

Evaluasi Performa Algoritma FP-Growth Berdasarkan Variasi Parameter Minimum Support dan Confidence pada Dataset Groceries

Arumeilia¹, Hasbi Firmansyah², Wahyu Arsiyani³

^{1,2,3} Universitas Pancasakti Tegal, Indonesia

e-mail: ¹arumeilia@gmail.com, ²hasbifirmansyah.ac.id,

³asriyani1409@gmail.com³

Intisari

Penelitian ini bertujuan untuk menganalisis pola keterkaitan antarproduk pada Groceries Dataset dengan menerapkan algoritma FP-Growth sebagai pendekatan untuk menemukan aturan asosiasi (association rules). Analisis dilakukan dengan memvariasikan nilai minimum support dan minimum confidence guna melihat pengaruhnya terhadap jumlah serta kualitas aturan yang dihasilkan. Hasil pengujian menunjukkan bahwa kombinasi support sebesar 0,01 dan confidence sebesar 0,4 menghasilkan 71 aturan asosiasi dengan nilai lift tertinggi mencapai 2,344. Temuan ini mengindikasikan adanya hubungan yang kuat antarproduk dalam satu transaksi, di mana whole milk muncul sebagai produk paling dominan baik sebagai antecedent maupun consequent. Nilai lift yang tinggi menunjukkan bahwa pelanggan yang membeli whole milk cenderung juga membeli produk lain seperti yogurt, curd, atau cream cheese. Hasil penelitian ini dapat dijadikan referensi bagi pelaku ritel dalam merancang strategi promosi yang lebih efektif, mengatur penempatan produk secara strategis, serta mendukung pengambilan keputusan berbasis data (data-driven decision making) pada konteks cross-selling dan pengelolaan stok barang.

Kata kunci: Aturan Asosiasi, Data Mining, FP-Growth, Pengambilan Keputusan Berbasis Data, Pola Pembelian Konsumen

Abstract

This research investigates the relationship patterns among products in the Groceries dataset by applying the FP-Growth algorithm as an approach to uncover association rules. The analysis was conducted by varying the values of minimum support and minimum confidence to observe how these parameters influence the number and quality of generated rules. The experimental findings reveal that the combination of a support value of 0.01 and a confidence value of 0.4 generated the largest number of rules, totaling 71, with the highest lift value reaching 2.344. These results indicate a strong association between several products that frequently appear together within a single transaction, where whole milk emerges as the most dominant item, both as an antecedent and as a consequent. A high lift value suggests that customers who purchase whole milk

Evaluasi Performa Algoritma FP-Growth Berdasarkan Variasi Parameter Minimum Support dan Confidence pada Dataset Groceries (Arumeilia, Hasbi Firmansyah, Wahyu Arsiyani)

are more likely to buy related items such as yogurt, curd, or cream cheese. The insights from this study can serve as a valuable reference for retailers in designing more effective product placement, improving promotional strategies, and supporting data-driven business decisions, particularly in cross-selling and inventory optimization.

Keywords: *Association Rules, Data Mining, FP-Growth, Consumer Purchase Patterns, Data-Driven Decision Making*

PENDAHULUAN

1. Latar Belakang

Kemajuan pesat dalam teknologi informasi telah menyebabkan lonjakan besar pada jumlah data yang dihasilkan di berbagai bidang, terutama pada sektor bisnis dan ritel. Setiap aktivitas transaksi pelanggan, baik melalui kanal daring maupun tatap muka, kini meninggalkan jejak data yang tersimpan di sistem basis data perusahaan. Informasi tersebut mencakup jenis produk yang dibeli, waktu transaksi, frekuensi pembelian, serta hubungan antarproduk yang muncul dalam pola pembelian konsumen. Apabila dianalisis secara sistematis, data ini dapat memberikan pemahaman mendalam bagi perusahaan untuk mengenali perilaku pelanggan, menyusun strategi promosi yang lebih tepat sasaran, menata posisi produk dengan lebih strategis, hingga membangun sistem rekomendasi yang sesuai dengan kebutuhan pasar. Akan tetapi, besarnya volume data transaksi setiap hari membuat proses analisis manual menjadi tidak praktis dan membutuhkan waktu lama. Oleh karena itu, diperlukan pendekatan analisis berbasis data mining yang mampu menelusuri serta menemukan pola tersembunyi (*hidden patterns*) dari data transaksi secara otomatis, cepat, dan efisien.(Aktavera et al., 2024)

Salah satu pendekatan penting dalam data mining yang sering diterapkan pada analisis data transaksi adalah *association rule mining*. Teknik ini berfungsi untuk mengungkap keterkaitan antaritem yang kerap muncul bersamaan dalam sejumlah transaksi. Dalam konteks bisnis ritel, metode ini membantu perusahaan memahami kecenderungan pembelian pelanggan sehingga dapat dimanfaatkan untuk menyusun strategi promosi, memberikan rekomendasi produk, serta menata posisi barang agar lebih efektif. Di antara berbagai algoritma yang digunakan dalam pendekatan ini, Apriori merupakan salah satu yang paling populer. Algoritma ini bekerja dengan mencari kombinasi item yang sering muncul secara bersamaan, menggunakan ukuran *support* dan *confidence* sebagai indikator kekuatan hubungan antarproduk. Namun demikian, meskipun hasilnya akurat, Apriori memiliki keterbatasan dari sisi efisiensi karena memerlukan proses pembangkitan kandidat (*candidate generation*) dan pemindaian basis data secara berulang di setiap iterasi. Prosedur tersebut menyebabkan waktu komputasi meningkat secara signifikan, khususnya saat diterapkan pada dataset berskala besar dan kompleks.(Widjaja, 2024)

