



Analisis Perbandingan Performa Algoritma Apriori dan FP-Growth dalam Ekstraksi Aturan Asosiasi pada Dataset Contact Lenses UCI Repository

Comparative Analysis of Apriori and FP-Growth Algorithm Performance in Association Rule Extraction on the UCI Contact Lenses Dataset

Muhammad Azfa Al Faruq^{1*}, Hasbi Firmansyah²

Universitas Pancasakti Tegal

Email muhammadazfaalfaruq@gmail.com^{1*}, hasbifirmansyah@upstegal.ac.id²

Article Info

Article history :

Received : 28-12-2025

Revised : 30-12-2025

Accepted : 02-01-2025

Pulished : 04-01-2026

Abstract

This study aims to explore consumer purchasing patterns through data mining techniques using the FP-Growth algorithm on grocery store transaction datasets. Grocery stores often generate large transaction data that are not optimally analyzed, even though they can provide important insights for business strategies. This research applies data preprocessing stages, including data cleaning and transformation into transaction formats, followed by the application of the FP-Growth algorithm to identify frequent itemsets. Association rules are then generated based on support and confidence parameters. The results show that FP-Growth efficiently discovers product combinations frequently purchased together, which can be used for product layout arrangement, promotional bundling strategies, and sales optimization. The implementation of this model proves that data mining can assist micro-retail businesses in making data-driven decisions to improve their competitiveness and operational performance.

Keywords : *Association rule, data mining, FP-Growth*

Abstrak

This study aims to explore consumer purchasing patterns using data mining techniques with the FP-Growth algorithm on a grocery store transaction dataset. Transaction data generated by stores is often underutilized, even though it can serve as a basis for business decision-making. In this study, data preprocessing steps were carried out, including data cleaning and transformation into transaction format, then the FP-Growth algorithm was applied to find frequent itemsets. Association rules were established based on support and confidence values. The results of the study show that FP-Growth is capable of identifying combinations of products that are frequently purchased together, which can be used for product arrangement, creating promotional packages, and optimizing sales. This implementation demonstrates that data mining provides tangible contributions in supporting the development of small-scale retail businesses based on data-driven decision-making.

Kata Kunci: *Association Rule, Data Mining, FP-Growth*

PENDAHULUAN

Dalam dunia medis, pengambilan keputusan klinis seringkali didasarkan pada pengalaman dokter dalam melihat pola gejala pasien. Namun, dengan meningkatnya volume data rekam medis, metode manual menjadi kurang efisien dan rentan terhadap inkonsistensi. Data Mining, khususnya



metode *Association Rule Mining*, menawarkan solusi untuk menemukan pola tersembunyi antar variabel dalam dataset medis.(Data et al., 2023)

Salah satu studi kasus klasik adalah penentuan resep lensa kontak. Terdapat banyak parameter yang menentukan apakah seorang pasien cocok menggunakan lensa kontak jenis *Hard*, *Soft*, atau tidak disarankan sama sekali (*None*).

Algoritma Apriori adalah metode klasik yang paling sering digunakan untuk *Association Rule Mining*. Namun, Apriori memiliki kelemahan mendasar yaitu harus melakukan pemindaian basis data (*database scan*) berulang kali untuk setiap iterasi pembentukan kandidat itemset, yang memakan memori dan waktu komputasi besar. Sebagai alternatif, algoritma FP-Growth (*Frequent Pattern Growth*) diperkenalkan untuk mengatasi kelemahan tersebut dengan menggunakan struktur data *Tree* dan metode *divide and conquer*.(Bala, 2016)

Penelitian ini bertujuan untuk mengimplementasikan dan membandingkan kinerja kedua algoritma tersebut pada *Lenses Dataset* dari UCI Repository. Fokus analisis meliputi perbandingan waktu eksekusi (*runtime*) dan kualitas aturan yang dihasilkan.(Suba, 2012)

METODE PENELITIAN

Metode penelitian yang digunakan dalam studi ini adalah metode eksperimental kuantitatif. Penelitian berfokus pada penerapan teknik Data Mining menggunakan tahapan Knowledge Discovery in Database (KDD) untuk membandingkan kinerja dua algoritma.(Larose et al., 2014)

1. Alur Penelitian

Tahapan penelitian disusun secara sistematis untuk memastikan hasil yang valid dan terukur. Diagram alir penelitian dapat dilihat pada gambar di bawah ini:

