Jurnal Pendidikan dan Teknologi Indonesia (JPTI)
Vol. 6, No. 2, Februari 2026, Hal. 334-342
DOI: https://doi.org/10.52436/1.jpti.992
p-ISSN: 2775-4227
e-ISSN: 2775-4219

Klasterisasi Data Transaksi Penjualan Menggunakan K-Means Dan Penerapan Association Rules Menggunakan FP Growth Pada Distributor Randy Jaya Berkah

Deni Patra Mandala[1], Sahrul[*2], Indra[3]

[1,2,3]Faculty of Information Technology, Universitas Budi Luhur, Indonesia
Email: [1]denipatra21@gmail.com, [2]sahrultaip@gmail.com, [3]indra@budiluhur.ac.id

Abstrak

Distributor randy jaya berkah menghadapi permasalahan ketidakseimbangan persediaan, seperti overstock dan stockout, akibat keterbatasan analisis data penjualan yang hanya berfokus pada jumlah barang terjual dan pendapatan. Penelitian ini menerapkan algoritma K-Means untuk klasterisasi data transaksi dan FP-Growth untuk menemukan pola pembelian dominan. Data yang digunakan mencakup informasi produk, jumlah pembelian, waktu transaksi, dan data pelanggan. Hasil klasterisasi K-Means menghasilkan tiga kelompok, yaitu Cluster 1 (953 faktur), Cluster 2 (3.848 faktur), dan Cluster 3 (9.010 faktur). Sementara itu, algoritma FP-Growth menemukan salah satu pola asosiasi kuat pada produk Sosro Kotak 200 ML (Karton 24) dengan nilai Support = 0,194. Temuan ini memberikan wawasan mengenai segmentasi pelanggan, tren pembelian, dan produk yang sering dibeli bersamaan, sehingga dapat dimanfaatkan untuk strategi promosi (bundling/diskon paket), optimalisasi persediaan, dan prediksi permintaan. Secara teoritis, penelitian ini berkontribusi dalam literatur data mining khususnya penerapan K-Means dan FP-Growth, serta secara praktis mendukung pengambilan keputusan manajemen distribusi yang lebih tepat.

Kata kunci: *Data Mining, FP-Growth, K-Means, Manajemen Distribusi, Pola Pembelian, Prediksi Permintaan.*

Clustering Sales Transaction Data Using K-Means and Applying Association Rules with FP-Growth at Distributor Randy Jaya Berkah

Abstract

Distributor Randy Jaya Berkah faces inventory imbalance problems, such as overstock and stockout, due to limited sales data analysis that only focuses on the number of products sold and total revenue. This study applies the K-Means algorithm for transaction data clustering and the FP-Growth algorithm to identify dominant purchasing patterns. The dataset includes product information, purchase quantity, transaction time, and customer data. The K-Means clustering resulted in three groups: Cluster 1 (953 invoices), Cluster 2 (3,848 invoices), and Cluster 3 (9,010 invoices). Meanwhile, FP-Growth discovered a strong association pattern for Sosro Kotak 200 ML (24-carton pack) with a Support value of 0.194. These findings provide insights into customer segmentation, purchasing trends, and frequently bought-together products, which can be utilized for promotion strategies (bundling/discount packages), inventory optimization, and demand forecasting. Theoretically, this study contributes to the literature on data mining, particularly in the application of K-Means and FP-Growth, while practically supporting more accurate decision-making in distribution management.

Keywords: *Data Mining, Demand Forecasting, Distribution Management, FP-Growth, K-Means, Purchasing Pattern.*

1. PENDAHULUAN

Teknologi informasi dan komunikasi (TIK) memainkan peran penting dalam lingkungan perusahaan, tidak hanya berfungsi sebagai alat untuk otomatisasi dan pengurangan biaya, tetapi juga sebagai pendorong strategis yang berpotensi secara radikal mengubah dinamika daya saing. TIK diakui dapat memberikan peluang bagi usaha kecil untuk memperoleh akses cepat ke pasar dan pelanggan global, yang secara signifikan dapat meningkatkan jangkauan serta daya saing [1]. Distribusi adalah kegiatan penyaluran hasil produksi berupa barang dan jasa dari produsen ke konsumen untuk memenuhi kebutuhan manusia. Pihak yang melakukan kegiatan distribusi disebut sebagai distributor [2]. Distributor dikelompokan menjadi tiga jenisberdasarkan distribusinya yaitu distributor barang, distributor jasa, dan distributor peroranngan. Distributor barang

p-ISSN: 2775-4227
e-ISSN: 2775-4219

merupakan badan usaha yang menyalurkan barang proses distribusinya barang diperoleh langsung dari produsen yang disalurkan kepada agen selain itu juga barang yang dikeluarkan harus sesuai yang diharapkan konsumen agar menimbulkan kepuasan konsumen [3]. Keberhasilan dalam sistim distribusi membuat produk (merek) dapat diterima oleh pasar artinya suatu produk harus diterima sampai dengan titik terakhir sesuai segmennya [2], seperti produk yang dipasarkan, harga yang ditentukan dan promosi, untuk layanan distribusi, untuk memenuhi kebutuhan pasar target [4] dan prediksi produk yang berkelanjutan dapat diintegrasikan melalui strategi *interactive bundle pricing*, di mana konsumen terdorong untuk membeli lebih dari satu produk karena digabungkan dalam satu paket promo atau bundling. Strategi ini terbukti mampu meningkatkan penjualan dan keuntungan penjual dengan tetap memperhatikan preferensi keberlanjutan konsumen [5].