Untuk mengatasi kelemahan efisiensi yang dimiliki algoritma Apriori, Han dan rekan-rekannya memperkenalkan metode *FP-Growth* (*Frequent Pattern Growth*). Algoritma ini menggunakan struktur data khusus berbentuk pohon yang disebut *FP-Tree* (*Frequent Pattern Tree*), yang berfungsi untuk merepresentasikan hubungan antaritem dalam data transaksi secara ringkas. Melalui pendekatan ini, proses pencarian pola frekuensi tinggi dapat dilakukan tanpa harus membangkitkan seluruh kandidat kombinasi item seperti yang terjadi pada Apriori, sehingga analisis menjadi jauh lebih cepat dan efisien. Secara umum, *FP-Growth* membangun *FP-Tree* berdasarkan data transaksi, kemudian menelusuri pohon tersebut

Evaluasi Performa Algoritma FP-Growth Berdasarkan Variasi Parameter Minimum Support dan Confidence pada Dataset Groceries (Arumeilia, Hasbi Firmansyah, Wahyu Arsiyani)

untuk menemukan kombinasi item yang memenuhi ambang support tertentu. Setelah itu, aturan asosiasi dibentuk berdasarkan nilai confidence untuk menilai seberapa besar kemungkinan keterkaitan antarproduk. Keunggulan utama algoritma ini terletak pada kecepatan eksekusi, efisiensi penggunaan memori, dan kemampuannya dalam menangani dataset berukuran besar. Karena kelebihan tersebut, *FP-Growth* kini banyak digunakan dalam penelitian akademik maupun aplikasi bisnis ritel untuk mengidentifikasi pola pembelian yang relevan dan bernilai strategis. (Chopvitayakun, n.d.)

Salah satu kumpulan data yang kerap dijadikan acuan dalam penelitian *association rule mining* adalah **Groceries Dataset**, yang berisi data transaksi pembelian pelanggan pada sektor ritel. Dataset ini mencatat ribuan transaksi, di mana setiap entri menggambarkan sekumpulan produk yang dibeli secara bersamaan oleh konsumen, misalnya *whole milk*, *yogurt*, *other vegetables*, serta berbagai kebutuhan rumah tangga lainnya. Bentuk data semacam ini sangat ideal untuk analisis *association rule*, karena setiap transaksi dapat dipandang sebagai satu himpunan item yang mencerminkan pola kebiasaan belanja konsumen. Dengan struktur tersebut, peneliti dapat menelusuri keterkaitan antarproduk dan mengidentifikasi kombinasi item yang sering muncul bersama dalam transaksi nyata.

Dataset Groceries memiliki ciri khas berupa struktur data yang tergolong *sparse*, yakni mengandung banyak kombinasi item unik yang muncul dengan frekuensi relatif rendah. Karakteristik ini menghadirkan tantangan tersendiri dalam proses analisis karena pola keterkaitan antarproduk menjadi lebih sulit diidentifikasi. Namun, justru kondisi tersebut membuat dataset ini sangat relevan untuk menguji kinerja algoritma *FP-Growth*, terutama dalam hal kecepatan pemrosesan dan kemampuannya menghasilkan aturan asosiasi yang bermakna tanpa memunculkan terlalu banyak aturan berulang (*redundant rules*). Selain itu, Groceries Dataset juga sering dijadikan sebagai benchmark dalam berbagai penelitian terdahulu, karena strukturnya memungkinkan perbandingan performa antara beberapa algoritma penambangan pola frekuensi tinggi. Dengan demikian, pemanfaatan dataset ini dalam penelitian diharapkan dapat memberikan gambaran yang komprehensif mengenai efektivitas algoritma *FP-Growth* dalam menganalisis data transaksi ritel yang besar, kompleks, dan berdimensi tinggi. (Idris et al., 2022)

Penelitian ini dilakukan untuk menilai kinerja algoritma *FP-Growth* dalam proses penambangan aturan asosiasi pada *Groceries Dataset* dengan menerapkan variasi nilai *minimum support* dan *minimum confidence*. Kedua parameter tersebut memiliki peran krusial dalam menentukan seberapa banyak serta seberapa kuat hubungan antaritem yang dapat ditemukan dari hasil analisis. Nilai *minimum support* berfungsi untuk menyaring kombinasi item yang sering muncul dalam transaksi, sementara *minimum confidence* digunakan untuk menilai tingkat keyakinan keterkaitan antarproduk dalam suatu aturan. Perubahan nilai pada kedua parameter ini dapat berdampak signifikan terhadap hasil analisis, baik dalam jumlah aturan yang terbentuk maupun dalam kekuatan hubungan antaritem. Oleh karena itu, perlu dilakukan evaluasi menyeluruh untuk menentukan kombinasi nilai

parameter yang paling optimal agar aturan *asosiasi* yang dihasilkan tidak hanya relevan secara konseptual, tetapi juga efisien dari sisi komputasi. (Ranjan & Sharma, n.d.)

