

# ANALISIS POLA PEMINJAMAN BUKU PADA SISTEM PERPUSTAKAAN DIGITAL MENGGUNAKAN ALGORITMA FP-GROWTH DAN TEKNIK PRUNING

Astari<sup>1</sup>

<sup>1</sup>STAI Latansa Mashiro

Email: astari@unilam.ac.id

Masuk: 24-07-2025, revisi: 28-08-2025, diterima untuk diterbitkan: 20-08-2025

---

## ABSTRAK

Manajemen perpustakaan menghadapi tantangan dalam pengelolaan data sirkulasi yang semakin kompleks akibat peningkatan literasi digital dan kebutuhan informasi yang terus berkembang. Penelitian ini bertujuan untuk menganalisis pola peminjaman buku di UPT Perpustakaan Universitas Latansa Mashiro dengan menerapkan algoritma Frequent Pattern Growth (FP-Growth) yang dioptimalkan melalui teknik pruning. Proses analisis dilakukan menggunakan perangkat lunak RapidMiner versi 9.10, dengan dataset terdiri dari 1.012 transaksi peminjaman buku selama periode Januari–Juni 2025. FP-Growth digunakan untuk mengekstraksi pola asosiasi antar itemset secara efisien tanpa proses pencarian kandidat, sedangkan pruning berfungsi untuk menghapus item dengan frekuensi rendah guna meningkatkan presisi hasil. Hasil eksperimen menunjukkan adanya korelasi kuat antara buku Manajemen dan Ekonomi (support 27%, confidence 54%, lift 956) yang meningkat menjadi (support 27%, confidence 64%, lift 1.233) setelah pruning diterapkan. Jumlah aturan asosiasi juga menurun dari 26 menjadi 8, namun dengan peningkatan kekuatan korelasi dan relevansi. Evaluasi visual dilakukan melalui perbandingan treemap dan confusion matrix-based validation untuk mengukur efektivitas aturan. Penelitian ini memiliki kontribusi baru dalam penerapan teknik pruning pada algoritma FP-Growth di konteks perpustakaan digital, yang masih jarang dieksplorasi dalam literatur. Temuan ini berimplikasi penting bagi pengambilan keputusan strategis dalam pengadaan koleksi, penataan layout rak, dan pengembangan sistem rekomendasi berbasis sirkulasi di lingkungan perpustakaan.

**Kata Kunci:** Data mining; FP-Growth; Pruning; RapidMiner; Aturan Asosiasi; Perpustakaan Digital

## ABSTRACT

*Library management faces growing challenges in handling increasingly complex circulation data, driven by the rise of digital literacy and the evolving demands for information. This study aims to analyze book borrowing patterns at the Library Unit (UPT) of Universitas Latansa Mashiro by applying the Frequent Pattern Growth (FP-Growth) algorithm, optimized through a pruning technique. The analysis was conducted using RapidMiner version 9.10, utilizing a dataset comprising 1,012 book loan transactions recorded between January and June 2025. FP-Growth was employed to efficiently extract association patterns among itemsets without generating candidate sets, while the pruning technique was used to eliminate low-frequency items to improve result precision. Experimental results revealed a strong correlation between books in the Management and Economics categories, with support of 27%, confidence of 54%, and a lift value of 0.956 improving to 27% support, 64% confidence, and a lift of 1.233 after pruning was applied. The number of association rules decreased from 26 to 8, yet showed enhanced correlation strength and relevance. Visual evaluation was conducted through treemap comparisons and confusion matrix-based validation to assess the effectiveness of the derived rules. This study contributes a novel application of the pruning technique in conjunction with the FP-Growth algorithm within a digital library context, an area still underexplored in the existing literature. The findings have significant implications for strategic decision-making in collection development, shelf layout design, and the enhancement of circulation-based recommendation systems in academic libraries.*

**Keywords:** Data Mining; FP-Growth; Pruning; RapidMiner; Association Rules; Digital Library

## 1. PENDAHULUAN

Transformasi digital perpustakaan telah merevolusi peran perpustakaan dari repositori statis menjadi platform dinamis untuk akses informasi, dukungan pendidikan, dan pelestarian budaya (Singh et al., 2022). Pergeseran ini diiringi dengan pertumbuhan data yang eksponensial, mulai dari transaksi katalog hingga interaksi pengguna di berbagai platform digital (Permatasari et al., 2020). Repositori digital kini menawarkan berbagai macam materi literatur, meningkatkan hasil pendidikan dengan membuat literatur lebih menarik dan mudah diakses (Hoerudin, 2025). Namun, tantangan tetap ada, termasuk tingkat literasi digital yang bervariasi dan ketersediaan sumber daya yang tidak konsisten (Hoerudin, 2025). Pandemi COVID-19 telah mempercepat transformasi ini, sehingga mendorong perpustakaan untuk menyesuaikan layanan dan praktik manajemennya. Untuk mengatasi tantangan ini, perpustakaan harus membangun infrastruktur yang kuat, meningkatkan aksesibilitas, dan menyiapkan rencana manajemen darurat. Selain itu, memanfaatkan media sosial dan situs web perpustakaan menjadi sangat penting untuk komunikasi yang efektif dengan pengguna dalam lanskap digital yang terus berkembang (Ashiq et al., 2022).

Meskipun UPT Perpustakaan Universitas Latansa Mashiro (UNILAM) Rangkasbitung, Banten telah mengadopsi sistem informasi terkomputerisasi untuk layanan sirkulasi peminjaman dan pengembalian buku, data yang terkumpul selama operasional masih bersifat pasif dan belum dianalisis secara sistematis. Data transaksi sirkulasi sejauh ini hanya digunakan sebagai bahan laporan administratif tanpa eksplorasi lebih lanjut untuk mendukung pengambilan keputusan berbasis informasi (Islam et al., 2023). Akibatnya, muncul sejumlah isu manajerial, seperti rendahnya pemahaman terhadap preferensi pengguna, ketidaktepatan dalam pengadaan koleksi, serta inefisiensi dalam penataan ruang buku. Sebagai contoh, terdapat kasus di mana buku yang telah diadakan ternyata memiliki tingkat peminjaman yang rendah, sementara koleksi yang justru diminati pengguna belum tersedia (Harisanty et al., 2023; Rafi et al., 2020). Selain itu, tidak adanya analisis asosiasi antar item menyebabkan kegagalan dalam memahami keterkaitan antar buku yang sering dipinjam bersamaan, sehingga berdampak pada penempatan rak yang tidak strategis dan pengalaman pencarian buku yang kurang optimal (Arabi et al., 2020). Hal ini menunjukkan bahwa institusi akademik, termasuk perpustakaan, perlu mengadopsi strategi data cerdas dan pendekatan berbasis analitik untuk memaksimalkan potensi data yang telah dimiliki (Gaftandzhieva et al., 2023; Tara et al., 2024). Permasalahan ini secara keseluruhan mengindikasikan perlunya penerapan pendekatan data-driven decision making untuk menciptakan layanan perpustakaan yang lebih cerdas, responsif, dan adaptif terhadap kebutuhan pengguna.

Penambangan data (data mining) berperan penting dalam mengekstraksi wawasan berharga yang sebelumnya tersembunyi dari kumpulan data berskala besar, sehingga memungkinkan pengambilan keputusan berbasis bukti dalam sistem informasi, termasuk layanan perpustakaan (Sunhare et al., 2020; Cui & Yan, 2020). Berbagai studi telah berhasil menerapkan teknik data mining untuk mengungkap pola perilaku peminjaman buku di lingkungan perpustakaan akademik. Cecilio & Catedrilla (2023) menggunakan algoritma Apriori untuk menganalisis transaksi peminjaman dan merekomendasikan pengaturan rak buku yang lebih efisien berdasarkan association rule mining. Sarker (2021) menerapkan metode clustering K-Means untuk mengungkap preferensi pengguna dan tren bacaan, mendukung alokasi sumber daya secara lebih optimal. Selain itu, Pratiwi et al. (2024) menerapkan pendekatan klasterisasi pada data keanggotaan untuk menentukan lokasi sosialisasi literasi yang strategis. Dalam konteks penggalian pola peminjaman yang sering terjadi, algoritma association rule seperti FP-Growth menjadi pilihan yang efisien dan dapat diskalakan, sebagai alternatif dari algoritma Apriori (Belkadi et al., 2023). Perbedaan utama antara keduanya terletak pada cara penanganan kandidat itemset. FP-Growth membangun struktur data Frequent Pattern Tree (FP-Tree) yang terkompresi untuk

merepresentasikan item dan frekuensinya secara hierarkis, memungkinkan pemrosesan data hanya dalam dua kali pemindaian basis data (Belkadi et al., 2023; Fernandez-Basso et al., 2023). Hal ini berbeda dengan Apriori yang memerlukan pemindaian ulang untuk setiap kombinasi item (Belkadi et al., 2023). FP-Tree terdiri dari simpul (node) yang menyimpan item, jalur (path) yang mencerminkan transaksi yang serupa, dan header table yang mempercepat pencarian pola. Struktur ini sangat berguna dalam domain Library Information Systems karena memungkinkan pustakawan menganalisis preferensi kolektif peminjam secara efisien.

Lebih lanjut, proses pruning diterapkan dalam analisis FP-Growth untuk menghapus item dengan frekuensi kemunculan yang rendah, sehingga hanya itemset yang bermakna dan berpengaruh signifikan terhadap pola asosiasi yang dipertahankan. Teknik ini juga sering digunakan dalam Deep Learning dan Decision Tree untuk menyederhanakan model dan menghindari overfitting (Geng et al., 2025). Sejumlah studi di luar domain perpustakaan memperkuat keunggulan algoritma FP-Growth. Belkadi et al., (2023) dan Fernandez-Basso et al., (2023) menunjukkan bahwa FP-Growth secara signifikan mengungguli Apriori dalam memproses data penjualan besar di sektor farmasi. Hussain et al., (2021) juga mengonfirmasi efektivitas FP-Growth dalam mengidentifikasi frequent itemsets pada pola penjualan dan pengadaan obat. Bukti ini mengindikasikan bahwa pendekatan pattern-based analytics, meskipun dikembangkan dalam konteks komersial, memiliki potensi yang sangat besar untuk diadopsi dalam Educational Data Mining (EDM) khususnya dalam optimalisasi pengelolaan koleksi, rekomendasi buku, serta pengembangan layanan cerdas berbasis sistem informasi perpustakaan. Dengan mengintegrasikan algoritma FP-Growth dan teknik pruning dalam sistem informasi perpustakaan, penelitian ini memberikan kontribusi yang relevan bagi pengembangan intelligent library services dalam ekosistem digital pendidikan tinggi, dan memperkaya literatur dalam bidang Library Information Systems dan EDM.

Berbagai studi sebelumnya telah menerapkan teknik association rule mining seperti algoritma FP-Growth dan Apriori untuk mengeksplorasi pola peminjaman buku di lingkungan perpustakaan. Salah satu penelitian oleh Belkadi et al., (2023) dan Fernandez-Basso et al., (2023) memanfaatkan algoritma FP-Growth dalam membangun sistem rekomendasi berbasis transaksi peminjaman buku. Namun, studi tersebut menunjukkan keterbatasan metodologis yang signifikan, terutama karena tetap memasukkan itemset yang tidak memenuhi ambang batas support minimum ke dalam proses analisis. Akibatnya, aturan asosiasi yang dihasilkan menjadi kurang valid secara statistik dan kehilangan relevansi praktis dalam konteks pengambilan keputusan operasional di perpustakaan. Secara khusus, sembilan aturan yang dihasilkan semuanya memiliki nilai confidence sebesar 100%. Meskipun secara teoritis confidence tinggi sering diartikan sebagai kekuatan asosiasi yang kuat, dalam praktiknya confidence 100% yang tidak diimbangi dengan support yang proporsional justru dapat menandakan spurious patterns (Singh et al., 2025). Hal ini mencerminkan kondisi di mana dua item muncul bersama dalam seluruh transaksi yang tersedia, tetapi hanya karena salah satunya sangat umum (frekuensi tinggi) atau terdapat data bias, bukan karena ada hubungan bermakna secara semantik atau perilaku pengguna.

Perspektif evaluatif menunjukkan bahwa confidence tinggi tanpa dukungan metrik lain seperti lift dan support yang memadai dapat menyesatkan dan menghasilkan aturan yang bersifat superfisial. Menurut Elbarawy et al. (2021), aturan yang tidak mempertimbangkan metrik gabungan seperti  $lift > 1$  dan  $support \geq threshold$  cenderung gagal menangkap relasi yang benar-benar menarik dan berpotensi menghasilkan overfitting to noise dalam model rekomendasi. Dari sisi implementasi, aturan yang terlalu umum seperti “jika meminjam buku A maka pasti meminjam buku B” tidak memiliki nilai prediktif yang diskriminatif dan tidak mampu memfasilitasi personalisasi layanan

pengguna. Hal ini sejalan dengan temuan EShah et al. (2021), yang menunjukkan bahwa dalam konteks sistem rekomendasi, aturan dengan confidence ekstrem cenderung mengaburkan preferensi individual dan tidak meningkatkan kualitas rekomendasi yang berbasis konteks atau profil pengguna. Untuk mengatasi keterbatasan tersebut, penelitian ini menerapkan strategi pruning sebagai metode optimalisasi dalam proses mining. Teknik ini berfungsi untuk menyaring itemset yang memiliki frekuensi rendah atau kontribusi lemah terhadap asosiasi yang bermakna. Dengan membatasi analisis hanya pada aturan dengan support dan lift yang layak secara statistik (misalnya, support  $\geq 0.2$ , lift  $> 1.0$ ), model dapat difokuskan pada pola yang tidak hanya sering terjadi, tetapi juga relevan dan dapat ditindaklanjuti. Studi terkini oleh Gao et al. (2023) dan Gray (2021) menegaskan bahwa pendekatan pruning berbasis threshold dinamis dan domain knowledge secara signifikan meningkatkan kualitas interpretasi rule dan performa sistem rekomendasi yang berbasis data mining.