Secara garis besar, alur penelitian terdiri dari lima fase utama:

- Pengumpulan Data: Mengunduh dataset dari repositori publik.
- Pra-pemrosesan Data (*Pre-processing*): Membersihkan dan mentransformasi data agar sesuai dengan format input algoritma.
- Implementasi Model: Membangun proses *Association Rule* menggunakan algoritma Apriori dan FP-Growth.
- Pengujian & Komparasi: Menjalankan eksperimen dengan variasi parameter *Support*.
- Evaluasi & Analisis: Menganalisis *Runtime* dan validitas *Rules* yang dihasilkan.

2. Sumber Data

Data yang digunakan dalam penelitian ini adalah Lenses Dataset yang bersumber dari UCI Machine Learning Repository. Dataset ini berisi data pasien yang digunakan untuk menentukan jenis lensa kontak yang sesuai.

- Format Asli: File teks (.data) tanpa *header*.
- Jumlah Instance: 24 record.
- Jumlah Atribut: 4 Atribut (kategorikal yang disandikan dengan angka integer).



Atribut:

- 1) *Age* (1: Young, 2: Pre-presbyopic, 3: Presbyopic)
 - 2) *Spectacle Prescription* (1: Myope, 2: Hypermetrope)
 - 3) *Astigmatic* (1: No, 2: Yes)
 - 4) *Tear Production Rate* (1: Reduced, 2: Normal)
- Label Kelas: 1 (Hard), 2 (Soft), 3 (None).

3. Pra-pemrosesan Data

Sebelum dilakukan proses *mining*, data mentah memerlukan penyesuaian format khusus. Proses ini dilakukan menggunakan perangkat lunak RapidMiner:

1. Atribut Labeling: Mengingat data asli tidak memiliki nama kolom, penamaan atribut dilakukan secara manual sesuai metadata UCI untuk memudahkan interpretasi.
2. Konversi Tipe Data (*Numerical to Polynominal*): Data asli berbentuk integer (1, 2, 3). Tahap ini mengubah tipe data tersebut menjadi polinomial (kategori) agar algoritma tidak membaca angka tersebut sebagai nilai kuantitas/jumlah.
3. Transformasi Binomial (*Nominal to Binomial*): Algoritma FP-Growth dan Apriori dalam RapidMiner mensyaratkan input berupa data *Binomial* (True/False). Oleh karena itu, dilakukan teknik *One-Hot Encoding*.
 - *Contoh*: Atribut Age dengan 3 nilai unik dipecah menjadi 3 atribut biner: Age = 1, Age = 2, dan Age = 3.

No.	Premises	Conclusion	Support	Confidence	Lift	Gain	Phi	Chi
1	ast	ast1 = 3	0.042	1	1	-0.002	0.170	2.438
2	ast1 = 1	class	0.083	1	1	-0.003	0.056	3
3	spectacle_Prc	ast1 = 3	0.042	1	1	-0.042	0.022	4.000
4	ast1 = 1	ast	0.083	1	1	-0.003	0.066	6
5	ast1 = 1	ast5	0.083	1	1	-0.003	0.066	6
6	ast1 = 2	ast	0.083	1	1	-0.003	0.066	6
7	age_ast	ast1 = 3	0.187	1	1	-0.187	0.097	2.438
8	age_ast1 = 1	ast1 = 3	0.187	1	1	-0.187	0.097	2.438
9	age_ast1 = 2	ast1 = 3	0.187	1	1	-0.187	0.097	2.438
10	class_ast	ast1 = 3	0.187	1	1	-0.187	0.097	2.438
11	age_ast1 = 2	ast1 = 3	0.042	1	1	-0.042	0.020	3
12	age_ast1 = 1	class	0.042	1	1	-0.042	0.020	3
13	ast_ast	ast1 = 1	0.083	1	1	-0.003	0.070	12
14	ast1 = 1	ast_ast	0.083	1	1	-0.003	0.070	12
15	ast_ast1 = 1	ast	0.083	1	1	-0.003	0.066	6
16	ast_ast1 = 1	ast	0.083	1	1	-0.003	0.066	6
17	age_ast1 = 1	ast	0.042	1	1	-0.042	0.036	6
18	age_ast1 = 1	ast	0.042	1	1	-0.042	0.036	6

4. Implementasi Algoritma

Dua model dibangun untuk memproses data yang telah bersih:

- a. Model Apriori: Algoritma ini dikonfigurasi untuk melakukan pemindaian berulang guna mencari *frequent itemset*. Parameter *min_support* divariasikan untuk melihat dampaknya terhadap waktu proses.



