	0.078261	0.052174	0.857143	10.952381	0.047410	6.452174
190	(BLUE POLKADOT CUP)	(RED RETROSPOT CUP)	0.060870	0.104348	0.052174	0.857143	8.214286	0.045822	6.269565

Gambar 3 Hasil algoritme FP-Growth data bulan januari

Tabel 2 Rules dan perbandingan efisiensi waktu algoritme Apriori dan FP- growth

Bulan	Jumlah rules support ≥ 0.05 dan waktu proses		Jumlah rules lift ≥ 6 dan confidence ≥ 0.08 dan waktu proses	
	apriori(rules-detik)	FP-growth(rules-detik)	apriori(rules-detik)	FP-growth(rules-detik)
Januari	5 - 0.0375	472 - 0.9009	6 - 0.0076	49 - 0.0234
Februari	5 - 0.0311	34 - 0.0338	4 - 0.0050	8 - 0.0069
Maret	5 - 0.4364	50 - 0.0363	11 - 0.0067	18 - 0.0068
April	5 - 0.0262	36 - 0.0292	0 - 0.0060	6 - 0.0050
Mei	5 - 0.0352	144 - 0.0463	3 - 0.0058	9 - 0.0089
Juni	5 - 0.1203	102 - 0.0803	19 - 0.0060	29 - 0.0069

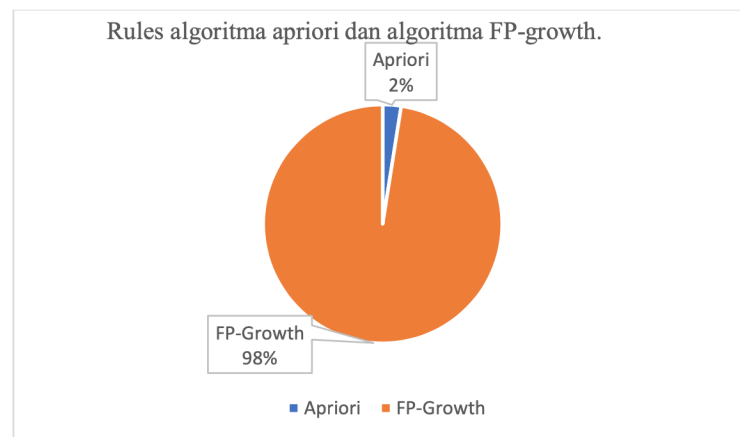
yang dibutuhkan dalam memproses dan menampilkan hasil. Untuk lebih jelas hasilnya dapat dilihat pada Table 2.

Jumlah rules dengan nilai *support* ≥ 0.05 jumlah rules yang didapatkan dari percobaan pertama bulan januari sampai percobaan ke tujuh bulan Juli untuk algoritme Apriori adalah 30 rules atau 2%. Sedangkan rules yang didapatkan oleh algoritme FP-Growth adalah 1.176 rules atau 98%. Ini menandakan bahwa algoritme FP-Growth lebih banyak menghasilkan pola pembelian atau item set dibandingkan dengan algoritme Apriori. Hasilnya bisa dilihat pada Gambar 4.