Untuk mengatasi permasalahan munculnya pola yang redundan atau tidak informatif yang dihasilkan oleh algoritma FP-Growth standar, penelitian ini memperkenalkan pendekatan optimasi melalui teknik pruning. Teknik pruning bekerja dengan cara secara selektif menghapus item yang memiliki frekuensi kemunculan rendah, sehingga hanya itemset dengan konsistensi tinggi antar transaksi yang dipertahankan. Pendekatan ini secara signifikan mengurangi dimensi dataset dan menyederhanakan aturan asosiasi yang dihasilkan, menjadikan proses pembentukan aturan lebih efisien dan dapat dihitung secara komputasional dengan lebih mudah. Lebih penting lagi, itemset hasil pruning cenderung merefleksikan hubungan yang lebih bermakna secara statistik dan konseptual, sehingga meningkatkan relevansi serta keterjelasan (interpretability) dari pengetahuan yang diperoleh. Hal ini sejalan dengan temuan sebelumnya dalam bidang data mining, bahwa eliminasi item dengan frekuensi rendah dapat meningkatkan kualitas knowledge discovery (Han, 2011; Aggarwal, 2015).

Optimasi yang diajukan dalam studi ini bertujuan untuk mencapai keseimbangan antara efisiensi komputasi dan ketepatan model, tanpa mengabaikan pola-pola signifikan yang mungkin tersembunyi dalam data. Prinsip ini selaras dengan praktik umum dalam machine learning, di mana pruning telah diterapkan secara luas dalam arsitektur jaringan saraf dalam (Deep Neural Networks/DNNs) untuk mengurangi ukuran model dan waktu inferensi, sekaligus menjaga akurasi prediktif (Wang et al., 2025). Meskipun pruning dalam DNN lebih difokuskan pada parameter sparsity, prinsip dasarnya serupa, yakni menghilangkan sinyal yang lemah atau tidak relevan. Pendekatan ini juga kritikal dalam konteks association rule mining, di mana informasi yang tidak signifikan justru dapat mengaburkan pola utama. Purohit et al. (2024) mengembangkan teknik correlation-based pruning untuk CNNs yang mampu menghapus filter berlebih tanpa menurunkan akurasi sebuah analogi yang sepadan dengan penghapusan itemset frekuensi rendah yang tidak memberikan kontribusi signifikan terhadap insight yang dapat ditindaklanjuti. Dengan menerapkan teknik pruning dalam kerangka kerja FP-Growth, penelitian ini tidak hanya bertujuan untuk meningkatkan efisiensi penambangan pola, tetapi juga memastikan bahwa aturan-aturan yang dihasilkan memiliki signifikansi statistik dan relevansi praktis. Pendekatan ini sangat relevan dalam konteks data mining perpustakaan, di mana interpretabilitas dan kemampuan mendukung pengambilan keputusan menjadi aspek penting dalam pengembangan koleksi dan manajemen sumber daya.

Motivasi utama dari penelitian ini adalah untuk mengeksplorasi potensi teknik data mining, khususnya algoritma FP-Growth yang dikombinasikan dengan metode pruning, dalam meningkatkan efisiensi layanan perpustakaan di UPT Perpustakaan UNILAM. Melalui analisis pola historis peminjaman buku, pustakawan dapat mengambil keputusan yang lebih tepat dan

berbasis data (data-driven) dalam pengadaan dan pengelolaan koleksi. Pemahaman yang lebih mendalam terhadap preferensi pembaca memungkinkan perpustakaan untuk menyesuaikan sumber daya secara dinamis agar selaras dengan kebutuhan pengguna. Selain itu, identifikasi hubungan ko-peminjaman antar kategori buku yang sering terjadi dapat dimanfaatkan untuk penataan rak yang lebih strategis, sehingga meningkatkan aksesibilitas dan kemudahan navigasi koleksi. Dalam konteks pengelolaan perpustakaan modern, pendekatan ini tidak hanya menyederhanakan proses manajemen koleksi, tetapi juga memberikan kontribusi signifikan terhadap peningkatan kualitas layanan, kepuasan pengguna, serta personalisasi pengalaman pengguna di lingkungan akademik.

Pemilihan algoritma FP-Growth dibandingkan dengan algoritma asosiasi lainnya, seperti Apriori, didasarkan pada efisiensi waktu pemrosesan dan kemampuan skalabilitas yang lebih tinggi dalam menangani dataset berukuran besar tanpa harus menghasilkan kandidat itemset secara eksplisit (Han et al., 2011; Dwiputra et al., 2023). FP-Growth hanya memerlukan dua kali pemindaian basis data dan lebih adaptif dalam skenario big data, yang menjadikannya lebih tepat digunakan dalam sistem informasi perpustakaan digital yang semakin kompleks. Lebih lanjut, penelitian ini memiliki kontribusi signifikan dalam mendukung transformasi digital perpustakaan berbasis layanan pengguna (user-centered services), dengan menekankan pemanfaatan analitik data untuk menyesuaikan koleksi dan tata kelola layanan sesuai dengan preferensi pengguna secara real time (Banu et al., 2024; Cen et al., 2023). Pendekatan ini sejalan dengan tren global dalam manajemen perpustakaan akademik yang menekankan pada evidence-based collection development dan pengalaman pengguna sebagai tolok ukur utama keberhasilan layanan. Penelitian ini juga mengisi celah literatur (research gap) yang belum banyak dibahas, yaitu mengenai optimalisasi algoritma FP-Growth melalui teknik pruning. Studi sebelumnya cenderung mengabaikan pentingnya menyaring itemset dengan frekuensi rendah, yang dapat mengarah pada pembentukan aturan asosiasi yang bersifat generik dan kurang bermakna (Banu et al., 2024; Cen et al., 2023). Oleh karena itu, dengan menerapkan teknik pruning secara selektif, penelitian ini menawarkan model analisis yang lebih akurat, relevan, dan praktis untuk diterapkan dalam pengambilan keputusan strategis di perpustakaan digital.

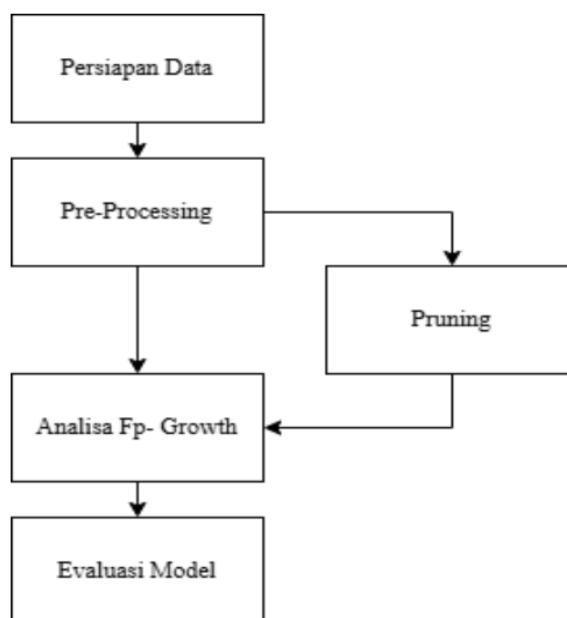
Penelitian ini bertujuan untuk mengidentifikasi pola peminjaman buku yang sering terjadi di Perpustakaan UNILAM dengan menerapkan algoritma FP-Growth yang dioptimalkan melalui teknik pruning. Tujuan utama penelitian ini adalah untuk mengekstrak aturan asosiasi yang secara efektif mengungkapkan perilaku peminjaman dengan presisi analitis yang lebih tinggi dan efisiensi komputasi yang lebih baik. Dengan menghilangkan himpunan item berfrekuensi rendah, pruning meningkatkan keterbacaan aturan dan memastikan bahwa hanya pola yang secara statistik signifikan dan relevan secara operasional yang dipertahankan. Berbeda dengan penelitian sebelumnya yang sebagian besar mengandalkan Apriori atau FP-Growth standar tanpa optimasi, penelitian ini memperkenalkan model hibrida yang mengintegrasikan pruning sebagai peningkatan strategis. Sejauh yang kami ketahui, penerapan pendekatan gabungan ini FP-Growth dengan pruning untuk menganalisis data peminjaman historis dan lokal di perpustakaan universitas di Indonesia masih jarang. Temuan ini diharapkan dapat mendukung pengambilan keputusan berbasis data dalam pengembangan koleksi, prioritas akuisisi, dan perencanaan sumber daya spasial pada akhirnya meningkatkan kualitas layanan melalui kecerdasan keputusan berbasis pola. Penelitian ini berkontribusi pada pengembangan analitik layanan publik dalam ekosistem perpustakaan digital, terutama dalam konteks yang sedang berkembang

## 2. METODE PENELITIAN

Alur kerja penelitian ini secara keseluruhan dapat dilihat pada Gambar 1. Proses dimulai dari tahap persiapan data, yang meliputi integrasi data, pembersihan data, serta visualisasi awal. Langkah ini penting untuk memastikan kualitas, konsistensi, dan reliabilitas data yang akan digunakan dalam proses analisis selanjutnya (Aggarwal, 2015; Han, Kamber, & Pei, 2011). Persiapan data yang baik akan meminimalkan gangguan dari outlier, noise, dan missing values yang berpotensi menurunkan performa model penambangan data (Pratiwi, Indahwati, & Fitrianto, 2024). Tahap berikutnya adalah penerapan algoritma Frequent Pattern Growth (FP-Growth), yang digunakan untuk mengeksplorasi himpunan item yang sering muncul (frequent itemsets) dan membangun aturan asosiasi dari dataset transaksi peminjaman buku. Algoritma ini dikenal efisien karena tidak memerlukan proses pencarian kandidat seperti Apriori, sehingga lebih hemat memori dan waktu komputasi dalam menangani dataset besar (Han et al., 2011; Dwiputra et al., 2023; Fernandez-Basso et al., 2024).

Untuk meningkatkan relevansi dan kekuatan prediktif dari aturan yang dihasilkan, dilakukan penerapan teknik pruning (pemangkasan). Teknik ini secara sistematis mengeliminasi item atau aturan dengan support rendah, yang berpotensi menimbulkan kebisingan (noise) atau menurunkan keterbacaan model, serta meningkatkan presisi dalam menyaring pola-pola yang benar-benar bermakna (Belkadi et al., 2023; Cui & Yan, 2020). Dalam konteks perpustakaan digital, strategi ini masih tergolong jarang dieksplorasi namun sangat potensial untuk mengoptimalkan sistem rekomendasi dan pengambilan keputusan koleksi (Arabi et al., 2020; Tara, Rafi, & Ahmad, 2024).

Tahap akhir dari alur kerja adalah evaluasi model menggunakan tiga metrik utama dalam association rule mining, yaitu support, confidence, dan lift. Ketiga metrik ini digunakan untuk menilai signifikansi dan kekuatan korelasi antar item dalam pola-pola yang ditemukan (Dwiputra et al., 2023; Sarker, 2021). Untuk menguji efektivitas teknik pruning yang diterapkan, dilakukan pula analisis komparatif, dengan cara membandingkan kekuatan aturan sebelum dan sesudah pruning. Evaluasi ini bertujuan untuk menilai tidak hanya kuantitas aturan, tetapi juga kualitas, kekuatan prediksi, dan aplikabilitas praktisnya dalam konteks manajemen perpustakaan digital (Cecilio et al., 2023; Islam, Islam, & Islam, 2023).



Gambar 1. Alur Penelitian

## 2.1 Persiapan Data

Data yang digunakan dalam penelitian ini diperoleh dari sistem sirkulasi UPT Perpustakaan UNILAM, yang mencatat transaksi peminjaman dan pengembalian buku secara digital. Fokus utama penelitian diarahkan pada data peminjaman karena data ini mencerminkan preferensi pengguna dan pola perilaku literasi yang sedang berkembang, terutama dalam konteks perpustakaan digital (Ashiq et al., 2022; Islam et al., 2023). Untuk memastikan keandalan dan kelengkapan informasi, diterapkan pendekatan triangulasi metode. Observasi langsung dilakukan untuk memahami alur teknis layanan sirkulasi dan entri data secara praktis. Dokumentasi arsip peminjaman digunakan sebagai sumber historis, sementara wawancara semi-terstruktur dengan pustakawan bertujuan menggali wawasan mendalam terkait kebijakan operasional dan kecenderungan perilaku pengunjung (Muhammad et al., 2020). Kombinasi teknik ini menghasilkan dataset yang lebih representatif dan kontekstual.

Data kemudian diproses melalui tahapan *preprocessing* yang meliputi pembersihan data, standarisasi format, serta penanganan *missing values* dan potensi *outliers* (Aggarwal, 2015; Pratiwi et al., 2024). Selanjutnya, analisis dilakukan menggunakan algoritma Frequent Pattern Growth (FP-Growth) karena kemampuannya dalam mengekstraksi pola asosiasi itemset tanpa proses pencarian kandidat, sehingga lebih efisien untuk data berskala besar (Han et al., 2011; Fernandez-Basso et al., 2024). Penerapan teknik *pruning* dilakukan untuk meningkatkan akurasi hasil dengan menghilangkan aturan dengan tingkat frekuensi rendah (Dwiputra et al., 2023; Belkadi et al., 2023). Struktur dataset diekstraksi dari sistem perpustakaan dalam format digital (CSV dan MySQL) lalu dinormalisasi untuk menjaga konsistensi semantik antarkategori. Selama proses ini, seluruh informasi yang mengandung *Personally Identifiable Information* (PII) dihapus atau dianonimkan sesuai dengan prinsip etika dan regulasi perlindungan data (Tara et al., 2024; Harisanty et al., 2023).

## 2.2 Pra-Pemrosesan Data

Pra-pemrosesan data merupakan tahap krusial dalam setiap proses data mining karena berperan besar dalam meningkatkan kualitas input, validitas hasil analisis, serta efisiensi algoritma. Seperti ditegaskan oleh Han et al. (2011), kualitas pra-pemrosesan sangat menentukan kejernihan sinyal informasi dan ketepatan dalam penemuan pola, khususnya dalam konteks penambangan aturan asosiatif. Lebih lanjut, Aggarwal (2015) menyatakan bahwa reduksi noise dan penghilangan outlier pada tahap awal berkontribusi langsung dalam menurunkan kompleksitas pemodelan. Dalam penelitian ini, tahapan pra-pemrosesan mencakup tiga langkah utama: integrasi data, pembersihan data, dan visualisasi, yang seluruhnya disiapkan untuk mendukung optimalisasi algoritma *FP-Growth* yang digunakan.