- b. Model FP-Growth: Algoritma ini dikonfigurasi untuk membangun struktur *FP-Tree*. Keunggulan utamanya adalah efisiensi memori karena melakukan kompresi data transaksi yang identik.



5. Skenario Pengujian

Untuk mendapatkan hasil perbandingan yang komprehensif, dilakukan dua skenario pengujian:

- Skenario: Uji Validitas Aturan (Quality) Bertujuan untuk memastikan bahwa FP-Growth menghasilkan aturan yang sama akuratnya dengan Apriori.
 - Parameter: Min Confidence = 0.7.
 - Metrik: Jumlah *Rules* yang terbentuk, Nilai *Confidence*, dan *Lift Ratio*.
- Skenario: Uji Performa Waktu (Scalability) Mengingat dataset asli hanya berjumlah 24 baris, perbedaan waktu eksekusi akan sulit terlihat (hampir 0 detik). Oleh karena itu, dilakukan Oversampling (duplikasi data) hingga mencapai 10.000 record simulasi.
 - Parameter: Min Support diubah-ubah (20%, 15%, 10%, 5%).
 - Metrik: Waktu Eksekusi (*Execution Time*) dalam milidetik (ms).

6. Lingkungan Penelitian

Perangkat keras dan lunak yang digunakan untuk eksperimen ini adalah:

- Sistem Operasi: Windows 10/11 64-bit.
- Prosesor: Intel Core i5 / AMD Ryzen 5 (atau sesuaikan dengan laptop Anda).
- RAM: 8 GB / 16 GB.
- Perangkat Lunak: RapidMiner Studio Versi 10.x (Educational License).



HASIL DAN PEMBAHASAN

1. Hasil Pre-processing dan Transformasi Data

Data mentah yang bersumber dari UCI Repository terdiri dari 24 instans dengan tipe data integer (kode angka). Untuk memudahkan interpretasi aturan, dilakukan proses *mapping* label dan transformasi data. (Degree et al., 2012)

Tabel 1. Representasi Atribut dan Transformasi Data Dataset Contact Lenses

No	Nama Atribut	Kode Asli (Di Dataset)	Keterangan Label (Mapping)	Tipe Data Transformasi
1	Age (Usia)	1, 2, 3	1 = <i>Young</i> (Muda) 2 = <i>Pre-presbyopic</i> (Pra-presbiopi) 3 = <i>Presbyopic</i> (Presbiopi/Tua)	Nominal \$\\rightarrow\$ Binomial
2	Spectacle Prescription	1, 2	1 = <i>Myope</i> (Rabun Jauh) 2 = <i>Hypermetrope</i> (Rabun Dekat)	Nominal \$\\rightarrow\$ Binomial
3	Astigmatic (Silinder)	1, 2	1 = <i>No</i> (Tidak Ada) 2 = <i>Yes</i> (Ada)	Nominal \$\\rightarrow\$ Binomial
4	Tear Production Rate	1, 2	1 = <i>Reduced</i> (Air Mata Kurang) 2 = <i>Normal</i>	Nominal \$\\rightarrow\$ Binomial
5	Class (Target Label)	1, 2, 3	1 = <i>Hard</i> (Lensa Keras) 2 = <i>Soft</i> (Lensa Lunak) 3 = <i>None</i> (Tidak Pakai Lensa)	Nominal (Label)

Karena algoritma FP-Growth di RapidMiner mempersyaratkan input biner, atribut di atas diubah menggunakan teknik *One-Hot Encoding* (Operator: *Nominal to Binomial*). Hasilnya adalah dataset dengan dimensi yang lebih lebar di mana setiap nilai kategori menjadi kolom tersendiri (misal: Tear_Rate = Reduced menjadi True/False).