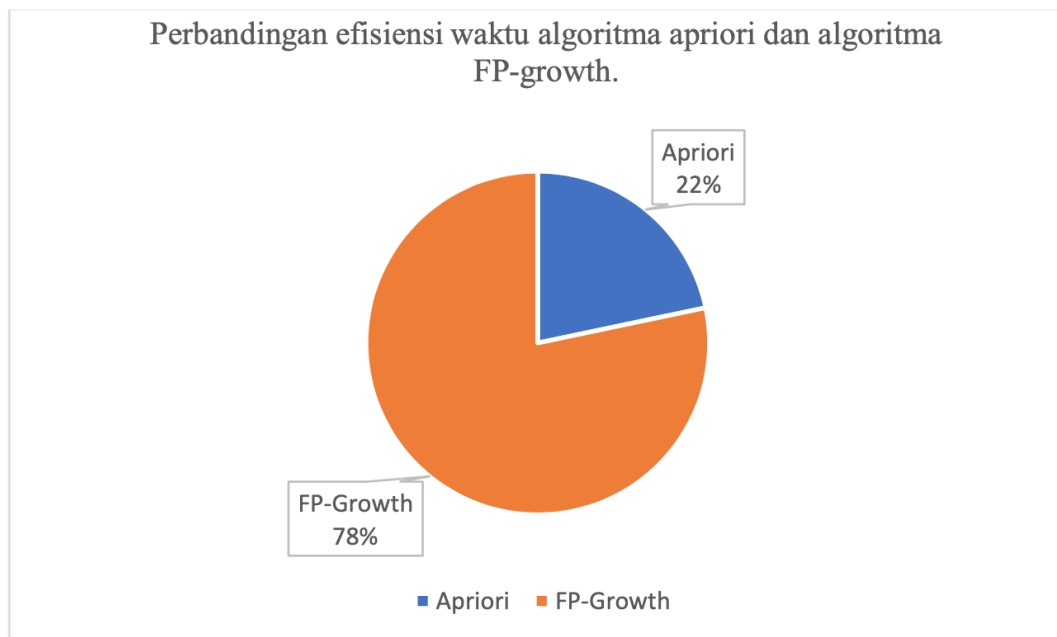
Untuk perbandingan efisiensi waktu yang dibutuhkan untuk menghasilkan rules dalam satuan detik jika dijumlahkan keseluruhan waktu pada tujuh percobaan, algoritme apriori membutuhkan waktu 0.3357 detik atau 22% dan FP-Growth 1.2110 detik atau 78%. Artinya algoritme Apriori membutuhkan waktu yang lebih sedikit untuk menampilkan hasil rules sedangkan FP-Growth membutuhkan waktu yang lebih lama 0.8752 detik dibandingkan algoritme Apriori. Perbandingan waktu efisiensi bisa dilihat pada Gambar 5.

Kemudian dilihat jumlah rules dengan nilai *lift* ≥ 6 dan nilai *confidence* ≥ 0.08 dan waktu proses pada seluruh percobaan. Jika dibandingkan jumlah rules yang diperoleh oleh kedua algoritme dari percobaan pada bulan januari sampai dengan bulan Juli, maka algoritme Apriori memperoleh 57 rules atau 18% sedangkan algoritme FP-Growth memperoleh 256 rules atau 82%. Ini artinya dengan nilai *lift* ≥ 6 dan nilai *confidence* ≥ 0.08 , algoritme FP-Growth masih lebih unggul dibandingkan algoritme Apriori dalam menghasilkan pola atau item set. Untuk lebih jelas bisa dilihat hasilnya pada Gambar 6.

Untuk perbandingan efisiensi waktu yang dibutuhkan untuk menghasilkan rules dalam