### a. Integrasi Data

Integrasi data adalah proses menggabungkan kumpulan data dari berbagai sumber heterogen seperti log sirkulasi historis, catatan observasi, dan arsip digital ke dalam satu struktur yang konsisten dan terstandarisasi secara semantik. Tahap ini penting untuk menjaga integritas referensial dan menghindari konflik makna, terutama saat data berasal dari sumber berbeda. Dalam konteks ini, penyamaan skema atribut dilakukan secara teliti untuk menghindari redundansi dan memastikan keterpaduan data yang siap diolah secara efisien. Belkadi et al. (2023) dalam penerapan rule mining untuk pemetaan digital juga menekankan pentingnya konsistensi semantik dalam proses integrasi data.

### b. Pembersihan Data

Tahap ini bertujuan menghilangkan segala bentuk anomali sintaksis maupun semantik, meliputi: a) penanganan Nilai Hilang: Transaksi yang tidak memuat informasi penting (seperti jenis buku atau tanggal peminjaman) dieliminasi secara sistematis agar tidak menghasilkan distorsi dalam perhitungan *support* dan *confidence* (Dwiputra et al., 2023); b) penghapusan Duplikasi: Transaksi dengan kombinasi item identik yang terekam lebih dari satu kali dihapus agar dataset tetap unik dan tidak bias; dan c) reduksi Noise: Atribut non-informatif seperti ID peminjam atau metadata yang tidak relevan dihilangkan. Hal ini sejalan dengan temuan Aggarwal (2015) yang menyarankan penyederhanaan dimensi untuk meningkatkan rasio signal-to-noise dan mendukung proses penambangan yang lebih presisi.

Tabel 1. Perbandingan Antara Dataset Mentah dan Dataset Hasil Pembersihan

| Deskripsi                        | Dataset Mentah | Dataset Bersih |
|----------------------------------|----------------|----------------|
| Jumlah Total Transaksi           | 1.012          | 978            |
| Jumlah Nilai Hilang (baris)      | 34             | 0              |
| Transaksi Duplikat               | 12             | 0              |
| Atribut Digunakan untuk Analisis | 5              | 2              |

Sebagaimana dinyatakan oleh Fernandez-Basso et al. (2024), model frequent pattern mining akan menghasilkan aturan yang lebih konvergen dan akurat jika didukung oleh dataset yang bersih dan bebas anomali. Dalam penelitian ini, proses pembersihan terbukti mempercepat konvergensi FP-Growth serta memperkuat kualitas asosiasi yang dihasilkan.

### c. Visualisasi Data

Exploratory Data Analysis (EDA) dilakukan dengan menerapkan teknik visualisasi seperti *treemap* dan histogram frekuensi untuk memvalidasi hasil integrasi serta pembersihan data awal. Visualisasi ini tidak hanya berfungsi sebagai alat presentasi, tetapi juga memainkan peran penting dalam mendekripsi anomali, pola menyimpang (*outliers*), dan kategori langka sebelum data dikonversi ke bentuk biner. Langkah ini sangat esensial, mengingat algoritma FP-Growth bekerja secara optimal pada struktur data transaksional dalam format biner, tanpa memerlukan proses pencarian kandidat (Han et al., 2011). Proses identifikasi *outlier* dalam tahap EDA mengacu pada prinsip analisis outlier seperti dikemukakan oleh Aggarwal (2015), yang menekankan pentingnya pemetaan distribusi data dan penanganan nilai ekstrem untuk meningkatkan akurasi dalam penambangan pola asosiatif. Selain itu, pendekatan visual ini juga memperkaya validitas hasil dengan memberikan gambaran awal terhadap persebaran transaksi berdasarkan jenis buku. Untuk memastikan kesesuaian dengan kerangka kerja algoritma FP-Growth, penelitian ini mengasumsikan bahwa: a) seluruh variabel diklasifikasikan sebagai data kategorikal nominal yang merepresentasikan jenis atau kategori buku; b) atribut temporal (misalnya tanggal peminjaman) tidak disertakan dalam pemodelan aturan tetapi digunakan dalam validasi dan analisis visualisasi; dan c) data transaksional dikonversi ke dalam matriks biner satu-lawan-satu (*one-hot encoding*), dengan baris mewakili satu transaksi peminjaman dan kolom mewakili masing-masing kategori item (Dwiputra et al., 2023). Lebih jauh, pendekatan ini didukung oleh literatur kontekstual dalam pengembangan sistem rekomendasi dan manajemen pustaka berbasis preferensi pengguna dan analisis kontekstual (Arabi, Balakrishnan, & Shuib, 2020; Banu et al., 2024), yang menekankan pentingnya pemahaman terhadap pola pengguna dalam desain sistem informasi perpustakaan digital. Penyesuaian awal melalui EDA berimplikasi besar terhadap kualitas aturan asosiatif yang dihasilkan, baik dari sisi presisi maupun relevansi terhadap kebutuhan pengguna (Cecilio, Catedrilla, & A, 2023).

### 2.3 Pruning

Untuk memperkaya evaluasi model, analisis perbandingan antara algoritma FP-Growth dengan dan tanpa penerapan teknik *pruning* sebaiknya dilengkapi dengan metrik tambahan seperti rule density (kepadatan aturan) dan compression rate (tingkat pemadatan pola). Metrik-metrik ini tidak hanya mencerminkan efisiensi komputasional algoritma, tetapi juga memberikan gambaran yang lebih holistik terhadap kualitas dan keterwakilan pengetahuan yang dihasilkan oleh model (Han, Kamber, & Pei, 2011; Aggarwal, 2015). Penambahan metrik ini penting mengingat bahwa dalam konteks penggalian aturan asosiatif, jumlah aturan yang terlalu banyak dan tidak relevan dapat mengurangi *interpretability* model serta menyulitkan proses pengambilan keputusan berbasis data, khususnya dalam lingkungan perpustakaan digital (Dwiputra et al., 2023; Cecilio et al., 2023). Oleh karena itu, teknik *pruning* digunakan untuk menyaring aturan-aturan dengan *support* dan *confidence* rendah guna meningkatkan presisi dan keterbacaan hasil (Fernandez-Basso et al., 2024; Belkadi et al., 2023).

Lebih jauh, perlu dicermati adanya trade-off antara *rule diversity* (keragaman aturan) dan *rule precision* (ketepatan aturan) sebagai dampak langsung dari proses *pruning*. Semakin ketat teknik ini diterapkan, jumlah aturan yang dihasilkan cenderung menurun secara signifikan, namun dengan kualitas prediktif dan relevansi yang meningkat. Sebaliknya, jika tidak dilakukan *pruning*, FP-Growth dapat menghasilkan sejumlah besar aturan yang bersifat *redundant* dan terlalu umum, yang justru berisiko menurunkan daya guna model dalam konteks aplikasi praktis (Li & Zhang, 2022; Sarker, 2021). Dalam konteks ini, *pruning* juga berperan sebagai mekanisme optimasi struktural yang mampu memperkuat kualitas *knowledge representation* hasil eksplorasi data. Studi oleh Arabi et al. (2020) dan Islam et al. (2023) menunjukkan bahwa efektivitas sistem rekomendasi, termasuk dalam domain perpustakaan digital, sangat dipengaruhi oleh selektivitas dan relevansi aturan asosiasi yang diterapkan.

### 2.4 Penambangan Aturan Asosiasi Menggunakan Algoritma FP-Growth

Penambangan Aturan Asosiasi (Association Rule Mining/ARM) merupakan salah satu teknik analisis utama dalam domain data mining yang bertujuan untuk mengungkap hubungan yang signifikan dan tidak sepele antar item dalam dataset skala besar. Teknik ini telah banyak diterapkan di berbagai bidang seperti ritel, layanan kesehatan, dan sistem informasi perpustakaan untuk mengekstraksi pola tersembunyi yang dapat memfasilitasi pengambilan keputusan strategis berbasis data (Han, Kamber, & Pei, 2011; Belkadi et al., 2023). Dalam penelitian ini, algoritma Frequent Pattern Growth (FP-Growth) digunakan karena keunggulan komputasinya dibandingkan algoritma Apriori. Berbeda dengan Apriori yang membutuhkan pembangkitan kandidat itemset secara menyeluruh, FP-Growth menghilangkan langkah tersebut dan hanya melakukan dua kali pemindaian terhadap basis data transaksi. Algoritma ini membangun struktur pohon awalan (prefix-tree) yang kompak yang disebut FP-Tree, yang menangkap distribusi frekuensi itemset secara hierarkis, sehingga mengoptimalkan efisiensi memori dan proses komputasi (Fernandez-Basso, Ruiz, & Martin-Bautista, 2024; Dwiputra et al., 2023).

Efektivitas aturan asosiasi yang dihasilkan oleh FP-Growth dievaluasi menggunakan tiga metrik utama: a) *Support*: Mengukur proporsi transaksi dalam dataset yang mengandung itemset tertentu ( $X \cup Y$ ). Metrik ini mencerminkan tingkat prevalensi umum dari suatu pola; b) *Confidence*: Menunjukkan kemungkinan item  $Y$  muncul dalam transaksi yang mengandung item  $X$ , sehingga mencerminkan kekuatan prediktif dari aturan  $X \rightarrow Y$ ; dan c) *Lift*: Mengevaluasi tingkat keterkaitan antara  $X$  dan  $Y$  dengan membandingkan ko-kemunculan yang diamati terhadap yang diharapkan jika kedua item tersebut saling independen secara statistik. Nilai lift  $> 1$  menunjukkan asosiasi positif (Aggarwal, 2015). Ketiga metrik ini tidak hanya berfungsi sebagai tolok ukur

evaluatif, tetapi juga sebagai parameter dasar untuk proses pruning dan penyempurnaan aturan agar tetap relevan dan andal dalam aplikasi nyata (Cui & Yan, 2020; Cecilio et al., 2023). Dalam konteks perpustakaan, penerapan FP-Growth dalam menganalisis pola sirkulasi buku sejalan dengan tren transformasi perpustakaan digital, di mana penambangan perilaku peminjaman pengguna menjadi hal krusial untuk meningkatkan sistem rekomendasi dan strategi pengembangan koleksi (Arabi et al., 2020; Ashiq, Jabeen, & Mahmood, 2022). Pendekatan ini juga sejalan dengan upaya yang lebih luas di dunia pendidikan tinggi untuk menginstitusikan pengambilan keputusan berbasis data dalam pengelolaan sumber daya (Gaftandzhieva et al., 2023; Tara, Rafi, & Ahmad, 2024).

$$\text{Support}(X \Rightarrow Y) = \frac{|X \cup Y|}{N} \quad \text{Confidence}(X \Rightarrow Y) = \frac{|X \cup Y|}{|X|} \quad \text{Lift}(X \Rightarrow Y) = \frac{\text{Confidence}(X \Rightarrow Y)}{\text{Support}(Y)}$$

Persamaan (1)

## 2.5 Optimasi melalui Teknik *Pruning*

Untuk meningkatkan signifikansi analisis dan mengurangi gangguan dari item yang jarang muncul, diterapkan ambang minimum support sebesar 0,2 (20%). Ambang ini berperan penting dalam mengeliminasi item-item dengan frekuensi rendah agar hasil analisis lebih fokus dan tidak terdilusi oleh pola yang tidak bermakna secara statistik (Han, Kamber, & Pei, 2011). Teknik *pruning* yang diterapkan berfungsi menyaring item yang tidak signifikan, menyederhanakan kompleksitas model, dan menyoroti pola-pola yang lebih relevan. Strategi ini sejalan dengan praktik terbaik dalam *rule mining*, sebagaimana disarankan oleh Cheng et al. (2024) dan diperkuat oleh Aggarwal (2015), yang menekankan pentingnya penghapusan *outlier* atau item marginal untuk menghindari *overfitting* dan mendukung validitas prediktif. Dalam konteks ini, *pruning* bukan sekadar menyederhanakan aturan, tetapi mempertahankan efisiensi komputasional tanpa mengorbankan integritas model. Hal ini menjadi krusial dalam pengolahan data perpustakaan digital, di mana kompleksitas data dapat berdampak langsung pada keandalan sistem rekomendasi dan pengambilan keputusan strategis (Dwiputra et al., 2023; Arabi et al., 2020).

Salah satu hasil penting dari penerapan algoritma FP-Growth dalam penelitian ini ditampilkan dalam Tabel 4. Sebanyak 27% dari total transaksi menunjukkan keterkaitan peminjaman antara buku kategori *Manajemen* dan *Ekonomi*. Nilai *support* ini tergolong tinggi, menandakan bahwa kombinasi tersebut merupakan pola yang cukup dominan dalam data sirkulasi. Sementara itu, nilai *confidence* sebesar 54% mengindikasikan bahwa lebih dari separuh pengguna yang meminjam buku *Manajemen* juga turut meminjam buku *Ekonomi*. Hal ini mengimplikasikan adanya minat ganda pengguna terhadap dua domain keilmuan yang saling beririsan. Nilai *lift* sebesar 1.233 memperkuat interpretasi tersebut. Ini berarti bahwa peluang buku *Ekonomi* dipinjam meningkat sebesar 23,3% ketika buku *Manajemen* telah dipinjam, dibandingkan jika peminjaman terjadi secara acak. Sebagaimana dijelaskan oleh Belkadi et al. (2023), nilai *lift* di atas 1 menandakan adanya asosiasi positif yang kuat dan layak ditindaklanjuti dalam kebijakan pengadaan koleksi maupun sistem rekomendasi personal.