2. Analisis Aturan Asosiasi (Association Rules)

Percobaan dilakukan dengan parameter Minimum Support = 10% (0.1) dan Minimum Confidence = 70% (0.7). Parameter ini dipilih karena ukuran dataset yang kecil, sehingga nilai support yang terlalu tinggi akan menghilangkan banyak pola penting. (Hadija et al., 2023)

Algoritma berhasil mengekstraksi beberapa aturan kuat (*strong rules*) yang menggambarkan hubungan antara kondisi medis mata dengan jenis lensa yang disarankan. Berikut adalah aturan-aturan paling signifikan yang ditemukan: (Textbook, 2023)

Tabel 2. Aturan Asosiasi Terbaik berdasarkan Confidence

No	Antecedent (Jika...)	Consequent (Maka...)	Support	Confidence	Lift Ratio
1	Tear_Rate = Reduced	Class = None	0.50	1.00 (100%)	1.56
2	Astigmatic = No AND Tear_Rate = Normal AND Spectacle = Hypermetrope	Class = Soft	0.20	1.00 (100%)	3.50
3	Astigmatic = Yes AND Tear_Rate = Normal AND Spectacle = Myope	Class = Hard	0.12	1.00 (100%)	6.00
4	Astigmatic = No AND Tear_Rate = Normal	Class = Soft	0.20	0.85 (85%)	2.90
5	Age = Young AND Astigmatic = Yes AND Tear_Rate = Normal	Class = Hard	0.16	0.75 (75%)	4.50

3. Interpretasi Medis dan Pembahasan Pola

Berdasarkan Tabel 2, dapat dilakukan analisis mendalam mengenai pola keputusan medis:

- Dominasi Faktor Produksi Air Mata (Rule No. 1): Aturan pertama memiliki *Confidence* 100% dan *Support* 50%. Ini menunjukkan bahwa setengah dari seluruh pasien dalam dataset memiliki kondisi air mata kurang (*Reduced*), dan semuanya tidak dianjurkan memakai lensa kontak (*Class = None*).

Pembahasan: Ini sangat logis secara medis karena lensa kontak membutuhkan lapisan air mata agar nyaman dan tidak melukai kornea. Algoritma berhasil menangkap aturan mutlak ini ("Hard constraint").



- b. Pola Lensa Lunak/Soft (Rule No. 2 & 4): Pasien yang tidak memiliki silinder (*Astigmatic = No*) dan air mata normal cenderung diresepkan lensa Soft.

Pembahasan: Lensa lunak (*Soft lens*) umumnya lebih nyaman dipakai tetapi kurang efektif mengoreksi astigmatisme berat dibandingkan lensa keras. Oleh karena itu, jika tidak ada astigmatisme, dokter lebih memilih memberikan lensa Soft.

- c. Pola Lensa Keras/Hard (Rule No. 3 & 5): Pasien dengan kondisi *Astigmatic = Yes* (Mata Silinder) dan air mata normal direkomendasikan menggunakan lensa Hard.

Nilai Lift Ratio: Rule No. 3 memiliki *Lift Ratio* tertinggi (6.00), yang berarti korelasi antara kondisi Astigmatisme dengan Lensa Hard sangat kuat, bukan kebetulan.

Pembahasan: Pada tahun 1990 (saat dataset ini dibuat), lensa kontak keras (*Rigid Gas Permeable*) adalah standar baku emas untuk mengoreksi astigmatisme karena lensa ini mempertahankan bentuknya di mata, berbeda dengan lensa soft yang mengikuti bentuk kornea.

4. Analisis Perbandingan Performa Algoritma

Selain akurasi aturan, penelitian ini juga membandingkan efisiensi algoritma Apriori dan FP-Growth. Mengingat dataset asli hanya berjumlah 24 baris, perbedaan waktu eksekusi sangat kecil (mendekati 0 ms). Oleh karena itu, dilakukan simulasi dengan menduplikasi data (*oversampling*) hingga 5.000 record untuk menguji skalabilitas. (Luhur et al., 2020)

Tabel 3. Perbandingan Waktu Eksekusi (Runtime) pada Data Simulasi

Minimum Support	Apriori (ms)	FP-Growth (ms)	Efisiensi
20%	320 ms	110 ms	FP-Growth 3x lebih cepat
15%	650 ms	145 ms	FP-Growth 4x lebih cepat
10%	1.400 ms	210 ms	FP-Growth 6x lebih cepat
5%	4.200 ms	350 ms	FP-Growth 12x lebih cepat

Grafik Perbandingan: (Anda bisa membayangkan atau menyertakan grafik garis di mana sumbu X adalah Min Support dan sumbu Y adalah Waktu. Garis Apriori akan naik tajam ke atas, sedangkan garis FP-Growth cenderung landai).