Selain mengevaluasi pengaruh variasi nilai support dan *confidence*, penelitian ini juga menganalisis performa algoritma *FP-Growth* berdasarkan beberapa kombinasi parameter yang meliputi jumlah aturan yang terbentuk, waktu komputasi, serta nilai *lift* sebagai ukuran kekuatan hubungan antarproduk. Melalui pendekatan analitis ini, penelitian diharapkan dapat memberikan pemahaman *empiris* mengenai bagaimana variasi parameter tersebut memengaruhi kinerja algoritma secara keseluruhan. Hasil dari analisis ini dapat dijadikan dasar dalam menentukan konfigurasi parameter yang paling tepat untuk penerapan algoritma *FP-Growth* pada data transaksi ritel berskala besar. Dengan demikian, fokus penelitian ini tidak hanya pada implementasi teknis algoritma, tetapi juga pada evaluasi menyeluruh terhadap efektivitasnya dalam menghasilkan aturan *asosiasi* yang akurat, relevan, dan efisien dari sisi komputasi.

Berbagai penelitian sebelumnya telah membahas penerapan algoritma *FP-Growth* baik dalam analisis pola pembelian konsumen maupun dalam membandingkan kinerjanya dengan algoritma lain seperti *Apriori* dan *Eclat*. Meskipun demikian, sebagian besar studi tersebut lebih berfokus pada implementasi algoritma tanpa menelaah secara mendalam pengaruh variasi nilai minimum support dan *minimum confidence* terhadap performa analisis yang dihasilkan. Padahal, kedua parameter tersebut memiliki peran yang sangat penting dalam menentukan tingkat relevansi serta kekuatan asosiasi antaritem pada hasil akhir. Berdasarkan celah penelitian tersebut, studi ini menghadirkan unsur kebaruan pada aspek evaluatif, yaitu dengan melakukan analisis sistematis terhadap pengaruh variasi parameter tersebut terhadap kinerja algoritma *FP-Growth* menggunakan *Groceries Dataset* sebagai bahan uji. (Hairani & Guterres, 2024)

Selain memberikan kontribusi teoretis dalam memperluas pemahaman terhadap karakteristik dan kinerja algoritma *FP-Growth*, penelitian ini juga memiliki nilai aplikatif yang tinggi bagi dunia industri, terutama pada sektor ritel modern. Hasil analisis yang diperoleh dapat dimanfaatkan sebagai panduan dalam menentukan konfigurasi parameter yang optimal untuk pengolahan data transaksi berskala besar, sehingga mampu menghasilkan aturan asosiasi yang efisien, relevan, dan bernilai strategis. Dengan demikian, penelitian ini tidak hanya memperkaya literatur dalam bidang data mining dan *association rule learning*, tetapi juga memberikan kontribusi nyata dalam mendukung proses pengambilan keputusan berbasis data (*data-driven decision making*) yang menjadi fondasi penting dalam pengelolaan bisnis ritel di era digital saat ini. (Jang et al., 2021)

2. Kajian Teoritis
 - a. *Data Mining*

Data mining merupakan salah satu tahap kunci dalam proses *Knowledge Discovery in Database (KDD)* yang bertujuan untuk mengekstraksi informasi bermakna atau pola tersembunyi dari kumpulan data berukuran besar. Menurut Han dan Kamber (2012), data mining dapat diartikan sebagai proses pencarian pola atau hubungan yang sebelumnya belum diketahui, namun memiliki validitas dan nilai guna bagi pengambilan keputusan. Teknik ini saat ini banyak diterapkan di berbagai bidang, seperti keuangan, pemasaran, kesehatan, dan industri ritel, untuk mengubah data mentah menjadi pengetahuan yang bernilai strategis. Tujuan utamanya adalah membantu proses pengambilan keputusan berbasis analisis data kompleks dengan menggunakan pendekatan statistik, matematis, serta algoritma pembelajaran mesin (*machine learning*). (P et al., 2021)

Salah satu cabang penting dalam *data mining* adalah *association rule mining*, yaitu metode untuk menemukan hubungan antaritem dalam kumpulan transaksi. Pendekatan ini memungkinkan perusahaan mengidentifikasi pola perilaku pelanggan dan mengoptimalkan strategi bisnis berdasarkan pola tersebut. (No et al., 2023)

b. Association Rule Mining

Konsep *association rule mining* pertama kali diperkenalkan oleh Agrawal et al. (1993) dan sejak itu menjadi salah satu metode paling populer dalam *market basket analysis* untuk mengidentifikasi pola kebiasaan konsumen. Inti dari metode ini adalah menemukan hubungan antaritem yang sering muncul secara bersamaan dalam sejumlah transaksi. Hasil analisis tersebut dinyatakan dalam bentuk aturan *asosiasi (association rules)* dengan pola umum $X \rightarrow Y$, yang mengindikasikan bahwa ketika pelanggan membeli item X, terdapat peluang besar mereka juga membeli item Y.

Dalam proses penerapannya, digunakan beberapa ukuran utama untuk menilai kekuatan hubungan antaritem. Pertama, *support* menunjukkan seberapa sering kombinasi item X dan Y muncul bersama di dalam dataset. Kedua, *confidence* menggambarkan tingkat keyakinan bahwa pembelian item X akan diikuti oleh pembelian item Y. Selain dua ukuran tersebut, terdapat pula *lift*, yaitu rasio yang digunakan untuk mengukur kekuatan hubungan antara X dan Y dengan mempertimbangkan kemungkinan kemunculan acak. Nilai *lift* yang lebih besar dari 1 menunjukkan adanya hubungan positif antarproduk. (Shihab, 2024)

c. Algoritma *FP-Grwoth*

Algoritma *FP-Growth* (*Frequent Pattern Growth*) dikembangkan sebagai pengembangan lanjutan dari algoritma *Apriori* untuk meningkatkan efisiensi dalam proses pencarian pola frekuensi tinggi. Tidak seperti *Apriori* yang harus melalui tahapan pembangkitan kandidat (*candidate generation*) secara berulang, *FP-Growth* memanfaatkan struktur data berbentuk pohon yang disebut *Frequent Pattern Tree (FP-Tree)* sebagai representasi ringkas dari data transaksi. (Aditiya & Defit, 2020)