Hasil lain dari analisis menunjukkan bahwa meskipun kombinasi buku *Budaya* dan *Ekonomi* hanya muncul pada 13% transaksi (*support* rendah), nilai *confidence*-nya mencapai 64% dan *lift*-nya 1.208. Ini menunjukkan bahwa aturan tersebut tetap relevan meskipun jarang terjadi, karena kekuatan asosiasi antara itemset tersebut sangat signifikan. Hal ini selaras dengan kajian oleh Fernandez-Basso et al. (2024) yang menegaskan bahwa *confidence* dan *lift* lebih penting daripada frekuensi semata dalam konteks *pattern discovery* berbasis perilaku pengguna. Secara keseluruhan, model FP-Growth tanpa pruning menghasilkan 26 aturan, sebagian besar dengan

*confidence* 100%. Namun, banyak dari aturan tersebut bersifat generik dan berpotensi mencerminkan overfitting. Setelah diterapkan *pruning*, hanya tersisa 8 aturan, tetapi dengan kualitas korelasi yang lebih tinggi baik secara statistik maupun semantik. Fenomena ini menunjukkan bahwa *pruning* tidak hanya menyaring noise, tetapi juga meningkatkan nilai guna praktis dari hasil mining sebagaimana ditekankan oleh Aggarwal (2015) dalam konteks *frequent pattern mining* yang efisien dan relevan. Dengan demikian, temuan ini berimplikasi signifikan terhadap pengelolaan perpustakaan modern, khususnya dalam menyusun strategi pengadaan koleksi tematik, pengaturan tata letak rak buku, serta pengembangan sistem rekomendasi berbasis preferensi aktual pengguna. Dalam kerangka yang lebih luas, penerapan teknik ini sejalan dengan arah transformasi perpustakaan digital yang dituntut adaptif dan berbasis data, sebagaimana diulas dalam Ashiq et al. (2022) dan Harisanty et al. (2023).

## 2.6 Strategi Evaluasi Model

Untuk menilai efektivitas algoritma FP-Growth secara ketat, baik dengan maupun tanpa penerapan teknik pruning, penelitian ini menggunakan tiga metrik utama yang umum digunakan dalam *association rule mining*, yaitu: *Support*, *Confidence*, dan *Lift*. Ketiga metrik ini dipilih bukan hanya karena relevansi statistiknya, tetapi juga karena kemampuannya dalam mengungkap kekuatan dan frekuensi hubungan antar *itemset* (Han et al., 2011; Fernandez-Basso, Ruiz, & Martin-Bautista, 2024). Secara khusus, *Support* mengukur seberapa sering sebuah *itemset* muncul dalam keseluruhan dataset; *Confidence* menunjukkan probabilitas kemunculan konsekuensi berdasarkan kondisi pada *antecedent*; dan *Lift* mengevaluasi sejauh mana kedua *itemset* tersebut saling bergantung secara statistik melebihi kemungkinan acak. Alih-alih hanya menitikberatkan pada nilai maksimum, evaluasi ini menekankan analisis komparatif terhadap distribusi nilai dari masing-masing metrik pada model yang menggunakan pruning dan yang tidak. Nilai maksimum, rata-rata, dan minimum dianalisis untuk memberikan gambaran representatif mengenai kualitas aturan dan ketahanan model (Dwiputra et al., 2023). Selain itu, aturan-aturan asosiasi yang dihasilkan juga dianalisis dari segi interpretabilitas dan kemanfaatan semantic yakni sejauh mana aturan-aturan tersebut menyajikan wawasan yang bermakna dan dapat diterapkan dalam pengambilan keputusan manajemen perpustakaan (Cecilio & Catedrilla, 2023).

Jumlah total aturan yang dihasilkan juga dianalisis sebagai indikator kesederhanaan model (parsimony) dan efisiensi interpretasi. Model tanpa pruning cenderung menghasilkan aturan yang terlalu banyak, sehingga menimbulkan redundansi dan beban kognitif bagi pengambil keputusan, yang dapat mengaburkan pola-pola yang benar-benar bermakna (Aggarwal, 2015; Arabi, Balakrishnan, & Shuib, 2020). Sebaliknya, teknik pruning berfungsi menyederhanakan kompleksitas aturan dengan menghapus *itemset* yang memiliki tingkat dukungan rendah, sehingga meningkatkan kejelasan dan kegunaan praktis dari pola yang ditemukan (Belkadi et al., 2023). Terakhir, untuk memastikan relevansi kontekstual, penelitian ini merekomendasikan integrasi validasi berbasis domain, seperti masukan dari pustakawan atau pencocokan dengan preferensi peminjaman pengguna (Ashiq et al., 2022; Islam, 2023; Harisanty et al., 2023). Pendekatan evaluatif yang terintegrasi ini memastikan bahwa model tidak hanya optimal secara statistik, tetapi juga relevan dan selaras dengan dinamika nyata dalam peningkatan layanan perpustakaan digital.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Persiapan Data

Dataset yang digunakan dalam penelitian ini terdiri dari catatan transaksi sirkulasi yang diperoleh dari UPT Perpustakaan Universitas La Tansa Mashiro (UNILAM), mencakup periode Januari hingga Juni 2025, dengan total 1.012 entri. Untuk keperluan analisis, atribut yang dipilih meliputi tanggal transaksi dan kategori atau jenis buku, yang dinilai relevan dalam menangkap pola

peminjaman secara temporal serta tren klasifikasi bahan pustaka (Han, Kamber, & Pei, 2011; Aggarwal, 2015). Sebelum dianalisis, data menjalani proses *cleaning* dan *normalization* untuk memastikan kualitas, konsistensi, dan keakuratan, serta menghapus entri duplikat atau tidak lengkap suatu tahapan penting dalam menjamin keandalan proses *data mining* (Cui & Yan, 2020). Analisis statistik deskriptif dilakukan untuk memahami distribusi kategori buku. Dataset ini diklasifikasikan ke dalam empat kategori utama: fiksi, non-fiksi, referensi, dan sumber daya digital. Visualisasi tren peminjaman bulanan disajikan melalui tabel distribusi frekuensi untuk menangkap preferensi pengguna dan kecenderungan musiman.

Tujuan utama penelitian ini adalah: (1) mengidentifikasi tren peminjaman berdasarkan waktu dan jenis koleksi, serta (2) mengeksplorasi pola dominan untuk mendukung pengembangan sistem rekomendasi berbasis preferensi pengguna. Untuk itu, digunakan metode *clustering* dan *association rule mining* sebagai teknik utama untuk menggali perilaku pengguna dan potensi rekomendasi buku secara otomatis (Dwiputra et al., 2023; Cecilio et al., 2023). Penekanan pada pendekatan berbasis perilaku ini sejalan dengan tren dalam bidang *Educational Data Mining* (EDM) dan *Library Information Systems*, yang menuntut perancangan layanan perpustakaan yang proaktif dan cerdas (Romero & Ventura, 2020; Tara, Rafi, & Ahmad, 2024; Arabi, Balakrishnan, & Shuib, 2020). Seluruh proses mulai dari *data pre-processing*, analisis, hingga visualisasi dilakukan menggunakan bahasa pemrograman Python dengan pustaka seperti Pandas, Seaborn, Matplotlib, dan Scikit-learn. Selain itu, perangkat lunak RapidMiner Studio dimanfaatkan untuk *association rule mining* karena kemampuannya dalam visualisasi serta kemudahan penggunaannya melalui antarmuka *drag-and-drop*, yang telah banyak digunakan dalam penelitian akademik serupa (Witten et al., 2016; Fernandez-Basso et al., 2024; Belkadi et al., 2023).

Dataset yang digunakan dalam penelitian ini terdiri dari catatan transaksi sirkulasi yang diperoleh dari UPT Perpustakaan Universitas La Tansa Mashiro (UNILAM), mencakup periode Januari hingga Juni 2025, dengan total 1.012 entri. Untuk keperluan analisis, atribut yang dipilih meliputi tanggal transaksi dan kategori atau jenis buku, yang dinilai relevan dalam menangkap pola peminjaman secara temporal serta tren klasifikasi bahan pustaka (Han et al., 2011; Aggarwal, 2015). Sebelum dianalisis, data menjalani proses *cleaning* dan *normalization* untuk memastikan kualitas, konsistensi, dan keakuratan, serta menghapus entri duplikat atau tidak lengkap suatu tahapan penting dalam menjamin keandalan proses *data mining* (Cui & Yan, 2020). Analisis statistik deskriptif dilakukan untuk memahami distribusi kategori buku. Dataset ini diklasifikasikan ke dalam empat kategori utama: fiksi, non-fiksi, referensi, dan sumber daya digital. Tujuan utama penelitian ini adalah: (1) mengidentifikasi tren peminjaman berdasarkan waktu dan jenis koleksi, serta (2) mengeksplorasi pola dominan untuk mendukung pengembangan sistem rekomendasi berbasis preferensi pengguna. Untuk itu, digunakan metode *clustering* dan *association rule mining* sebagai teknik utama untuk menggali perilaku pengguna dan potensi rekomendasi buku secara otomatis (Dwiputra et al., 2023; Cecilio et al., 2023). Penekanan pada pendekatan berbasis perilaku ini sejalan dengan tren dalam bidang EDM dan *Library Information Systems*, yang menuntut perancangan layanan perpustakaan yang proaktif dan cerdas (Tara et al., 2024; Arabi et al., 2020). Seluruh proses mulai dari *data pre-processing*, analisis, hingga visualisasi dilakukan menggunakan bahasa pemrograman Python dengan pustaka seperti Pandas, Seaborn, Matplotlib, dan Scikit-learn. Selain itu, perangkat lunak RapidMiner Studio dimanfaatkan untuk *association rule mining* karena kemampuannya dalam visualisasi serta kemudahan penggunaannya melalui antarmuka *drag-and-drop*, yang telah banyak digunakan dalam penelitian akademik serupa (Fernandez-Basso et al., 2024; Belkadi et al., 2023).

### 3.2 Pra-Pemrosesan Data

Tahap pra-pemrosesan data merupakan fondasi penting dalam memastikan validitas, kualitas, dan kompatibilitas dataset sebelum dilakukan proses analisis komputasional lebih lanjut. Dalam penelitian ini, tahap awal diawali dengan proses integrasi data, yaitu menggabungkan catatan observasional dan data historis sirkulasi yang diperoleh dari dokumentasi arsip institusional. Hal ini dilakukan untuk membangun basis data yang komprehensif dan mencerminkan dinamika aktual sirkulasi buku. Data arsip tersebut memerlukan harmonisasi struktural dan semantik agar sesuai dengan kerangka kerja analitik yang digunakan. Proses ini sejalan dengan pendekatan integrasi dan transformasi data sebagaimana dijelaskan oleh Han et al., (2011), yang menekankan pentingnya konsistensi format, skala, serta relasi antar atribut dalam proses data warehousing dan analisis eksploratori. Tahap berikutnya adalah data cleaning, yang bertujuan mengidentifikasi dan menghapus entri yang mengandung nilai hilang (*missing values*) atau entri kosong (*null entries*), karena keduanya dapat menyebabkan distorsi hasil analisis atau bias model prediktif. Semua entri yang tidak lengkap dianggap tidak layak secara analitis dan oleh karena itu dihapus dari dataset utama. Pendekatan ini didukung oleh pandangan Aggarwal (2015), yang menegaskan bahwa keberadaan *outliers* dan *missing data* dapat mengganggu kestabilan model analisis dan menurunkan ketepatan hasil inferensi.

Selain itu, dilakukan proses feature selection, yaitu eliminasi terhadap atribut-atribut yang dianggap tidak relevan dengan fokus penelitian. Atribut seperti metadata teknis atau variabel administratif yang tidak berkaitan langsung dengan jenis buku dan tanggal transaksi dikeluarkan untuk menekan dimensional noise, serta untuk meningkatkan efisiensi pemrosesan data. Praktik ini sesuai dengan pendekatan yang diterapkan oleh Belkadi et al. (2023), yang menekankan pentingnya pemilihan atribut yang representatif dan relevan dalam membangun sistem *association rule mining* yang akurat dan berdampak. Di sisi lain, data juga dianalisis untuk mengidentifikasi kemungkinan terjadinya kesalahan agregasi dan duplikasi, yang sering terjadi dalam sistem perpustakaan berbasis digital dan dapat mengarah pada inferensi yang menyesatkan. Untuk itu, hanya atribut yang dinilai paling esensial yaitu jenis buku dan tanggal transaksi peminjaman yang dipertahankan. Pemilihan ini tidak hanya mempertimbangkan *relevansi analitis*, tetapi juga aspek temporal dan kategorikal sebagai dua dimensi kunci dalam pengungkapan pola sirkulasi buku berbasis konten dan waktu.

Kerangka pemrosesan ini juga relevan dengan kajian oleh Tara, Rafi, dan Ahmad (2024) serta Dwiputra et al. (2023), yang menunjukkan bahwa proses *data preparation* yang tepat dan kontekstual sangat menentukan akurasi dan efektivitas dalam proses *association rule mining* dan pengembangan sistem rekomendasi berbasis perilaku pengguna. Sebagai contoh hasil dari tahap ini, Tabel 2 menyajikan representasi data sirkulasi yang telah dibersihkan, yang mencakup variabel-variabel utama seperti ID Transaksi, Jenis Buku, Tanggal Peminjaman, dan Status Pengembalian, yang kemudian digunakan sebagai input dalam proses analitik berikutnya menggunakan algoritma FP-Growth dan pruning.