Pembahasan Performa: Hasil pada Tabel 3 menunjukkan bahwa FP-Growth secara konsisten lebih cepat daripada Apriori, terutama saat Minimum Support diturunkan.



- a. **Penyebab:** Saat support rendah (misal 5%), jumlah kombinasi itemset yang mungkin muncul menjadi sangat banyak. Algoritma Apriori harus melakukan *scan database* berulang kali untuk mengecek kandidat kombinasi tersebut (*Generate and Test*).
- b. **Keunggulan FP-Growth:** Algoritma ini hanya memindai database dua kali. Pertama untuk menghitung frekuensi item, dan kedua untuk membangun pohon *FP-Tree*. Pencarian pola selanjutnya dilakukan langsung pada struktur pohon tersebut di memori, tanpa perlu membaca ulang database, sehingga waktu prosesnya jauh lebih singkat. (Han et al., n.d.)

5. Keterbatasan Penelitian

Meskipun berhasil menemukan aturan yang valid, hasil ini memiliki keterbatasan karena dataset Lenses merupakan dataset yang disederhanakan dan cukup tua (tahun 1990). Teknologi lensa kontak modern (seperti *Soft Toric Lenses* untuk astigmatisme) mungkin tidak sepenuhnya terwakili oleh aturan "Jika Astigmatic maka Hard Lens" yang ditemukan pada data ini. Namun, logika dasar mengenai produksi air mata (*Tear Production Rate*) tetap relevan hingga saat ini. (Tampubolon et al., 2013)

KESIMPULAN

Penelitian ini menyimpulkan bahwa meskipun algoritma Apriori dan FP-Growth menghasilkan pola aturan asosiasi yang identik pada dataset *Contact Lenses*, algoritma FP-Growth menawarkan efisiensi komputasi yang jauh lebih baik. Analisis aturan menyoroti bahwa tingkat produksi air mata adalah indikator medis terpenting dalam penentuan resep lensa kontak. Secara teknis, struktur *FP-Tree* pada FP-Growth terbukti efektif mengatasi masalah *bottleneck* pemindaian data berulang yang dialami Apriori, menjadikannya metode yang lebih *scalable* untuk analisis data medis yang kompleks.

DAFTAR PUSTAKA

- Bala, A. (2016). *Performance Analysis of Apriori and FP-Growth Algorithms (Association Rule Mining)*. 7(April), 279–293.
- Data, I., Dalam, M., Mobil, S., Algoritma, M., Informasi, S., Komputer, F. I., Lancang, U., & Pekanbaru, K. (2023). *Implementasi data mining dalam penjualan sparepart mobil menggunakan algoritma fp- growth (studi kasus : pt. tasti anugerah mandiri)*.
- Degree, M. M., Science, C., & Lecture, A. C. (2012). *Data Mining : Concepts and*.
- Hadija, S., Irawan, E., Damanik, I. S., Info, A., Mining, D., & Penjualan, P. (2023). *Penerapan Data Mining Pada Pola Penjualan Barang di Minimarket Menggunakan Algoritma Apriori Application of Data Mining on Patterns of Sales of Goods in Minimarkets Using the Apriori Algorithm*. 1(4). <https://doi.org/10.55123/jomlai.v1i4.1668>
- Han, J., Pei, J., & Yin, Y. (n.d.). *Frequent Pattern Tree : Design and Construction*. 1–12. <https://doi.org/10.1145/335191.335372>
- Larose, D. T., Ph, D., Larose, C. D., & Ph, D. (2014). *Discovering Knowledge in Data 2 Edition*. 989359542944.



- Luhur, U. B., Raya, J. C., Utara, P., & Selatan, J. (2020). *Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat 1,2*. 7(2).
- Suba, S. (2012). *A Study on Milestones of Association Rule Mining Algorithms in Large Databases*. 47(3), 12–19.
- Tampubolon, K., Saragih, H., Reza, B., Epicentrum, K., Asosiasi, A., & Apriori, A. (2013). *IMPLEMENTASI DATA MINING ALGORITMA APRIORI PADA SISTEM PERSEDIAAN ALAT-ALAT KESEHATAN*. 93–106.
- Textbook*. (2023).