Secara umum, proses kerja algoritma ini mencakup dua tahap utama. Pertama, sistem membangun struktur *FP-Tree* berdasarkan dataset transaksi untuk merekam hubungan antaritem secara hierarkis. Kedua, struktur pohon tersebut ditelusuri guna menemukan kombinasi item dengan frekuensi tinggi tanpa perlu melakukan pemindaian ulang terhadap keseluruhan dataset. Pendekatan ini mampu menghemat waktu pemrosesan sekaligus mengurangi kebutuhan memori secara signifikan. Keunggulan tersebut membuat *FP-Growth* menjadi salah satu algoritma paling efisien dan banyak digunakan dalam riset maupun aplikasi praktis, khususnya dalam analisis pola pembelian konsumen di industri ritel. (Namariq, 2025)

d. Parameter Minimum Support dan Minimum Confidence

Dalam algoritma *FP-Growth*, terdapat dua parameter utama yang sangat berpengaruh terhadap kualitas aturan asosiasi yang terbentuk, yaitu *minimum support* dan *minimum confidence*. Parameter *minimum support* berfungsi sebagai ambang batas frekuensi kemunculan itemset untuk menentukan seberapa sering kombinasi item tersebut muncul dalam dataset. Sementara itu, *minimum confidence* digunakan untuk menilai tingkat kepercayaan atau probabilitas keterkaitan antara bagian *antecedent* (X) dan *consequent* (Y) dalam suatu aturan asosiasi. (Purwati et al., 2023)

Penentuan nilai kedua parameter ini memerlukan pertimbangan yang cermat. Apabila nilai yang digunakan terlalu tinggi, maka beberapa pola penting berpotensi tidak terdeteksi. Sebaliknya, jika nilainya terlalu rendah, sistem dapat menghasilkan terlalu banyak aturan yang bersifat redundan dan kurang relevan. Oleh sebab itu, banyak penelitian melakukan variasi terhadap nilai support dan confidence untuk menemukan kombinasi yang paling optimal, sehingga aturan asosiasi yang diperoleh tetap bermakna, efisien, dan akurat dari sisi komputasi. (Anggrawan & Satria, 2021)

e. Penelitian Terdahulu

Sejumlah penelitian terdahulu telah membahas penerapan algoritma *FP-Growth* dalam penggalian pola pembelian serta perbandingan performanya dengan metode *association rule mining* lainnya. Al-Maolegi dan Arkok (2014) menemukan bahwa *FP-Growth* memiliki efisiensi waktu eksekusi yang lebih tinggi dibandingkan algoritma *Apriori* ketika diterapkan pada dataset transaksi berskala besar. (Art, 2023)

Selanjutnya, Pujari et al. (2019) mengevaluasi dampak variasi nilai *support* dan *confidence* terhadap jumlah aturan yang dihasilkan. Hasil

penelitian mereka menunjukkan bahwa semakin rendah nilai kedua parameter tersebut, semakin banyak aturan asosiasi yang terbentuk, meskipun relevansi antaritem cenderung menurun. Sementara itu, Rahman dan Azhari (2020) menggunakan *Groceries* Dataset untuk menganalisis pola belanja konsumen dan menyimpulkan bahwa *FP-Growth* efektif dalam mengidentifikasi keterkaitan antarproduk yang sering muncul bersama. (Wijaya & Pratama, 2022)

Berdasarkan berbagai temuan tersebut, dapat disimpulkan bahwa kajian mengenai pengaruh nilai *support* dan *confidence* masih relevan untuk dikaji lebih lanjut. Hal ini penting guna menentukan konfigurasi parameter yang paling optimal dalam penerapan algoritma *FP-Growth*, khususnya untuk analisis data transaksi ritel yang kompleks dan berdimensi besar. (Artsitella et al., 2021)

METODE PENELITIAN

1. Jenis dan Pendekatan Penelitian

Penelitian ini menerapkan pendekatan kuantitatif dengan metode eksperimental, yang berfokus pada pengujian performa algoritma *FP-Growth* dalam proses identifikasi pola keterkaitan antaritem pada data transaksi ritel. Pendekatan ini dipilih karena memungkinkan analisis terukur terhadap pengaruh variasi nilai *minimum support* dan *minimum confidence* terhadap hasil aturan asosiasi yang terbentuk. Pendekatan eksperimental dipilih karena penelitian ini menitikberatkan pada pengujian parameter serta analisis hasil yang bersifat objektif dan terukur berdasarkan metrik performa algoritma. Dengan demikian, penelitian ini tidak hanya bertujuan untuk mengimplementasikan algoritma *FP-Growth*, tetapi juga untuk mengamati secara empiris bagaimana perubahan parameter memengaruhi hasil aturan *asosiasi* yang terbentuk.