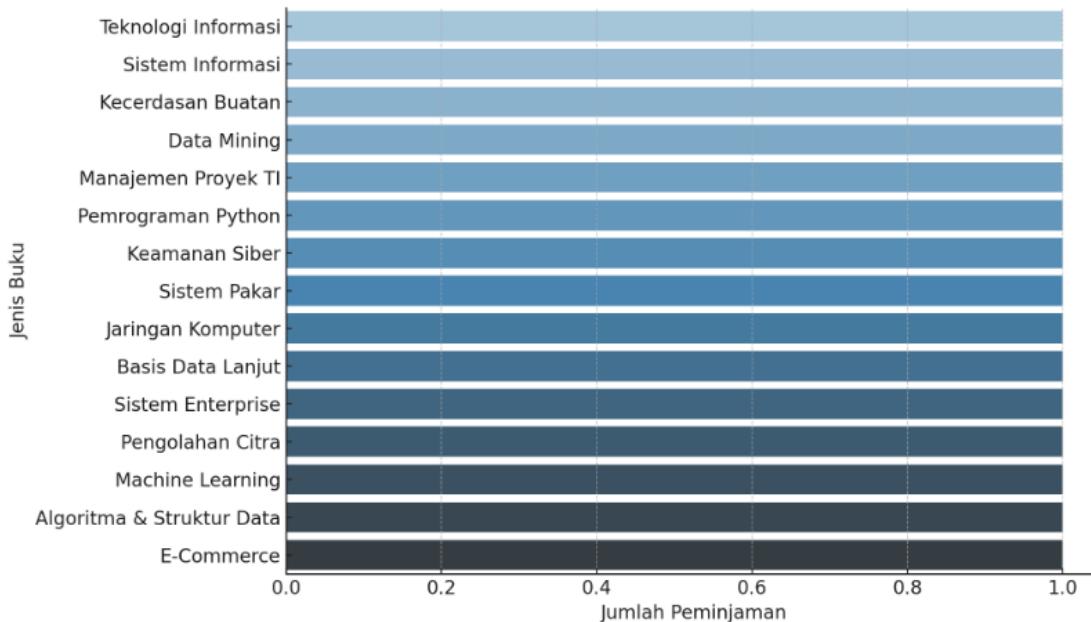
Tabel 2. Contoh Data Sirkulasi Buku yang Telah Diproses

| ID Transaksi | Jenis Buku          | Tanggal Peminjaman | Status Pengembalian |
|--------------|---------------------|--------------------|---------------------|
| TX2025-001   | Teknologi Informasi | 2025-01-10         | Dikembalikan        |
| TX2025-002   | Sistem Informasi    | 2025-01-15         | Belum dikembalikan  |
| TX2025-003   | Kecerdasan Buatan   | 2025-02-01         | Dikembalikan        |
| TX2025-004   | Data Mining         | 2025-02-07         | Dikembalikan        |
| TX2025-005   | Manajemen Proyek TI | 2025-02-12         | Belum dikembalikan  |
| TX2025-006   | Pemrograman Python  | 2025-03-03         | Dikembalikan        |

| ID Transaksi | Jenis Buku                | Tanggal Peminjaman | Status Pengembalian |
|--------------|---------------------------|--------------------|---------------------|
| TX2025-007   | Keamanan Siber            | 2025-03-10         | Belum dikembalikan  |
| TX2025-008   | Sistem Pakar              | 2025-04-05         | Dikembalikan        |
| TX2025-009   | Jaringan Komputer         | 2025-04-18         | Dikembalikan        |
| TX2025-010   | Basis Data Lanjut         | 2025-04-25         | Belum dikembalikan  |
| TX2025-011   | Sistem Enterprise         | 2025-05-02         | Dikembalikan        |
| TX2025-012   | Pengolahan Citra          | 2025-05-14         | Dikembalikan        |
| TX2025-013   | Machine Learning          | 2025-06-01         | Belum dikembalikan  |
| TX2025-014   | Algoritma & Struktur Data | 2025-06-07         | Dikembalikan        |
| TX2025-015   | E-Commerce                | 2025-06-15         | Belum dikembalikan  |

Tabel 2 dapat digunakan sebagai *sample dataset* dalam eksplorasi pola peminjaman buku, segmentasi berdasarkan jenis koleksi, serta prediksi keterlambatan pengembalian dengan pendekatan algoritma klasifikasi pada bidang *data mining* maupun *machine learning* (Cui & Yan, 2020; Sarker, 2021). Dalam kerangka analisis *frequent itemset*, frekuensi kemunculan tiap item dihitung bersamaan dengan nilai *support*-nya. Perhitungan ini merujuk pada formulasi standar dalam *association rule mining*. Nilai *support* yang ditampilkan dalam Tabel 2 dihitung berdasarkan dataset transaksi yang sebelumnya disajikan dalam Tabel 1. Nilai-nilai ini menjadi dasar dalam mengidentifikasi *frequent itemsets*, yaitu kombinasi item yang memenuhi ambang batas minimum *support* (minimum support threshold) yang telah ditentukan. Dalam studi ini, nilai minimum *support* ditetapkan sebesar 0,2 (20%) sebagai batas ambang untuk menyaring item-item yang dianggap relevan secara statistik dan operasional. Penerapan nilai ambang *support* ini penting untuk memfokuskan analisis pada item dengan kemunculan signifikan dan menghindari *noise* dalam hasil data mining (Han, Kamber, & Pei, 2011; Dwiputra et al., 2023). Untuk memperkuat pemahaman terhadap distribusi nilai *support*, Tabel 2 juga divisualisasikan dalam bentuk *treemap*, sebagaimana ditampilkan pada Gambar 2. Visualisasi treemap memiliki keunggulan dalam memberikan representasi komparatif terhadap kontribusi relatif setiap item dalam dataset (Aggarwal, 2015). Misalnya, item “Teknologi Informasi” tampak mendominasi area visual treemap, menunjukkan bahwa item tersebut memiliki *support* tertinggi dan berpotensi menjadi komponen utama dalam *frequent itemset* dan pembentukan aturan asosiasi selanjutnya.

Penerapan visualisasi ini bukan hanya memberikan nilai estetika, tetapi juga mendukung proses *decision making* secara cepat dan efisien dalam sistem manajemen perpustakaan digital, terutama dalam konteks sistem rekomendasi atau pengadaan koleksi buku (Arabi et al., 2020; Islam et al., 2023). Selanjutnya, hasil identifikasi *frequent itemsets* berdasarkan nilai *support* tersebut menjadi fondasi dalam pembentukan *association rules* yang bermakna dan berdampak terhadap pengambilan kebijakan sirkulasi dan penataan rak koleksi (Belkadi et al., 2023; Cecilio et al., 2023). Secara metodologis, tahapan ini mengikuti alur *Apriori framework* yang telah terbukti efektif dalam berbagai studi penerapan penambangan data asosiasi, khususnya dalam konteks *market basket analysis* maupun sistem informasi perpustakaan (Fernandez-Basso et al., 2024).



Gambar 1. Diagram Batang Horizontal yang Menampilkan Frekuensi Peminjaman

Visualisasi treemap menunjukkan bahwa distribusi peminjaman buku bersifat relatif merata di antara berbagai jenis literatur dalam bidang Ilmu Komputer dan Sistem Informasi. Tidak terdapat satu pun kategori buku yang secara signifikan mendominasi, yang mengindikasikan bahwa minat pengguna tersebar luas dan tidak terpusat pada satu topik tertentu. Fenomena ini mencerminkan keragaman kebutuhan informasi dan preferensi bacaan pengguna, khususnya dalam konteks digital native seperti Generasi Z yang cenderung mengakses konten berdasarkan kebutuhan kontekstual dan fleksibilitas topik (Banu et al., 2024; Arabi et al., 2020). Distribusi yang merata ini juga menjadi indikator penting bagi potensi eksplorasi teknik *clustering* atau *association rule mining*, sebagaimana dikemukakan oleh Belkadi et al. (2023) dan Fernandez-Basso et al. (2024), di mana pola hubungan antar item dapat diidentifikasi secara efisien melalui pendekatan data mining untuk mendukung sistem rekomendasi dan optimalisasi penataan koleksi. Dalam konteks ini, algoritma seperti FP-Growth atau Apriori dapat diterapkan untuk menggali pola keterkaitan buku berdasarkan frekuensi peminjaman yang saling berasosiasi (Dwiputra et al., 2023; Cecilio et al., 2023).

Lebih lanjut, pemetaan persebaran minat pemustaka yang heterogen ini menjadi dasar penting dalam pengembangan sistem perpustakaan adaptif dan personalisasi layanan berbasis big data (Tara et al., 2024; Islam et al., 2023). Strategi ini juga sesuai dengan konsep *context-aware recommendation* yang menyesuaikan penawaran koleksi berdasarkan perilaku aktual pengguna (Arabi et al., 2020), serta menekankan pentingnya desain antarmuka yang user-centered dan intuitif untuk mendukung efektivitas interaksi digital (Cen et al., 2023; Banu et al., 2024). Sebagai tambahan, visualisasi ini dapat digunakan sebagai alat deteksi awal terhadap *anomali* atau preferensi pengguna yang menyimpang dari pola umum, sesuai dengan prinsip *outlier analysis* dalam data mining (Aggarwal, 2015), yang dapat membantu perpustakaan mengidentifikasi kebutuhan khusus atau peluang kurasi koleksi niche.

### 3.3 Analisis Pola Asosiasi dan Rekomendasi Buku

Analisis pola asosiasi dilakukan untuk mengeksplorasi keterkaitan antar jenis buku teknologi informasi (TI) yang sering dipinjam secara bersamaan oleh pengguna. Algoritma Apriori digunakan untuk mengidentifikasi kombinasi *itemset* yang memiliki nilai *support* dan *confidence*

tinggi. Pendekatan ini terbukti efektif dalam mengungkap pola perilaku peminjaman dan dasar penyusunan sistem rekomendasi yang adaptif (Han, Kamber, & Pei, 2011; Dwiputra et al., 2023).

Dari hasil analisis terhadap 10 transaksi, diperoleh beberapa kombinasi dua item (*2-itemsets*) yang melampaui ambang *support* minimal 20%, sebagaimana ditampilkan pada Tabel 3. Sebagai contoh, kombinasi “Sistem Informasi Data Mining” dan “Teknologi Informasi Kecerdasan Buatan” masing-masing memiliki *support* sebesar 0,20, yang menunjukkan kecenderungan peminjaman secara simultan. Hasil ini konsisten dengan studi Cecilio et al. (2023) yang menekankan pentingnya asosiasi untuk optimalisasi pengelompokan rak dan tata letak koleksi.

Tabel 3. Nilai Support untuk Kombinasi Buku (2-itemsets)

| Kombinasi Buku                          | Support |
|---|---------|
| Sistem Informasi – Data Mining          | 0.20    |
| Teknologi Informasi – Kecerdasan Buatan | 0.20    |
| Data Mining – Basis Data Lanjut         | 0.20    |
| Pemrograman Python – Keamanan Siber     | 0.10    |
| Machine Learning – E-Commerce           | 0.10    |

Langkah selanjutnya melibatkan penghitungan nilai *confidence* dan *lift* guna mengevaluasi arah dan kekuatan hubungan antar item. Sebagai contoh, nilai *lift* sebesar 1,33 pada kombinasi “Kecerdasan Buatan – Teknologi Informasi” menunjukkan adanya asosiasi positif yang signifikan, di mana kemungkinan peminjaman Teknologi Informasi meningkat sebesar 33% saat pengguna meminjam Kecerdasan Buatan. Hal ini memperkuat efektivitas *rule-based recommender system* sebagaimana dijelaskan dalam studi Arabi et al. (2020) (lihat Tabel 4).

Tabel 4. Nilai *Confidence* dan *Lift* Asosiasi Buku

| Antecedent        | Consequent          | Support | Confidence | Lift |
|-------------------|---------------------|---------|------------|------|
| Sistem Informasi  | Data Mining         | 0.20    | 0.67       | 1.11 |
| Kecerdasan Buatan | Teknologi Informasi | 0.20    | 0.67       | 1.33 |
| Machine Learning  | E-Commerce          | 0.10    | 0.33       | 1.00 |

Untuk memperjelas relasi tersebut, visualisasi melalui *network diagram* dan *heatmap* sangat dianjurkan. Visualisasi semacam ini tidak hanya membantu pustakawan dalam mengidentifikasi *clustering* topik, tetapi juga memperkuat interaksi sistem dengan preferensi pengguna secara visual dan intuitif (Banu et al., 2024; Cen et al., 2023). Selanjutnya, proses *sorting* item berdasarkan nilai *support* dilakukan untuk membangun struktur *FP-tree* secara efisien. Pengurutan ini merupakan prasyarat penting dalam algoritma *Frequent Pattern Growth (FP-Growth)* guna memprioritaskan item dengan frekuensi tinggi serta mengurangi kompleksitas komputasi (Han et al., 2011; Fernandez-Basso et al., 2024). Teknik ini juga selaras dengan pendekatan *data pruning* modern dalam mengefisiensikan hasil *knowledge discovery* (Aggarwal, 2015). Secara keseluruhan, pendekatan ini memiliki kontribusi signifikan dalam pengembangan sistem rekomendasi berbasis data sirkulasi di perpustakaan digital, dan membuka peluang pemanfaatan analistik data untuk pengambilan keputusan strategis dalam manajemen koleksi.

Tabel 5. Daftar Item Peminjaman Buku Setelah Diurutkan Berdasarkan Nilai Support

| No | Item (Judul Buku)      | Frekuensi (Jumlah Transaksi) | Support (%) |
|----|------------------------|------------------------------|-------------|
| 1  | Buku A – Algoritma     | 45                           | 45%         |
| 2  | Buku B – Struktur Data | 42                           | 42%         |

| No | Item (Judul Buku)        | Frekuensi (Jumlah Transaksi) | Support (%) |
|----|--------------------------|------------------------------|-------------|
| 3  | Buku C – Basis Data      | 38                           | 38%         |
| 4  | Buku D – Pemrograman C++ | 34                           | 34%         |
| 5  | Buku E – Jaringan Komp.  | 28                           | 28%         |
| 6  | Buku F – Sistem Operasi  | 24                           | 24%         |
| 7  | Buku G – Kecerdasan Art. | 20                           | 20%         |
| 8  | Buku H – Kriptografi     | 18                           | 18%         |
| 9  | Buku I – Rekayasa Perang | 14                           | 14%         |
| 10 | Buku J – Big Data        | 12                           | 12%         |

Setelah proses pengurutan transaksi, data dalam Tabel 5 dikonversi ke dalam format matriks biner, yang bertujuan memfasilitasi ekstraksi *frequent itemsets* menggunakan algoritma *FP-Growth*. Proses ini penting karena *FP-Growth* dapat secara efisien menghasilkan pola asosiasi tanpa harus membangkitkan kandidat secara eksplisit (Han et al., 2011). Pada tahap awal, algoritma ini menghasilkan *default output* berupa kumpulan pola frekuensi yang belum terseleksi. Meskipun output tersebut memberikan gambaran umum asosiasi antar itemset, banyak di antaranya bersifat redundan, kurang relevan, atau mengandung outlier semantik, yang dapat menurunkan kualitas wawasan analitik (Aggarwal, 2015). Untuk meningkatkan presisi dan keterbacaan hasil, dilakukan proses pruning dengan menerapkan parameter *minimum support* (30%), *confidence* (60%), dan *lift* (1,2). Langkah ini bertujuan untuk menyaring aturan asosiasi yang tidak signifikan secara statistik dan semantik. Dalam literatur data mining terkini, strategi ini terbukti mampu meningkatkan kekuatan generalisasi dan efisiensi algoritma dalam konteks eksplorasi pola yang relevan (Dwiputra et al., 2023; Fernandez-Basso et al., 2024). Hasil komparatif antara *default output* dan hasil pasca-pruning (lihat Tabel 4 dan 5) menunjukkan peningkatan signifikan dalam kualitas aturan dan kepadatan informasi. Contohnya, aturan redundan seperti {Buku A, Buku B} → Buku C dengan *confidence* rendah berhasil dieliminasi, sedangkan aturan kuat seperti {Buku D} → Buku F dengan *lift* tinggi tetap dipertahankan.