Distributor Randy Jaya Berkah merupakan salah satu distributor minuman yang beroperasi di beberapa wilayah Kabupaten Bogor menghasilkan data transaksi penjualan dalam jumlah besar setiap harinya. Namun terdapat masalah yang dihadapi Ketidaktepatan dalam penyusunan strategi pemasaran seperti tanpa adanya analisis klaster penjualan dan aturan asosiasi yang terukur, manajemen kesulitan menentukan prioritas produk apa yang sebaiknya dipromosikan secara bersamaan (bundling) atau didorong penjualannya pada segmen tertentu dan minimnya pemanfaatan data transaksi untuk rekomendasi penjualan seperti meski data penjualan melimpah, belum ada upaya optimal untuk menggali asosiasi (hubungan/aturan) antarkategori atau produk. Kondisi ini menghambat pembuatan strategi cross-selling, up-selling, atau bundling produk yang efektif.

Sehubungan dengan hal tersebut, dibutuhkan suatu metode data mining atau penambangan data yang menjadi salah satu teknologi yang berkembang pesat dan memungkinkan pengguna untuk menggali informasi berharga dari kumpulan data yang kompleks. Data mining tidak hanya bermanfaat dalam mengolah data secara efisien tetapi juga mampu mendukung pengambilan keputusan berbasis data, yang kini menjadi dasar dalam berbagai bidang, seperti bisnis, kesehatan, keuangan, dan pendidikan [6]. Salah satunya teknik algoritma *K-Means* clustering yang dirancang untuk memisahkan data ke dalam kelompok atau cluster berdasarkan tingkat kesamaan seperti mengelompokkan data transaksi, risiko kerugian akibat kelebihan stok atau kekurangan pasokan dapat diminimalkan [7] dan tehnik algoritma Frequent Pattern Growth (FP-Growth) lebih efisien diterapkan pada dataset transaksi berskala besar, mampu menghindari proses generasi kandidat yang memakan waktu dan sumber daya, sehingga strategi penjualan yang memerlukan analisis pola asosiasi produk secara cepat dan efektif [8].

Penelitian sebelumnya yang mengimplementasi algoritma FP-Growth di koperasi KAOCHEM Sinergi Mandiri yang bergerak di bidang kebutuhan pokok (sembako), menggunakan algoritma FP-Growth untuk mengetahui tingkat pembelian barang bersamaan oleh pelanggan serta dapat menentukan strategi cross-selling dan bundle/promo produk yang banyak dibeli bersama. Hasil perhitungan algoritma FP-Growth menghasilkan 3 strong rules dengan nilai support sekitar 3% dan confidence minimal 50% [9]. Toko HAS menggunakan algoritma FP-Growth untuk menganalisis data penjualan pakaian dan mengidentifikasi produk yang paling sering dibeli oleh pelanggan. Hasil analisis menunjukkan asosiasi kuat antara produk seperti Gamis dan Jilbab, dengan nilai support 53,33% dan confidence 100%. Temuan ini memungkinkan toko untuk menyusun strategi pemasaran yang lebih fokus dan efisien, serta mengoptimalkan tata letak produk untuk meningkatkan pelayanan dan kenyamanan pelanggan dalam berbelanja [10]. PT Global Lighting Indonesia bergerak dibidang penjualan lampu, menggunakan algoritma K-Means untuk membantu persiapan stok yang tidak stabil karena kekurangan stok lampu saat permintaan konsumen. Hasilnya mendapatkan 3 Cluster yaitu cluster 0 dengan 39 data kategori laku, Cluster 2 dengan 7 data kategori kurang laku dan Cluster 2 dengan 7 data dengan menggunakan 192 kategori paling laku [11]. Toko Asrafi Raya merupakan suatu toko yang bergerak di bidang penjualan tuppeware yang berada di daerah Kabupaten Pasaman Barat, menggunakan metode K-Means untuk meningkatkan Penjualan Tupperware karena kesulitan kurangnya stok produk yang laku karena penjualan tinggi dan menumpuknya produk yang tidak laku karena penjualannya rendah, hasilnya membantu pemilik toko Asrafi Raya dalam menentukan strategi penjualan mendapatkan 3 cluster yaitu cluster 1(C1)Sangat Laris,Cluster 2 (C2) Laris, Cluster 3 (C3) Tidak Laris [12]. Sports Station perusahaan retail yang menjual perlengkapan olahraga menggunakan penerapan metode K-Means Clustering karena sering mengalami kesulitan mengelompokkan produk. Hasilnya menunjukkan pengelompokkan menjadi 2 yaitu cluster 0 dengan nilai 995 sebanyak 121 produk dengan kategori laris dan cluster 1 dengan nilai 327 sebanyak 2.279 produk dengan kategori kurang laris. Serta hasil DBI yang paling mendekati 0 adalah K 2 menghasilkan nilai 0,10 [13]. Fashion Viral Solo bergerak dibidang distributor perlengkapan fashion, menggunakan algoritma K-Means untuk membantu segmentasi dan mengelompokkan pelanggan berdasarkan kemiripannya. Hasilnya mendapatkan 2 cluster yaitu cluster 1 dengan jumlah 343 pelanggan kategori jumlah anggota terbanyak dan cluster 2 dengan jumlah 8 pelanggan kategori jumlah anggota sedikit [14].