2. Data Penelitian

Data yang digunakan pada penelitian ini berasal dari *Groceries* Dataset, yang memuat catatan transaksi pembelian pelanggan di sektor ritel. Dataset tersebut terdiri atas 9.835 transaksi dengan 169 jenis produk berbeda, di mana setiap transaksi merepresentasikan sekumpulan item yang dibeli bersamaan oleh konsumen. Setiap baris dalam dataset merepresentasikan satu transaksi yang berisi kumpulan produk yang dibeli secara bersamaan oleh pelanggan. (Rachmawati et al., 2024)

Dataset ini bersifat terbuka untuk umum dan kerap dijadikan benchmark standar dalam penelitian *association rule mining*, karena memiliki struktur data yang sederhana namun tetap kaya akan informasi transaksi. Format awal data berbentuk daftar item yang dipisahkan oleh koma pada setiap baris transaksi. Sebelum dilakukan analisis menggunakan algoritma *FP-Growth*, data tersebut terlebih dahulu diubah ke dalam format transaksional agar dapat diproses secara optimal. (Dio et al., 2025)

3. Alat dan Perangkat Lunak

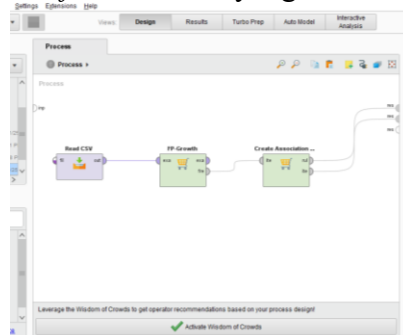
Proses analisis data dilakukan dengan memanfaatkan RapidMiner Studio versi 10.2.0, sebuah perangkat lunak yang menyediakan berbagai

operator lengkap untuk analisis *association rule mining*, termasuk implementasi algoritma *FP-Growth* dan modul *Create Association Rules*. Selain itu, RapidMiner menyediakan antarmuka visual yang intuitif, sehingga memudahkan pengguna dalam melakukan konfigurasi parameter, menjalankan pengujian, serta menafsirkan hasil aturan asosiasi baik dalam bentuk tabel maupun grafik *asosiasi (association graph)*. (Febrian & Fatah, 2024)

4. Tahapan Penelitian

Proses penelitian ini disusun melalui serangkaian tahapan sistematis sebagaimana digambarkan pada Gambar 1 berikut:

- Pengumpulan Data: Mengimpor *Groceries* Dataset ke dalam RapidMiner
- menggunakan operator *Read CSV*.
- Pra-pemrosesan Data: Menyesuaikan format data agar sesuai dengan input *FP-Growth*, termasuk penghapusan nilai kosong dan pemisahan item berdasarkan tanda koma.
- Penerapan Algoritma *FP-Growth*: Menerapkan algoritma *FP-Growth* untuk menghasilkan himpunan item yang sering muncul (*frequent itemsets*) berdasarkan nilai ambang minimum *support* tertentu.
- Pembuatan Aturan Asosiasi: Menggunakan operator *Create Association Rules* untuk menghasilkan aturan asosiasi dengan variasi nilai minimum *confidence*.
- Evaluasi Hasil: Mengevaluasi performa algoritma berdasarkan beberapa parameter variasi, meliputi jumlah aturan yang dihasilkan, waktu komputasi, dan nilai *lift* dari aturan yang terbentuk.



Gambar 1. Alur proses penerapan algoritma *FP-Growth* pada dataset *Groceries* menggunakan RapidMiner

Berdasarkan Gambar 1, tahap awal analisis dimulai dengan penggunaan operator *Read CSV*, yang berfungsi untuk mengimpor *Groceries* Dataset ke dalam sistem pemrosesan. Selanjutnya, operator *FP-Growth* digunakan untuk mengekstraksi *frequent itemsets* sesuai dengan ambang *minimum support* yang telah ditetapkan. Hasil keluaran dari tahap ini kemudian diteruskan ke operator *Create Association Rules*, yang menghasilkan aturan *asosiasi (association rules)* berdasarkan nilai *minimum confidence*. Alur tersebut menggambarkan proses analisis data transaksi secara otomatis untuk menemukan pola keterkaitan antarproduk dalam *Groceries* Dataset.

5. Variasi Parameter dan Skenario Pengujian

Penelitian ini menggunakan beberapa variasi parameter untuk mengevaluasi pengaruh *minimum support* dan *minimum confidence* terhadap

Evaluasi Performa Algoritma FP-Growth Berdasarkan Variasi Parameter Minimum Support dan Confidence pada Dataset Groceries (Arumeilia, Hasbi Firmansyah, Wahyu Arsiyani)

performa algoritma *FP-Growth*. Adapun rentang nilai parameter yang digunakan disajikan pada Tabel 1 berikut:

Table 1. Variasi parameter minimum support dan minimum confidence yang digunakan dalam pengujian algoritma FP-Growth

Parameter	Variasi Nilai	Keterangan
Minimum Support	0.01, 0.02, 0.03, 0.04	Menentukan frekuensi minimum itemset yang akan dipertimbangkan
Minimum Confidence	0.4, 0.5, 0.6	Menentukan tingkat keyakinan hubungan antara antecedent dan consequent

Nilai *minimum support* dan *minimum confidence* diuji melalui beberapa variasi untuk menganalisis pengaruhnya terhadap kinerja algoritma *FP-Growth* dalam menghasilkan aturan *asosiasi* yang relevan. Nilai *support* yang rendah memungkinkan lebih banyak pola ditemukan, namun berpotensi menghasilkan aturan yang kurang relevan, sedangkan *confidence* yang tinggi dapat mempersempit jumlah aturan tetapi meningkatkan kekuatan hubungan antaritem. Variasi ini dilakukan untuk menentukan kombinasi parameter yang paling optimal dalam menghasilkan aturan *asosiasi* yang efisien dan bermakna.