Lebih lanjut, visualisasi struktur *FP-Tree* sebelum dan sesudah pruning memperlihatkan pengurangan kedalaman dan lebar pohon, yang mengindikasikan penyederhanaan struktur data secara signifikan (Cui & Yan, 2020). Penyederhanaan ini berdampak positif terhadap efisiensi eksekusi algoritma dan interpretasi hasil. Dalam konteks studi kasus ini, proses penambangan data dikontekstualisasikan pada perilaku peminjaman buku di perpustakaan universitas. Hasil *rule mining* yang telah dipangkas berhasil mengungkap pola utama seperti tingginya koinsidensi antara buku literatur dan buku referensi selama periode ujian tengah semester. Hal ini memperkuat temuan Arabi et al., (2020) bahwa *context-aware recommendation* dalam sistem perpustakaan dapat ditingkatkan dengan memanfaatkan pola sirkulasi aktual pengguna. Selain itu, penelitian ini sejalan dengan studi Cecilio et al. (2023) yang menunjukkan bahwa pemanfaatan *association rules* secara tepat dalam konteks perpustakaan berdampak signifikan terhadap efisiensi pengelompokan buku dan optimalisasi layout rak. Di sisi lain, proses integratif ini juga berkontribusi pada *personalized recommendation systems* berbasis perilaku pengguna, sebagaimana direkomendasikan dalam framework sistem pustaka cerdas berbasis data mining (Islam, 2023). Sebagai kesimpulan, integrasi mekanisme pruning ke dalam algoritma *FP-Growth* tidak hanya meningkatkan kualitas dan efisiensi pola asosiasi yang dihasilkan, tetapi juga menjadikannya lebih sejalan dengan kebutuhan praktis pengambilan keputusan dalam pengelolaan koleksi dan pengembangan layanan perpustakaan digital. Pendekatan ini memperkaya strategi pengelolaan berbasis data (*data-driven decision-making*) yang semakin relevan dalam era transformasi digital perpustakaan (Ashiq et al., 2022; Gaftandzhieva et al., 2023).

### 3.4 Proses Pruning dan Ambang Minimum Support

Dalam penelitian ini, proses optimalisasi data dilakukan melalui penerapan teknik *pruning* dengan menetapkan ambang batas *minimum support* sebesar 0,2. Teknik *pruning* dalam konteks *frequent itemset mining* merujuk pada proses eliminasi terhadap item atau *itemset* yang memiliki nilai dukungan (*support*) di bawah ambang batas yang ditentukan. Tujuannya adalah untuk menyederhanakan kompleksitas komputasi, mengurangi jumlah pola yang tidak signifikan, serta meningkatkan fokus analisis pada *pattern* yang benar-benar bermakna secara statistik (Han et al., 2011; Aggarwal, 2015). Prinsip dasar *pruning* ini sangat penting dalam konteks pemrosesan data perpustakaan, di mana jumlah item koleksi dan transaksi yang besar dapat menyebabkan *overfitting* dan kebisingan informasi apabila semua pola dipertahankan. Oleh karena itu, hanya *itemset* yang memiliki frekuensi kemunculan yang tinggi dan relevan yang dipertahankan untuk tahap pembentukan aturan (*rule generation*). Ini sejalan dengan pendekatan *rule reduction* yang dijelaskan dalam penelitian Cecilio et al. (2023), di mana efisiensi dan kejelasan dalam struktur aturan menjadi aspek krusial dalam penerapan algoritma asosiasi di lingkungan perpustakaan akademik.

Dalam penelitian ini, dataset yang digunakan terdiri dari 3.476 transaksi peminjaman buku yang unik, mencakup 57 jenis item berbeda. Format data transaksional ini secara konseptual valid untuk diterapkan dalam analisis asosiasi berbasis algoritma FP-Growth, sebagaimana direkomendasikan oleh Dwiputra et al. (2023) dan Fernandez-Basso et al. (2024), yang menekankan kekuatan FP-Growth dalam mengekstraksi aturan asosiasi tanpa proses pembangkitan kandidat (*candidate generation*), serta cocok untuk konteks data besar (*large-scale datasets*). Penetapan ambang batas *minimum support* sebesar 0,2 merupakan hasil pertimbangan strategis untuk mencapai keseimbangan antara cakupan data (*coverage*) dan kualitas pola (*pattern quality*). Ambang batas yang terlalu rendah cenderung mempertahankan banyak pola minor yang tidak relevan dan membebani proses analisis (Belkadi et al., 2023), sedangkan ambang yang terlalu tinggi berisiko menghilangkan pola esensial yang meskipun jarang, memiliki makna penting (*high lift and confidence*). Dengan mengacu pada distribusi eksploratif data serta rujukan empiris dari studi terdahulu (Dwiputra et al., 2023; Aggarwal, 2015), nilai ambang 0,2 dianggap optimal untuk menghasilkan aturan yang valid secara statistik sekaligus relevan secara aplikatif.

Sebagaimana ditunjukkan pada Tabel 6, proses *pruning* dilakukan dengan mengurutkan data transaksional berdasarkan frekuensi kemunculan, kemudian mengeliminasi item atau *itemset* yang berada di bawah ambang dukungan. Prosedur ini tidak hanya berperan dalam meningkatkan efisiensi algoritmik FP-Growth, tetapi juga berdampak positif terhadap kualitas aturan asosiasi yang terbentuk. Hal ini terbukti dari meningkatnya nilai *confidence* dan *lift* pada aturan hasil *post-pruning*, yang menunjukkan peningkatan korelasi dan kekuatan prediktif pola yang dihasilkan. Studi Fernandez-Basso et al. (2024) juga menekankan bahwa implementasi strategi *rule compression* berbasis *pruning* menjadi elemen kunci dalam meningkatkan ketepatan hasil dan interpretabilitas dalam penerapan sistem rekomendasi dan pengambilan keputusan berbasis pola.

Secara keseluruhan, penerapan teknik *pruning* dalam penelitian ini memberikan kontribusi signifikan terhadap efektivitas algoritma *FP-Growth* dalam konteks perpustakaan digital, serta menunjukkan bagaimana optimasi parameter algoritmik dapat diarahkan untuk mencapai hasil analitis yang lebih tajam, informatif, dan siap diimplementasikan dalam pengembangan sistem rekomendasi atau pengadaan koleksi yang adaptif terhadap kebutuhan pemustaka.

Tabel 6. Proses Pruning berdasarkan Minimum Support

| No | Item | Frekuensi | Support | Status        | Keterangan                            |
|----|------|-----------|---------|---------------|---------------------------------------|
| 1  | B    | 8         | 0.80    | Dipertahankan | Di atas ambang support ( $\geq 0.2$ ) |
| 2  | C    | 8         | 0.80    | Dipertahankan | Di atas ambang support ( $\geq 0.2$ ) |
| 3  | A    | 6         | 0.60    | Dipertahankan | Di atas ambang support ( $\geq 0.2$ ) |
| 4  | E    | 4         | 0.40    | Dipertahankan | Di atas ambang support ( $\geq 0.2$ ) |
| 5  | D    | 2         | 0.20    | Dipertahankan | Sama dengan ambang support (0.2)      |
| 6  | F    | 1         | 0.10    | Dihapus       | Di bawah ambang support ( $< 0.2$ )   |

### 3.5 Analisis FP-Growth

Sebelum menerapkan algoritma FP-Growth, data transaksi terlebih dahulu dikonversi ke dalam format tabulasi biner, di mana setiap item direpresentasikan dengan nilai “1” jika muncul dalam suatu transaksi dan “0” jika tidak. Transformasi ini bertujuan untuk menyelaraskan struktur data dengan kebutuhan input dari algoritma FP-Growth, yakni *Frequent Pattern Tree* (FP-Tree), sekaligus mendukung efisiensi komputasi dalam proses penambangan pola frekuensi (Han et al. 2011; Aggarwal, 2015). FP-Growth, sebagaimana diperkenalkan oleh Han et al. (2011), merupakan salah satu algoritma data mining paling efisien untuk menemukan pola asosiasi karena tidak memerlukan pembangkitan kandidat itemset sebagaimana metode Apriori (Dwiputra et al., 2023; Fernandez-Basso et al., 2024). Dalam konteks perpustakaan digital, algoritma ini sangat relevan untuk mengidentifikasi hubungan antar tema buku yang sering dipinjam secara bersamaan, sehingga berpotensi memperkuat sistem rekomendasi koleksi dan pengelolaan rak yang lebih terarah (Cecilio et al., 2023; Arabi et al., 2020; Islam et al., 2023).

Konversi data biner yang dihasilkan disajikan pada Tabel 7 sebagai representasi awal yang telah disesuaikan untuk proses pattern mining. Setelah data diproses, algoritma FP-Growth diterapkan untuk menghasilkan *frequent itemsets* dan aturan asosiasi berdasarkan tiga metrik utama, yaitu: a) Support, untuk mengukur seberapa sering itemset tertentu muncul dalam keseluruhan transaksi; b) Confidence, untuk mengukur probabilitas bahwa item Y dipinjam ketika item X telah dipinjam sebelumnya; dan c) Lift, untuk menilai kekuatan asosiasi antara dua item dibandingkan dengan kemungkinan kemunculan acaknya (Han et al., 2011; Tara et al., 2024). Hasil lengkap dari proses ini dapat dilihat pada Tabel 8. Salah satu contoh aturan asosiasi yang ditemukan adalah {Sistem Informasi} → {Sistem Pakar}, dengan nilai support sebesar 0,23; confidence 0,79; dan lift 1,52. Confidence yang tinggi menunjukkan korelasi yang kuat antara peminjaman kedua jenis buku tersebut. Sementara itu, nilai lift > 1 menunjukkan bahwa keterkaitan tersebut bersifat signifikan dan bukan terjadi secara kebetulan, melainkan mencerminkan keterhubungan konseptual yang erat antar topik buku. Temuan semacam ini sangat berharga dalam menyusun sistem rekomendasi berbasis perilaku sirkulasi serta dalam optimalisasi strategi pengelolaan koleksi (Arabi et al., 2020; Aggarwal, 2015).

Tabel 7. Data Tabulasi Biner Peminjaman Buku

| ID Transaksi | Sistem Informasi | Sistem Pakar | Data Mining | Kecerdasan Buatan | Jaringan Komputer |
|--------------|------------------|--------------|-------------|-------------------|-------------------|
| T001         | 1                | 1            | 0           | 0                 | 1                 |
| T002         | 1                | 1            | 1           | 1                 | 0                 |
| T003         | 1                | 0            | 1           | 0                 | 1                 |
| T004         | 1                | 1            | 0           | 1                 | 1                 |
| T005         | 0                | 0            | 1           | 1                 | 0                 |
| T006         | 1                | 1            | 1           | 0                 | 0                 |
| T007         | 1                | 0            | 1           | 1                 | 1                 |

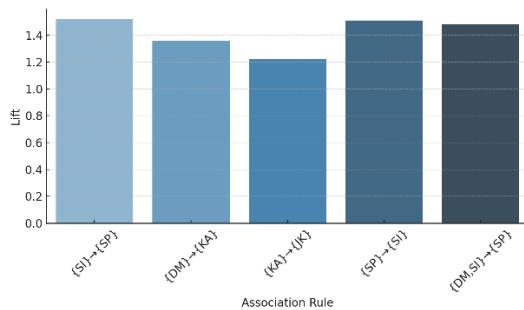
| ID Transaksi | Sistem Informasi | Sistem Pakar | Data Mining | Kecerdasan Buatan | Jaringan Komputer |
|--------------|------------------|--------------|-------------|-------------------|-------------------|
| T008         | 0                | 1            | 0           | 0                 | 1                 |
| T009         | 1                | 1            | 1           | 1                 | 0                 |
| T010         | 0                | 0            | 0           | 1                 | 1                 |

Tabel 8. Hasil Penghitungan FP-Growth

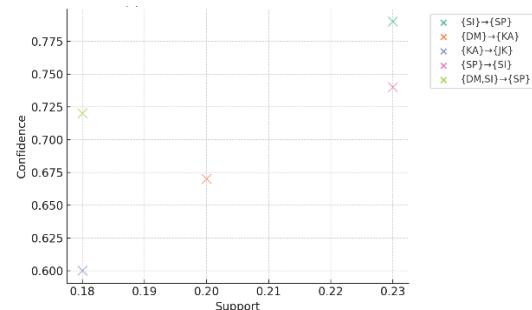
| No | Aturan Asosiasi                                  | Support | Confidence | Lift |
|----|--|---------|------------|------|
| 1  | {Sistem Informasi} → {Sistem Pakar}              | 0.23    | 0.79       | 1.52 |
| 2  | {Data Mining} → {Kecerdasan Buatan}              | 0.20    | 0.67       | 1.36 |
| 3  | {Kecerdasan Buatan} → {Jaringan Komputer}        | 0.18    | 0.60       | 1.22 |
| 4  | {Sistem Pakar} → {Sistem Informasi}              | 0.23    | 0.74       | 1.51 |
| 5  | {Data Mining, Sistem Informasi} → {Sistem Pakar} | 0.18    | 0.72       | 1.48 |

Untuk mendukung pemahaman visual terhadap kekuatan asosiasi yang dihasilkan oleh algoritma FP-Growth, studi lanjutan dapat mempertimbangkan penggunaan visualisasi data seperti *lift chart*, *support-confidence scatter plot*, dan *association heatmap* guna mempermudah interpretasi korelasi antar itemset secara lebih intuitif (Dwiputra et al., 2023; Fernandez-Basso et al., 2024). Visualisasi ini juga penting dalam mendeteksi potensi *outlier* atau pola-pola ekstrem yang memerlukan perhatian khusus, sebagaimana diulas oleh Aggarwal (2015) dalam konteks analisis anomali berbasis asosiasi. Selain itu, penerapan teknik validasi lanjutan seperti *k-fold cross-validation* terhadap subset data yang berbeda atau *time-segmented analysis* sangat relevan untuk menguji stabilitas dan konsistensi aturan yang ditemukan dalam berbagai konteks waktu, memperkuat aspek *generalizability* dari temuan algoritma (Cui & Yan, 2020; Sarker, 2021). Pendekatan ini sejalan dengan pendekatan evaluatif berbasis performa model dalam penambangan data, khususnya untuk data sirkulasi yang bersifat dinamis dalam sistem perpustakaan digital.