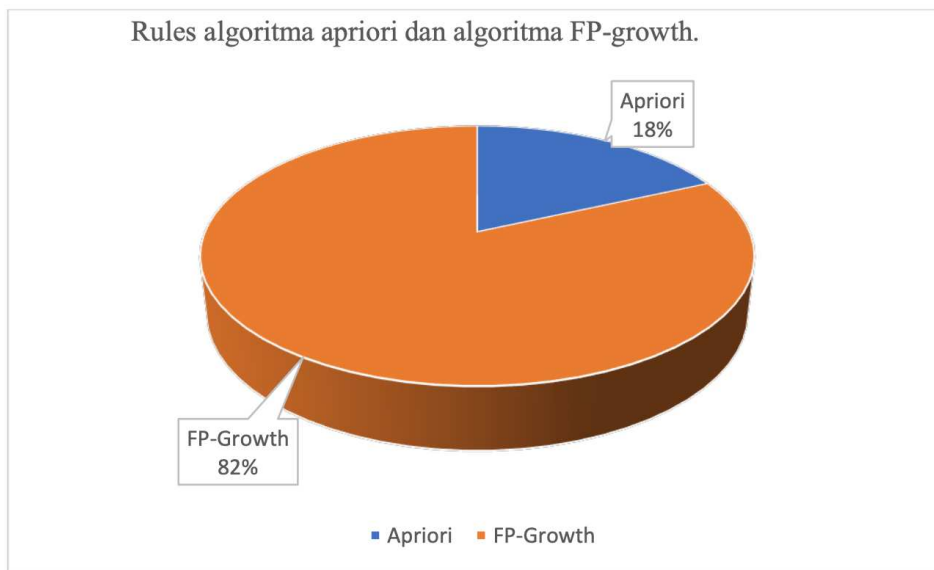


■ **Gambar 4** Perbandingan rules dengan nilai support ≥ 0.05

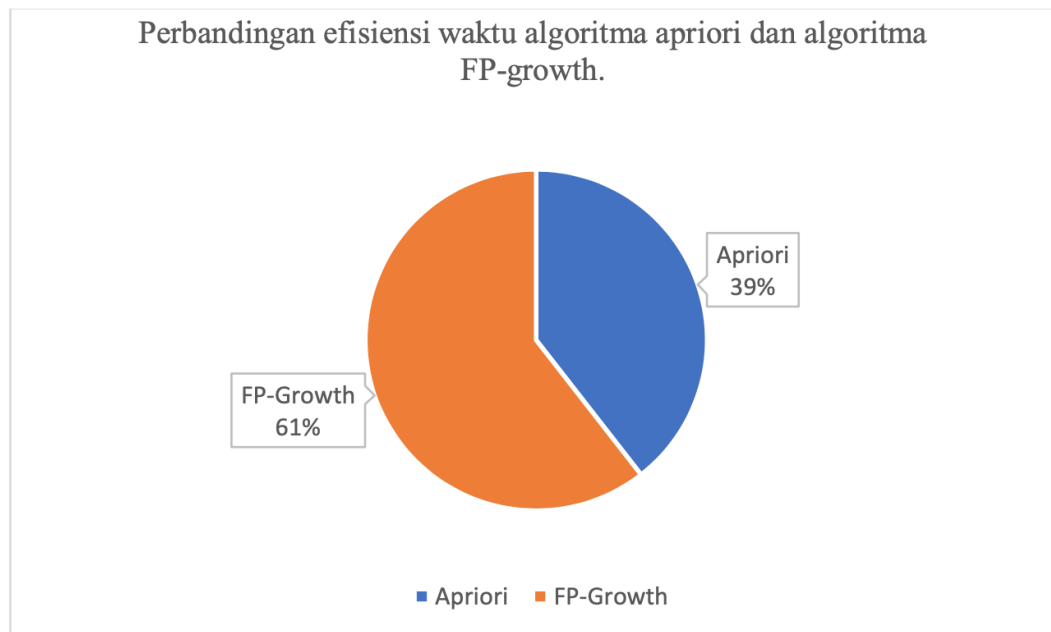


■ **Gambar 5** Perbandingan rules dengan nilai support ≥ 0.05

satuan detik jika dijumlahkan keseluruhan waktu pada tujuh percobaan, algoritme apriori membutuhkan waktu 0.0424 detik atau 39% sedangkan algoritme FP-Growth 0.0650 detik atau 61%. Artinya algoritme Apriori membutuhkan waktu yang lebih sedikit untuk menampilkan hasil rules sedangkan FP-Growth membutuhkan waktu yang lebih lama 0.0226 detik dibandingkan algoritme Apriori. Pada penelitian sebelumnya[21], algoritme Apriori membutuhkan waktu yang lebih lama dibandingkan algoritme FP-Growth. Perbandingan waktu efisiensi bias dilihat pada Gambar 7



Gambar 6 Perbandingan rules dengan nilai support ≥ 0.05



Gambar 7 Perbandingan rules dengan nilai support ≥ 0.05

4 Kesimpulan dan saran

Hasil dari penelitian ini adalah menemukan perbedaan pola pembelian barang yang dilakukan oleh pembeli berdasarkan bulan dengan menggunakan algoritme Apriori dan FP-Growth. Algoritme FP-Growth pada tiap bulannya lebih banyak menampilkan pola yang dihasilkan dibandingkan dengan Apriori. Untuk rules yang dihasilkan oleh algoritme Apriori masih

lebih sedikit dibandingkan dengan FP-Growth sedangkan pada efisiensi waktu Apriori lebih cepat atau membutuhkan waktu yang relatif sedikit dibandingkan dengan FP-Growth dalam mengolah data dan menampilkan hasil.

Untuk meningkatkan hasil pola yang didapatkan dan untuk menyempurnakan hasil penelitian yang telah dibuat ini, maka peneliti memberikan saran sebagai berikut :

1. menambah algoritme data mining yang digunakan untuk MBA untuk menemukan pola item set yang lebih bervariasi.
2. Diharapkan untuk seleksi dataset per bulan menggunakan kode, karena pada penelitian ini transaksi dipisahkan filenya berdasarkan bulan.