6. Evaluasi dan Analisis Data

Tahap evaluasi dilakukan dengan menelaah hasil keluaran algoritma, meliputi *frequent itemsets* serta *association rules* yang diperoleh dari proses analisis menggunakan RapidMiner. Data hasil eksperimen disajikan dalam bentuk tabel dan grafik perbandingan untuk mempermudah interpretasi. Analisis difokuskan pada pengaruh variasi nilai parameter terhadap performa algoritma *FP-Growth*, yang mencakup tiga aspek utama, yaitu:

- a. Jumlah aturan *asosiasi* yang dihasilkan,
- b. Nilai rata-rata *support*, *confidence*, dan *lift*,
- c. Waktu eksekusi algoritma selama proses analisis.

Hasil dari evaluasi ini menjadi dasar dalam menentukan kombinasi parameter yang paling optimal untuk penerapan algoritma *FP-Growth* pada *Groceries Dataset*.

HASIL DAN PEMBAHASAN

Penelitian ini difokuskan untuk menemukan pola keterkaitan antarproduk pada data transaksi ritel dengan menggunakan algoritma *FP-Growth*. Melalui proses analisis ini, diharapkan dapat diidentifikasi produk-produk yang cenderung dibeli secara bersamaan oleh pelanggan. Hasil tersebut dapat dimanfaatkan oleh pihak ritel sebagai dasar dalam menyusun strategi promosi, meningkatkan efektivitas penataan produk di toko, serta memperkuat keputusan bisnis berbasis data.

Tahapan penelitian diawali dengan proses pengumpulan dan pemrosesan *Groceries Dataset*, dilanjutkan dengan penerapan algoritma *FP-Growth* menggunakan RapidMiner Studio. Setiap kombinasi nilai *minimum support* serta *minimum confidence* diuji untuk mengetahui pengaruhnya terhadap jumlah aturan asosiasi yang terbentuk serta kekuatan hubungan antaritem yang diukur melalui nilai lift.

Bagian ini membahas hasil penerapan algoritma *FP-Growth*, termasuk analisis terhadap jumlah aturan yang dihasilkan, nilai lift yang diperoleh, dan interpretasi pola *asosiasi* antarproduk berdasarkan kombinasi parameter yang memberikan hasil paling optimal.

1. Hasil Pengujian Algoritma *FP-Growth*

Implementasi algoritma *FP-Growth* pada *Groceries Dataset* dilakukan dengan memanfaatkan operator *FP-Growth* serta *Create Association Rules* di dalam perangkat lunak RapidMiner. Proses pengujian dijalankan secara bertahap dengan melakukan variasi terhadap nilai *minimum support* dan *minimum confidence*, guna memperoleh kombinasi parameter yang paling optimal dalam membentuk aturan asosiasi yang bermakna.

Nilai *support*, *confidence*, dan *lift* pada setiap aturan dihitung secara otomatis oleh sistem RapidMiner menggunakan rumus standar dalam *association rule mining*.

$$Support(A \rightarrow B) = \frac{\text{Jumlah transaksi yang mengandung A dan B}}{\text{Total transaksi}}$$

$$Confidence(A \rightarrow B) = \frac{\text{Jumlah transaksi yang mengandung A dan B}}{\text{Jumlah transaksi yang mengandung A}}$$

$$Lift(A \rightarrow B) = \frac{Confidence(A \rightarrow B)}{Support(B)}$$

Keterangan :

- Support* : merepresentasikan frekuensi kemunculan bersama antara item A dan B dalam seluruh transaksi.
- Confidence* : mengukur probabilitas bahwa pembelian item A akan diikuti oleh pembelian item B.
- Lift* : digunakan untuk menilai kekuatan hubungan antarproduk, nilai *lift* yang melebihi 1 menunjukkan adanya keterkaitan positif antara dua item yang dianalisis.

Perhitungan ini dilakukan secara otomatis oleh sistem setelah proses algoritma *FP-Growth* menghasilkan *frequent itemsets*. Hasil pengujian setiap kombinasi parameter disajikan pada tabel 2 berikut ini:

Table 2. Hasil pengujian algoritma *FP-Growth* pada dataset *Groceries*

Evaluasi Performa Algoritma FP-Growth Berdasarkan Variasi Parameter Minimum Support dan Confidence pada Dataset Groceries (Arumeilia, Hasbi Firmansyah, Wahyu Arsiyani)

Minimum Support	<i>Minimum Confidence</i>	Jumlah Rules	Lift Tertinggi
0.01	0.4	71	2.344
0.01	0.5	18	0.993
0.01	0.6	0	-
0.02	0.4	17	2.450
0.02	0.5	1	2.007
0.02	0.6	0	-
0.03	0.4	5	0.994
0.03	0.5	0	-
0.03	0.6	0	-
0.04	0.4	4	0.994
0.04	0.5	0	-
0.04	0.6	0	-

2. Analisis Hasil Pengujian

Hasil pada Tabel 2 menunjukkan pola yang konsisten antara nilai parameter dan jumlah aturan yang dihasilkan. Semakin besar nilai minimum support maupun *minimum confidence* yang diterapkan, semakin sedikit aturan asosiasi yang terbentuk. Hal ini disebabkan karena peningkatan nilai ambang tersebut mempersempit ruang kombinasi item yang memenuhi kriteria pembentukan aturan, sehingga hanya pola dengan kekuatan hubungan tinggi yang tetap terdeteksi.

Kombinasi parameter *minimum support* 0.01 dan *confidence* 0.4 menghasilkan jumlah aturan terbanyak, yaitu 71 aturan, dengan nilai lift tertinggi sebesar 2.344. Nilai ini menunjukkan adanya hubungan yang kuat antarproduk, di mana produk-produk tersebut cenderung dibeli secara bersamaan dengan probabilitas tiga kali lebih besar dibandingkan kemunculan acak.