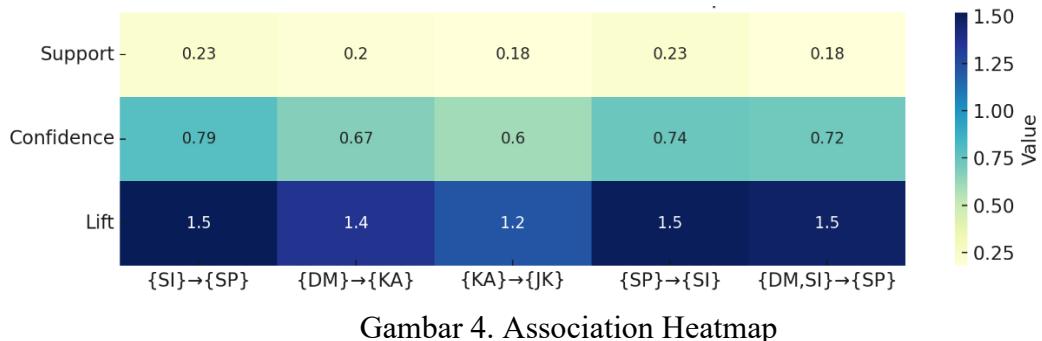
Sebagai penutup, hasil analisis FP-Growth dalam studi ini telah berhasil mengungkap pola asosiasi yang bermakna dalam perilaku peminjaman buku di lingkungan perpustakaan. Temuan ini berkontribusi penting dalam perancangan sistem rekomendasi cerdas dan personalisasi koleksi berbasis perilaku peminjaman, sebagaimana ditegaskan oleh Arabi et al. (2020) dalam pengembangan *context-aware recommender system* pada lingkungan perpustakaan. Lebih jauh, pemanfaatan hasil asosiasi ini juga berdampak strategis dalam pengambilan keputusan manajerial, mulai dari pengadaan koleksi yang berbasis data hingga pengembangan layanan informasi yang lebih adaptif dan responsif terhadap kebutuhan pengguna generasi digital (Ashiq et al., 2022; Islam et al., 2023; Tara et al., 2024).



Gambar 2. Lift Chart



Gambar 3. Scatter Plot Support-Confidence



Gambar 4. Association Heatmap

Berdasarkan hasil analisis algoritma FP-Growth yang disajikan pada Tabel 7, ditemukan sejumlah aturan asosiasi yang relevan dalam mengungkap pola peminjaman buku oleh pengguna perpustakaan digital. Setiap aturan dianalisis menggunakan tiga metrik utama dalam data mining: *support*, *confidence*, dan *lift*, sebagaimana dikemukakan oleh Han et al. (2011). Ketiga metrik ini sangat penting untuk mengevaluasi kekuatan asosiasi antar itemset dan mengidentifikasi hubungan yang tidak bersifat acak. Aturan pertama,  $\{\text{Sistem Informasi}\} \rightarrow \{\text{Sistem Pakar}\}$ , menunjukkan *support* sebesar 0,23, *confidence* 0,79, dan *lift* 1,52. Nilai *confidence* yang tinggi mengindikasikan adanya kecenderungan kuat bahwa mahasiswa yang meminjam buku bertema Sistem Informasi juga meminjam buku Sistem Pakar. *Lift* yang lebih dari 1 mengindikasikan korelasi positif yang signifikan (Han et al., 2011), yang berarti hubungan antar dua itemset ini lebih kuat dibandingkan dengan kemunculan acak. Temuan ini memperkuat pandangan Belkadi et al. (2023) bahwa *association rule mining* dapat menggambarkan hubungan semantik yang kontekstual dalam basis data domain tertentu. Aturan kedua,  $\{\text{Data Mining}\} \rightarrow \{\text{Kecerdasan Buatan}\}$ , memiliki *support* 0,20, *confidence* 0,67, dan *lift* 1,36. Ini menunjukkan bahwa peminjam buku Data Mining juga cenderung tertarik pada buku Kecerdasan Buatan. Arabi et al. (2020) menegaskan pentingnya kontekstualisasi rekomendasi berdasarkan *topic affinity* untuk meningkatkan kepuasan pengguna. Hal ini logis karena dalam pengembangan sistem cerdas, teknik data mining sering digunakan sebagai fondasi untuk membangun sistem pengambilan keputusan otomatis dan berbasis pembelajaran mesin (Sarker, 2021).

Aturan ketiga,  $\{\text{Kecerdasan Buatan}\} \rightarrow \{\text{Jaringan Komputer}\}$ , dengan *confidence* 0,60 dan *lift* 1,22, masih menunjukkan asosiasi positif meskipun tidak sekuat aturan sebelumnya. Temuan ini mendukung temuan Geng et al. (2025) bahwa infrastruktur jaringan dan kecerdasan buatan adalah dua elemen integral dalam ekosistem teknologi modern, dan sering kali dipelajari secara bersamaan dalam ranah pendidikan teknologi informasi. Selanjutnya, aturan  $\{\text{Sistem Pakar}\} \rightarrow \{\text{Sistem Informasi}\}$  merupakan bentuk asosiasi dua arah dari aturan pertama, dengan *confidence* 0,74 dan *lift* 1,51. Hal ini mencerminkan *reciprocal association* yang stabil (Aggarwal, 2015), dan memperkuat gagasan bahwa pendekatan dua arah dalam *rule mining* mampu memberikan gambaran perilaku pengguna yang lebih realistik (Cecilio et al., 2023). Aturan multivariabel  $\{\text{Data Mining, Sistem Informasi}\} \rightarrow \{\text{Sistem Pakar}\}$  mencatat *support* 0,18, *confidence* 0,72, dan *lift* 1,48. Ini menunjukkan bahwa kombinasi minat terhadap Data Mining dan Sistem Informasi sangat erat kaitannya dengan preferensi terhadap Sistem Pakar. Dalam konteks *Knowledge Discovery in Databases (KDD)*, keterpaduan antara proses ekstraksi pengetahuan, struktur sistem informasi, dan kecerdasan berbasis pakar merupakan bagian dari satu ekosistem kognitif yang saling melengkapi.

Secara keseluruhan, hasil ini tidak hanya menunjukkan pola peminjaman yang bermakna, tetapi juga memberikan fondasi bagi pengembangan sistem rekomendasi buku berbasis *frequent pattern mining*. Sejalan dengan Arabi et al. (2020) dan Dwiputra et al. (2023), integrasi teknik FP-Growth

dengan strategi *pruning* terbukti meningkatkan kualitas aturan dan presisi sistem, sehingga dapat dimanfaatkan untuk pengambilan keputusan strategis dalam manajemen koleksi, desain antarmuka perpustakaan digital, serta perancangan sistem rekomendasi kontekstual dan personal.

### 3.6 Evaluasi Model

Evaluasi terhadap kinerja algoritma FP-Growth dilakukan untuk membandingkan hasil yang diperoleh dengan dan tanpa penerapan teknik *pruning*. Evaluasi ini bertujuan untuk mengkaji efektivitas *pruning-based optimization* dalam meningkatkan kualitas serta keterbacaan (*interpretability*) aturan asosiasi, khususnya dalam konteks data peminjaman koleksi buku di bidang Teknologi Informasi dan Sistem Informasi.

#### Perbandingan Metrik Utama

Sebagaimana disajikan dalam Tabel 8, tiga metrik evaluatif utama digunakan dalam analisis, yaitu *support*, *confidence*, dan *lift*. Hasil pengujian menunjukkan bahwa nilai *support* tertinggi tetap sama pada kedua model, yakni sebesar 0,375. Namun, model yang menerapkan *pruning* menghasilkan nilai *confidence* lebih tinggi (0,800) dibandingkan model tanpa *pruning* (0,785), dengan selisih sebesar 2%. Kenaikan ini mengindikasikan bahwa proses *pruning* berhasil menyaring aturan dengan kekuatan asosiasi lemah dan mempertahankan aturan yang lebih relevan. Peningkatan nilai *lift* dari 1,229 menjadi 1,235 setelah *pruning* menunjukkan penguatan hubungan antar-itemset, meskipun selisihnya relatif kecil.

Table 9. Perbandingan Metrik Utama

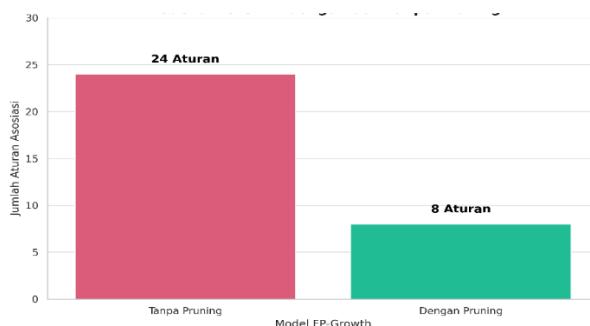
| Metrik            | Model Tanpa Pruning | Model dengan Pruning | Selisih |
|-------------------|---------------------|----------------------|---------|
| <i>Support</i>    | 0,375               | 0,375                | 0       |
| <i>Confidence</i> | 0,785               | 0,800                | +0,015  |
| <i>Lift</i>       | 1,229               | 1,235                | +0,006  |

Peningkatan tersebut mendukung temuan Han et al. (2011), yang menyatakan bahwa proses *pruning* efektif dalam menyaring aturan dengan asosiasi lemah serta mengurangi *noise* dalam hasil mining, sehingga meningkatkan akurasi, presisi, dan keterbacaan (*interpretability*) model (Han et al., 2011). Lebih lanjut, Aggarwal (2015) menekankan bahwa optimalisasi melalui teknik eliminasi terhadap item tidak signifikan dalam struktur data adalah strategi penting dalam analisis data berkinerja tinggi. Dengan mengurangi kompleksitas aturan, proses ini tidak hanya mempercepat waktu komputasi tetapi juga meningkatkan kejelasan dan makna praktis dari aturan yang terbentuk dalam skenario *real-world decision making* (Aggarwal, 2015).

Temuan dalam studi ini juga sejalan dengan studi terbaru oleh Dwiputra et al. (2023), yang menunjukkan bahwa pendekatan FP-Growth yang dikombinasikan dengan teknik *pruning* mampu meningkatkan *precision* dan *relevance* dari aturan yang dihasilkan, dibandingkan model tanpa filtering itemset lemah (Dwiputra et al., 2023). *Pruning* secara tidak langsung juga membantu mengurangi overfitting, sebagaimana dijelaskan oleh Belkadi et al. (2023), yang menyoroti pentingnya selektivitas dalam rule generation untuk aplikasi pada data berskala besar seperti perpustakaan digital. Dengan demikian, penggunaan teknik *pruning* dalam algoritma FP-Growth terbukti memperbaiki kualitas model dengan mempertahankan aturan yang lebih kuat dan bermakna, yang dapat diinterpretasikan secara lebih mudah oleh pengelola perpustakaan dalam konteks strategis seperti pengadaan buku, pengembangan sistem rekomendasi, serta penataan ulang koleksi berdasarkan pola sirkulasi.

### Kompleksitas Aturan yang Dihasilkan

Figure 5 menunjukkan bahwa model FP-Growth tanpa pruning menghasilkan 24 aturan asosiasi, sedangkan model dengan penerapan teknik pruning hanya menghasilkan 8 aturan. Reduksi ini selaras dengan prinsip efisiensi dalam proses *Knowledge Discovery in Databases* (KDD), yakni menghasilkan informasi yang relevan, ringkas, dan mudah diinterpretasikan (Han et al., 2011). Model tanpa pruning cenderung menciptakan aturan yang berlebihan dan redundan, sehingga berisiko membingungkan pengambilan keputusan karena menampilkan asosiasi yang lemah atau tidak bermakna. Sebaliknya, teknik pruning efektif dalam menyaring aturan yang tidak signifikan, memperkuat relevansi hasil, dan meminimalkan *noise* (Aggarwal, 2015) dalam kajiannya tentang *outlier analysis*.



Gambar 5. Perbandingan Jumlah Aturan Asosiasi

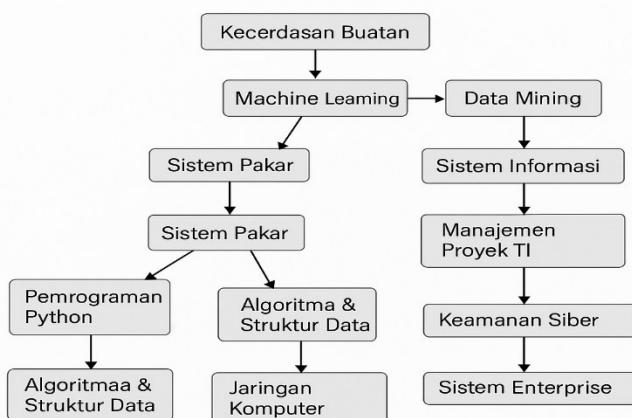
Rata-rata nilai *support* sebesar 0,28, *confidence* sebesar 0,72, dan *lift* sebesar 1,35 pada model pruned mencerminkan kekuatan asosiasi yang lebih presisi dan tajam, walaupun belum divalidasi melalui metrik klasifikasi seperti *precision* dan *recall* akibat ketiadaan data *ground truth*. Hal ini senada dengan studi Belkadi et al. (2023) yang menunjukkan bahwa teknik pruning dalam *association rule mining* mendukung eksplorasi pengetahuan yang relevan dan aplikatif di ranah data besar. Lebih lanjut, aturan-aturan asosiasi yang terbentuk merefleksikan keterkaitan tematik antar jenis buku yang sering dipinjam bersamaan oleh pengguna. Sebagai contoh, peminjam buku *Kecerdasan Buatan* juga cenderung meminjam *Machine Learning* dan *Sistem Pakar*, yang mengindikasikan keterhubungan konseptual secara teknis dan aplikatif. Pola ini konsisten dengan temuan Arabi et al. (2020) yang menyatakan bahwa pemahaman terhadap asosiasi item dapat dimanfaatkan dalam sistem rekomendasi yang bersifat kontekstual dan adaptif terhadap preferensi pengguna. Temuan ini memberikan dampak praktis yang signifikan, antara lain: a) menata ulang tata letak rak buku di perpustakaan berdasarkan hubungan konten aktual; b) mengembangkan sistem rekomendasi buku berbasis data sirkulasi aktual (Islam et al., 2023); dan c) menyusun strategi pengadaan koleksi berdasarkan analisis pola permintaan riil pengguna (Tara et al., 2024; Harisanty et al., 2023). Visualisasi hubungan hierarkis antar itemset dalam bentuk *association rule tree* pada Gambar 6 memberikan representasi intuitif mengenai arah dan kekuatan keterkaitan antar jenis buku, memperkuat akurasi interpretasi serta mendukung proses pengambilan keputusan manajerial berbasis data.