Pustaka

- 1 N. Isa, N. Kamaruzzaman, M. Ramlan, N. Mohamed, and M. Puteh, "Market basket analysis of customer buying patterns at corm café," *International Journal of Engineering & Technology*, vol. 7, no. 4.42, pp. 119–123, 2018.
- 2 I. Surjandari and A. Seruni, "Design of product placement layout in retail shop using market basket analysis," *Makara Journal of Technology*, vol. 9, no. 2, pp. 43–47, 2010.
- 3 N. M. Kawale and D. Snehil, "Market basket analysis using apriori algorithm in r language," *nternational Journal of Trend in Scientific Research and Development*, vol. 2, no. 4, pp. 2628–2633, 2018.
- 4 A. Valarmathi, M. Durga, and M. Fathima, "Market basket analysis for mobile showroom," *International Journal for Research in Applied Science & Engineering Technology*, vol. 5, no. X, pp. 1279–1284, 2017.
- 5 M. Kaur and S. Kang, "Market basket analysis: Identify the changing trends of market data using association rule mining," *Procedia computer science*, vol. 85, pp. 78–85, 2016.
- 6 N. Maheshwari, N. K. Pandey, and P. Agarwal, "Market basket analysis using association rule learning," *International Journal of Computer Applications*, vol. 975, p. 8887, 2016.
- 7 A. Mansur and T. Kuncoro, "Product inventory predictions at small medium enterprise using market basket analysis approach-neural networks," *Procedia Economics and Finance*, vol. 4, pp. 312–320, 2012.
- 8 A. Izang, U. Okoro, T. Olarewaju, T. Fasanu, and A. Adeyinka, "Automated market basket analysis system," *International Journal of Computer Applications*, vol. 180, no. 39.
- 9 K. S. Kumar and R. M. Chezian, "A survey on association rule mining using apriori algorithm," *International Journal of Computer Applications*, vol. 45, no. 5, pp. 47–50, 2012.
- 10 R. Gupta, A. Ratan, C. Rajesh, R. Chen, H. L. Kim, R. Burhans, W. Miller, S. Santhosh, R. V. Davuluri, A. J. Butte *et al.*, "Sequencing and analysis of a south asian-indian personal genome," *BMC genomics*, vol. 13, no. 1, p. 440, 2012.
- 11 P. Devi, M. Rao, A. Sigamani, A. Faruqui, M. Jose, R. Gupta, P. Kerkar, R. Jain, R. Joshi, N. Chidambaram *et al.*, "Prevalence, risk factors and awareness of hypertension in india: a systematic review," *Journal of human hypertension*, vol. 27, no. 5, pp. 281–287, 2013.
- 12 B. Patel, V. K. Chaudhari, R. K. Karan, and Y. Rana, "Optimization of association rule mining apriori algorithm using aco," *International Journal of Soft Computing and Engineering*, vol. 1, no. 1, pp. 24–26, 2011.
- 13 S. S. Kumar, S. S. Karanth, K. Akshay, A. Prabhu, and B. M. Kumar, "Improved aprori algorithm based on bottom up approach using probability and matrix," *International Journal of Computer Science Issues (IJCSI)*, vol. 9, no. 2, p. 242, 2012.

- 14 F. Fatihatul, A. Setiawan, and R. Rosadi, "Asosiasi data mining menggunakan algoritma fp-growth untuk market basket analysis," *Jatinangor Univ. Padjadjaran*, pp. 1–8, 2011.
- 15 A. N. Sagin and B. Ayvaz, "Determination of association rules with market basket analysis: Application in the retail sector," *Southeast Europe Journal of Soft Computing*, vol. 7, no. 1, 2018.
- 16 K. Tatiana and M. Mikhail, "Market basket analysis of heterogeneous data sources for recommendation system improvement," *Procedia Computer Science*, vol. 136, pp. 246–254, 2018.
- 17 S. F. Rodiyansyah, "Algoritma apriori untuk analisis keranjang belanja pada data transaksi penjualan," *INFOTECH journal*, vol. 1, no. 2, 2015.
- 18 G. K. ADALI and M. E. BALABAN, "A dynamic application of market basket analysis with r and shiny in the electric materials sector," *Bilişim Teknolojileri Dergisi*, vol. 12, no. 2, pp. 93–102.
- 19 R. Fitria, W. Nengsih, and D. H. Qudsi, "Implementasi algoritma fp-growth dalam penentuan pola hubungan kecelakaan lalu lintas," *Jurnal Sistem Informasi*, vol. 13, no. 2, pp. 118–124, 2017.