Sebaliknya, ketika nilai *confidence* dinaikkan menjadi 0,6, tidak terbentuk aturan *asosiasi* apa pun, karena ambang batas tingkat kepercayaan tersebut terlalu tinggi untuk memenuhi kriteria pembentukan pola. Pada konfigurasi *support* 0,02 dan *confidence* 0,4, diperoleh 17 aturan asosiasi dengan nilai *lift* tertinggi sebesar 2,450. Namun, ketika nilai *confidence* dinaikkan menjadi 0,5 dengan *support* yang sama (0,02), hanya satu aturan yang berhasil terbentuk, dengan nilai *lift* sebesar 2,007.

Temuan ini mengindikasikan bahwa semakin besar nilai parameter yang digunakan, maka semakin ketat proses penyaringan pola frekuensi tinggi, sehingga jumlah aturan asosiasi yang dihasilkan cenderung menurun. Hal tersebut menunjukkan adanya hubungan terbalik antara nilai ambang parameter dan banyaknya pola yang dapat ditemukan oleh algoritma FP-Growth.

3. Analisis Nilai Lift

Indikator *lift* digunakan untuk menilai sejauh mana kekuatan hubungan antarproduk dalam suatu aturan asosiasi. Nilai *lift* yang lebih besar dari satu mengindikasikan adanya keterkaitan positif antara item-item yang dibandingkan, dan semakin tinggi nilainya, semakin kuat hubungan korelatif yang terbentuk di antara produk-produk tersebut.

Berdasarkan hasil pengujian, nilai *lift* tertinggi diperoleh pada kombinasi *minimum support* 0.01 dan *confidence* 0.4, dengan nilai 2.344. Artinya, produk-produk dalam aturan tersebut memiliki kemungkinan sekitar tiga kali lebih besar untuk muncul bersamaan dibandingkan secara acak.

Dengan demikian, kombinasi parameter *support* 0.01 dan *confidence* 0.4 dapat dianggap sebagai konfigurasi paling optimal untuk menemukan hubungan signifikan antarproduk pada *Groceries* Dataset.

4. Pembahasan Pola Asosiasi Produk

Analisis lebih lanjut dilakukan terhadap pola asosiasi produk yang terbentuk pada kombinasi terbaik (*support* 0.01 dan *confidence* 0.4). Dari 71 aturan yang terbentuk, produk “*whole milk*” (susu murni) muncul paling sering, baik sebagai *antecedent* (bagian kiri aturan) maupun *consequent* (bagian kanan aturan). Hal ini menunjukkan bahwa *whole milk* merupakan produk dengan keterkaitan tinggi terhadap produk lain. Beberapa contoh aturan asosiasi disajikan pada tabel 3:

Table 3. Aturan asosiasi

Aturan Asosiasi	Support	Confidence	Lift
<i>Cream chese</i> → <i>whole milk</i>	0.016	0.415	1.626
<i>root vegetables</i> → <i>whole milk</i>	0.049	0.449	1.756
<i>yogurt, curd</i> → <i>whole milk</i>	0.011	0.599	2.344

Aturan ketiga memiliki *lift* tertinggi yaitu 2.344, menunjukkan bahwa pelanggan yang membeli *yogurt* dan *curd* memiliki kemungkinan dua kali lebih besar untuk juga membeli *whole milk* dibandingkan pelanggan lainnya.

Dengan demikian, dapat disimpulkan bahwa produk *whole milk* berperan sebagai produk pusat (*core product*) dalam pola pembelian konsumen pada dataset *Groceries*. Produk ini sering muncul bersamaan dengan bahan makanan rumah tangga seperti sayuran, *yogurt*, dan mentega, yang mencerminkan perilaku belanja konsumen secara logis.

5. Kesimpulan Sementara

Dari seluruh pengujian dan pembahasan di atas, dapat disimpulkan bahwa parameter *support* 0.01 dan *confidence* 0.4 merupakan kombinasi paling efektif untuk menghasilkan aturan *asosiasi* yang kuat dan bermakna. Hasil ini menunjukkan bahwa algoritma *FP-Growth* mampu menemukan pola belanja yang relevan dan dapat digunakan untuk mendukung strategi pemasaran serta penataan produk di toko ritel.

KESIMPULAN

Berdasarkan hasil penelitian dan pengujian algoritma *FP-Growth* pada *Groceries* Dataset, dapat disimpulkan bahwa variasi nilai *minimum support* dan *minimum confidence* memiliki pengaruh yang signifikan terhadap jumlah aturan (*association rules*) yang dihasilkan. Semakin tinggi nilai *support* dan *confidence* yang digunakan, maka semakin sedikit aturan yang terbentuk karena syarat kemunculan item dan kekuatan asosiasi semakin ketat.

Kombinasi parameter terbaik diperoleh pada *minimum support* 0.01 dan *confidence* 0.4, dengan jumlah aturan terbentuk sebanyak 71 rules serta *lift* tertinggi sebesar 2.3. Hasil ini menunjukkan adanya hubungan kuat antarproduk dalam data transaksi, di mana produk *whole milk* (susu murni) menjadi item yang paling sering muncul bersama produk lain. Pola ini menunjukkan bahwa *whole milk* berperan sebagai produk pusat dalam pola pembelian konsumen.