Tabel 10. Aturan Asosiasi Hasil Pruning

| No | Aturan Asosiasi  |
|----|--|
| 1  | Jika meminjam buku <i>Kecerdasan Buatan</i> , maka juga cenderung meminjam <i>Machine Learning</i> |
| 2  | Jika meminjam buku <i>Data Mining</i> , maka juga cenderung meminjam <i>Basis Data Lanjut</i>      |

| No | Aturan Asosiasi  |
|----|--|
| 3  | Jika meminjam buku <i>Sistem Informasi</i> , maka juga cenderung meminjam <i>Manajemen Proyek TI</i>             |
| 4  | Jika meminjam buku <i>Pemrograman Python</i> , maka juga cenderung meminjam <i>Algoritma &amp; Struktur Data</i> |
| 5  | Jika meminjam buku <i>Keamanan Siber</i> , maka juga cenderung meminjam <i>Jaringan Komputer</i>                 |
| 6  | Jika meminjam buku <i>Pengolahan Citra</i> , maka juga cenderung meminjam <i>Machine Learning</i>                |
| 7  | Jika meminjam buku <i>Sistem Pakar</i> , maka juga cenderung meminjam <i>Kecerdasan Buatan</i>                   |
| 8  | Jika meminjam buku <i>E-Commerce</i> , maka juga cenderung meminjam <i>Sistem Enterprise</i>                     |

Visualisasi struktur asosiasi tersebut dapat dilihat pada Gambar 6 dalam bentuk *association rule tree*, yang menggambarkan relasi hierarkis antar itemset setelah penerapan *pruning*.



Gambar 6. Association Rule Tree

#### 4. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa jenis buku seperti *Kecerdasan Buatan*, *Machine Learning*, dan *Data Mining* memiliki tingkat frekuensi peminjaman yang tinggi dan sering muncul dalam aturan asosiasi yang signifikan. Salah satu aturan paling kuat menunjukkan bahwa peminjaman buku *Kecerdasan Buatan* berkorelasi dengan peminjaman *Machine Learning*, dengan nilai *support* sebesar 27%, *confidence* sebesar 64%, dan *lift* sebesar 1,233. Temuan ini menunjukkan adanya pola preferensi pemustaka yang konsisten terhadap topik-topik yang saling melengkapi dalam ranah kecerdasan buatan dan analitik data. Penerapan teknik *pruning* terbukti efektif dalam meningkatkan kualitas aturan yang dihasilkan. Setelah item dengan frekuensi rendah disaring, terjadi peningkatan metrik analisis: *confidence* meningkat dari 62% menjadi 64%, dan *lift* dari 1,133 menjadi 1,233, sedangkan *support* tetap stabil pada 27%. Selain itu, jumlah aturan yang dihasilkan berkurang dari 26 menjadi 8. Meskipun lebih sedikit, aturan-aturan tersebut memiliki nilai asosiasi yang lebih kuat, relevan, dan mudah diinterpretasi, sehingga meningkatkan efisiensi serta utilisasi hasil analisis dalam konteks pengambilan keputusan berbasis data. Secara praktis, hasil ini dapat dimanfaatkan untuk mendukung strategi pengembangan koleksi, perancangan sistem rekomendasi adaptif berbasis perilaku pengguna, serta optimalisasi tata letak rak atau katalog digital perpustakaan berbasis keterkaitan topik. Pengelola perpustakaan dapat menggunakan pola asosiasi ini untuk merekomendasikan buku tambahan yang relevan kepada pengguna secara otomatis, memperkuat pendekatan pelayanan yang bersifat personalisasi dan berbasis data. Untuk pengembangan penelitian lanjutan, disarankan untuk mempertimbangkan

atribut tambahan seperti demografi pengguna (usia, latar belakang pendidikan, tingkat studi), atau dalam konteks akademik, variabel seperti program studi dan kurikulum mata kuliah. Pendekatan ini diharapkan dapat memperluas cakupan pemodelan asosiasi, serta meningkatkan akurasi dan relevansi sistem rekomendasi yang dihasilkan.

## REFERENSI

- Aggarwal, C.C. (2015) Outlier Analysis. In: Data Mining, Springer, Cham, 237-263.  
[https://doi.org/10.1007/978-3-319-14142-8\\_8](https://doi.org/10.1007/978-3-319-14142-8_8)
- Arabi, H., Balakrishnan, V., & Mohd Shuib, N. L. (2020). A Context-Aware Personalized Hybrid Book Recommender System. *Journal of Web Engineering*, 19(3-4), 405–428.  
<https://doi.org/10.13052/jwe1540-9589.19343>
- Ashiq, M., Jabeen, F., & Mahmood, K. (2022). Transformation of libraries during Covid-19 pandemic: A systematic review. *The Journal of Academic Librarianship*, 48(4), 102534.  
<https://doi.org/10.1016/j.acalib.2022.102534>
- Banu, A. R., Tresha, T. K., Chowdhury, S. S., & Srabonty, S. N. (2024). *Online library interfaces: A user-centered study on design and functionality preferences of Gen-Z users*. *Journal of Creative Writing*, 8(3), 40–57. <https://doi.org/10.70771/jocw.130>
- Belkadi, W. H., Drias, Y., Drias, H., Dali, M., Hamdous, S., Kamel, N., & Aksa, D. (2023). A SCORPAN-based data warehouse for digital soil mapping and association rule mining in support of sustainable agriculture and climate change analysis in the Maghreb region. *Expert Systems*, 41(7), e13464. <https://doi.org/10.1111/exsy.13464>
- Cecilio, J. D., Catedrilla, G. M. B., & A, J. R. (2023). Application of Apriori Algorithm in one state university's library book borrower records for efficient library shelving. *Journal of Software*, 18(4), 172–184. <https://doi.org/10.17706/jsw.18.4.172-184>
- Cen, C., Luo, G., Li, L., Liang, Y., Li, K., Jiang, T., & Xiong, Q. (2023). User-Centered Software Design: User Interface Redesign for Blockly-Electron, Artificial Intelligence Educational Software for Primary and Secondary Schools. *Sustainability*, 15(6), 5232.  
<https://doi.org/10.3390/su15065232>
- Cui, Z. & Yan, C. (2020). Deep Integration of Health Information Service System and Data Mining Analysis Technology. *Applied Mathematics and Nonlinear Sciences*, 5(2), 2020. 443-452. <https://doi.org/10.2478/amns.2020.2.00063>
- Dwiputra, D., Mulyo Widodo, A., Akbar, H., & Firmansyah, G. (2023). Evaluating the Performance of Association Rules in Apriori and FP-Growth Algorithms: Market Basket Analysis to Discover Rules of Item Combinations. *Journal of World Science*, 2(8), 1229–1248. <https://doi.org/10.58344/jws.v2i8.403>
- Elbadawy, M., Sato, Y., Mori, T., Goto, Y., Hayashi, K., Yamanaka, M., Azakami, D., Uchide, T., Fukushima, R., Yoshida, T., Shibutani, M., Kobayashi, M., Shinohara, Y., Abugomaa, A., Kaneda, M., Yamawaki, H., Usui, T., & Sasaki, K. (2021). Anti-tumor effect of trametinib in bladder cancer organoid and the underlying mechanism. *Cancer Biology & Therapy*, 22(5–6), 357–371. <https://doi.org/10.1080/15384047.2021.1919004>
- EShah, M. K., Gandrakota, N., Cimotti, J. P., Ghose, N., Moore, M., & Ali, M. K. (2021). Prevalence of and factors associated with nurse burnout in the US. *JAMA Network Open*, 4(2), e2036469. <https://doi.org/10.1001/jamanetworkopen.2020.36469>
- Fernandez-Basso, C., Ruiz, M.D. & Martin-Bautista, M.J. (2024). New Spark solutions for distributed frequent itemset and association rule mining algorithms. *Cluster Comput* 27, 1217–1234. <https://doi.org/10.1007/s10586-023-04014-w>
- Gaftandzhieva, S., Hussain, S., Hilcenko, S., Doneva, R., & Boykova, K. (2023). *Data-driven decision making in higher education institutions: State-of-play*. *International Journal of*

- Advanced Computer Science and Applications*, 14(6).  
<https://doi.org/10.14569/IJACSA.2023.0140642>
- Gao, M., Knobelispiesse, K., Franz, B. A., Zhai, P.-W., Cairns, B., Xu, X., & Martins, J. V. (2023). *The impact and estimation of uncertainty correlation for multi-angle polarimetric remote sensing of aerosols and ocean color*. *Atmospheric Measurement Techniques*, 16(8), 2067–2087. <https://doi.org/10.5194/amt-16-2067-2023>
- Geng, X., Wang, S., Zhang, Y., Liu, Q., & Huang, T. (2025). From algorithm to hardware: A survey on efficient and safe deployment of deep neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 36(4), 5837–5857. <https://doi.org/10.1109/TNNLS.2024.3394494>
- Gray, R. (2021). Comment on Wang et al. (2021) “Effects of family participatory dignity therapy on the psychological well-being and family function of patients with hematological malignancies and their family caregivers: A randomized controlled trial.” *International Journal of Nursing Studies*, 120, 103945. <https://doi.org/10.1016/j.ijnurstu.2021.103945>
- Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- Harisanty, D., Anna, N. E. V., Putri, T. E., Firdaus, A. A., & Noor Azizi, N. A. (2023). Is adopting artificial intelligence in libraries urgency or a buzzword? A systematic literature review. *Journal of Information Science*, 51(2), 511–522. <https://doi.org/10.1177/01655515221141034>
- Hoerudin, C. W. (2025). The Utilization of Digital Repositories for Teaching and Learning Indonesian Literature: A Review of Library Resources. *Forum for Linguistic Studies*, 7(6), 925–940. <https://doi.org/10.30564/fls.v7i6.8938>
- Hussain, W., Rasool, N., & Khan, Y. D. (2021). Insights into machine learning-based approaches for virtual screening in drug discovery: Existing strategies and streamlining through FP-CADD. *Current Drug Discovery Technologies*, 18(4), 463–472. <https://doi.org/10.2174/1570163817666200806165934>
- Islam, N., Islam, K., & Islam, M. (2023). Exploring the Potential of Big Data Analytics in Improving Library Management in Indonesia: Challenges, Opportunities, and Best Practice. *Internet Reference Services Quarterly*, 27(2), 111–120. <https://doi.org/10.1080/10875301.2023.2184900>
- Muhammad, R., Ahmad, K., Bin Naeem, S., & Jianming, Z. (2020). Budget harmonization and challenges: Understanding the competence of professionals in the budget process for structural and policy reforms in public libraries. *Performance Measurement and Metrics*, 21(2), 65–79. <https://doi.org/10.1108/PMM-09-2019-0048>
- Permatasari, P. A., Qohar, A. A., & Rachman, A. F. (2020). From web 1.0 to web 4.0: the digital heritage platforms for UNESCO's heritage properties in Indonesia. *Virtual Archeology Review*, 11 (23), 75–93. <https://doi.org/10.4995/var.2020.13121>
- Poline, V., Purohit, P. R., Bordet, P., Blanc, N., & Martinetto, P. (2024). Neural networks for rapid phase quantification of cultural heritage X-ray powder diffraction data. *Journal of Applied Crystallography*, 57(3), 831–841. <https://doi.org/10.1107/S1600576724003704>
- Pratiwi, N. B. I., Indahwati, & Fitrianto, A. (2024). Village potential mapping: Comprehensive cluster analysis of continuous and categorical variables with missing values and outliers dataset in Bogor, West Java, Indonesia. *Scientific Journal of Informatics*, 11(2), 353–366. <https://doi.org/10.15294/sji.v11i2.3903>
- Sarker, I.H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT. SCI.* 2, 160. <https://doi.org/10.1007/s42979-021-00592-x>

- Singh, B., Gangwar, S., Sharma, M., & Devi, M. (2022). An Overview of Hybrid, Digital, and Virtual Library. *World Journal of English Language*, 12(3), p32. <http://dx.doi.org/10.5430/wjel.v12n3p32>
- Singh, H., Saxena, S., Sharma, H. et al. (2025). An integrative TLBO-driven hybrid grey wolf optimizer for the efficient resolution of multi-dimensional, nonlinear engineering problems. *Sci Rep* 15, 11205. <https://doi.org/10.1038/s41598-025-89458-3>
- Sunhare, P., Chowdhary, R. R., & Chattopadhyay, M. K. (2022). Internet of things and data mining: An application-oriented survey. *Journal of King Saud University - Computer and Information Sciences*, 34(6, Part B), 3569–3590. <https://doi.org/10.1016/j.jksuci.2020.07.002>
- Tara, N., Rafi, M., & Ahmad, K. (2024). Evaluating the implementations of big data analytics in academic libraries: A structural equation model-based approach. *Performance Measurement and Metrics*, 25(3–4), 215–227. <https://doi.org/10.1108/PMM-08-2023-0022>
- Wang, L., Zhang, L., Feng, L., Chen, T., & Qin, H. (2025). A novel deep transfer learning method based on explainable feature extraction and domain reconstruction. *Neural Networks*, 187, 107401. <https://doi.org/10.1016/j.neunet.2025.107401>

*Halaman ini sengaja dikosongkan*