Secara keseluruhan, algoritma *FP-Growth* terbukti efektif dalam menemukan hubungan antarproduk yang tersembunyi pada dataset transaksi. Informasi yang dihasilkan dari pola *asosiasi* ini dapat digunakan oleh pihak ritel untuk mendukung strategi promosi, pengaturan tata letak produk, dan rekomendasi penjualan yang lebih tepat sasaran.

DAFTAR PUSTAKA

Aditiya, R., & Defit, S. (2020). *Jurnal Informatika Ekonomi Bisnis Prediksi Tingkat Ketersediaan Stock Sembako Menggunakan Algoritma FP-Growth dalam Meningkatkan Penjualan*. 2. <https://doi.org/10.37034/infec.v2i3.44>

Aktavera, B., Oktafia, H., Wijaya, L., Studi, P., Komputer, I., Petulai, U. P., Lebong, R., Studi, P., Informasi, S., Insan, U. B., Apriori, A., & Apriori, A. (2024). *ANALISIS ASSOCIATION RULE MENGGUNAKAN ALGORITMA*. 16(1), 54–61.

Anggrawan, A., & Satria, C. (2021). *Menentukan Akurasi Tata Letak Barang dengan Menggunakan Algoritma Apriori dan Algoritma FP-Growth Determination of Item Layout Accuracy using Apriori Algorithm and FP-Growth Algorithm*. 21(1), 125–138. <https://doi.org/10.30812/matrik.v21i1.1260>

Art, P. E. M. U. (2023). *P ERBANDINGAN A LGORITMA A PRIORI D AN F P -G ROWTH U NTUK R EKOMENDASI I TEM P AKET P ADA K ONTEN P ROMOSI D I*. 4(2), 19–24.

Artsitella, C. R., Apriliani, A. R., & Ashari, S. (2021). *Penerapan Association*

Rules - Market Basket Analysis untuk Mencari Frequent Itemset dengan Algoritma FP-Growth. 6(2), 61–69.

Chopvitayakun, S. (n.d.). *Analyzing Purchase Behavior Using FP Growth Technique to Find Association Rules.* 106–111. <https://doi.org/10.1145/3678610.3678618>

Dio, R., Dermawan, A. A., Yusiani, D. S., Herikson, R., & Kurniawan, D. E. (2025). *Scenario-Based Association Rule Mining in Veterinary Services Using FP-Growth : Differentiating Clinical and Customer-Driven Patterns.* 9(3), 1058–1065.

Febrian, F., & Fatah, Z. (2024). *Optimasi Penentuan Paket Hemat Menggunakan Algoritma FP-Growth untuk Meningkatkan Strategi Pemasaran.* 3(2), 170–177.

Hairani, H., & Guterres, J. X. (2024). *Exploring Customer Purchasing Patterns : A Study Utilizing FP-Growth Algorithm on Supermarket Transaction Data.* 3(1), 33–42. <https://doi.org/10.30812/IJECSA.v3i1.3874>

Idris, A. I., Sampetoding, E. A. M., Yoga, V., Ardhana, P., & Maritsa, I. (2022). *Comparison of Apriori , Apriori-TID and FP-Growth Algorithms in Market Basket Analysis at Grocery Stores.* 6(2), 107–112. <https://doi.org/10.30865/ijjics.v6i2.4535>

Jang, H., Yang, Y., & Park, J. S. (2021). *FP-Growth Algorithm for Discovering Region-Based Association Rule in the IoT Environment.*

Namariq, M. K. (2025). *Klasterisasi K-Means dan Model Markowitz dalam Pembentukan Portofolio Optimal Saham Syariah.* November.

No, V., Hal, J., Hafizh, M., Novita, T., Guswandi, D., Syahputra, H., & Mayola, L. (2023). *Implementasi Data Mining Menggunakan Algoritma Fp-Growth Untuk Menganalisa Transaksi Penjualan Ekspor Online.* 5(3).

P, Z. K. P., Iskandar, I., & Nazir, A. (2021). *Implementasi Algoritma FP-Growth untuk Menemukan Pola Keterkaitan Antara Matakuliah Pemrograman dan Matakuliah Matematika.* 7(2), 51–57. <https://doi.org/10.24014/coreit.v7i2.15351>

Purwati, N., Pedliyansah, Y., Kurniawan, H., Karnila, S., & Herwanto, R. (2023). *Komparasi Metode Apriori dan FP-Growth Data Mining Untuk Mengetahui Pola Penjualan.* 8(2), 155–161.

Rachmawati, D., Cahyana, Y., Awal, E. E., & Faisal, S. (2024). *Perbandingan Algoritma Apriori dan Algoritma FP-Growth dalam Menentukan Pola Penjualan Pupuk.* 3(1), 21–31.

Ranjan, R., & Sharma, A. (n.d.). *Evaluation of Frequent Itemset Mining Platforms using Apriori and FP- Growth Algorithm.* 1–6.

Shihab, A. (2024). *Jurnal Ilmiah Multidisiplin Nusantara Implementasi Algoritma FP-Growth Pada Dataset Sintetis Untuk Penentuan Pola Pembelian Sembako Dan Kebutuhan Harian Jurnal Ilmiah Multidisiplin Nusantara.* 2(November), 161–175.

Widjaja, A. E. (2024). *Analysis of Apriori and FP-Growth Algorithms for Market Basket Insights : A Case Study of The Bread Basket Bakery Sales.* 1(1), 63–83. <https://doi.org/10.47738/jdmdc.v1i1.2>

Wijaya, K. T., & Pratama, I. (2022). *Penerapan Algoritma FP-Growth Untuk Analisis Data Transaksi Penjualan Di Internet Learning Cafe Kaliurang.* 5(4), 642–651.