Secara keseluruhan, hasil penelitian sebelumnya meningkatkan efisiensi strategi penjualan, merancang rekomendasi promosi produk yang lebih efektif dan mengoptimalkan peluang keberhasilan dengan menargetkan produk yang paling diminati oleh konsumen [15]. Melalui proses data mining dengan *K-Means dan FP-Growth*

p-ISSN: 2775-4227
e-ISSN: 2775-4219

pada Distributor Randy Jaya Berkah dapat membantu mengidentifikasi item-item yang sering dibeli bersama dalam transaksi pembelian, sehingga membantu menetapkan aturan asosiasi antara item-item tersebut [Fajrawati]. serta dapat menyiapkan strategi untuk meningkatkan strategi pemasaran, kepuasan pelanggan, serta meningkatkan optimalisasi persediaan produk Distributor Randy Jaya Berkah [16].

2. METODE PENELITIAN

Metode penelitian ini didukung dengan tahapan data mining dan kerangka kerja CRISP-DM. Pada tahapan data mining untuk mencari informasi-informasi yang berharga, pola yang ada di dalam data, yang melibatkan algoritma untuk mengidentifikasi pola pada data [17], data mining tersebut menggunakan pemodelan algoritma K-Means untuk melakukan clustering dan algoritma Frequent Pattern Growth (FP-Growth) untuk menemukan pola, relasi, atau tren yang tidak terlihat secara langsung dan dapat memberikan wawasan berharga untuk pengambilan keputusan [18] serta menggunakan kerangka kerja CRISP-DM memberikan pendekatan sistematis dari pemahaman bisnis hingga implementasi, memastikan bahwa setiap fase, dari pemahaman konteks bisnis hingga penerapan solusi [19]. Adapun penjelasan terkait metode penelitian ini sebagai berikut:

2.1. Metode Algoritma K-Means dan Algoritma Frequent Pattern Growth (FP-Growth)

Pemilihan algoritma K-Means dalam penelitian ini didasarkan pada beberapa pertimbangan metodologis. K-Means merupakan algoritma klasterisasi yang relatif sederhana, efisien, dan mampu menangani dataset berukuran besar dengan waktu komputasi yang singkat. Hal ini sesuai dengan karakteristik data transaksi Distributor Randy Jaya Berkah yang terdiri dari puluhan ribu baris data. Dibandingkan dengan metode lain seperti DBSCAN, K-Means lebih mudah diterapkan karena tidak memerlukan parameter kepadatan (epsilon dan minimum points) yang sensitif terhadap variasi data. DBSCAN dapat mendeteksi klaster dengan bentuk tidak beraturan, namun cenderung kurang optimal jika data berukuran besar dan memiliki distribusi yang relatif homogen seperti transaksi penjualan. Oleh karena itu, K-Means dinilai lebih sesuai untuk menghasilkan segmentasi pelanggan atau transaksi berdasarkan kesamaan pola pembelian dengan hasil yang mudah diinterpretasikan.

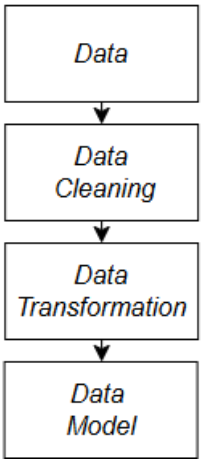
Sementara itu, algoritma FP-Growth dipilih karena kemampuannya yang efisien dalam menemukan frequent itemset tanpa harus menghasilkan kandidat aturan secara eksplisit, sebagaimana dilakukan oleh algoritma Apriori. Apriori bekerja dengan pendekatan generate-and-test yang sering menimbulkan ledakan kombinasi kandidat, sehingga tidak efisien pada dataset besar. Sebaliknya, FP-Growth menggunakan struktur FP-Tree yang lebih ringkas dan hanya memerlukan dua kali pemindaian data, sehingga mampu memproses data dalam skala besar dengan lebih cepat. Dengan karakteristik data transaksi Distributor Randy Jaya Berkah yang kompleks dan berjumlah besar, penggunaan FP-Growth terbukti lebih tepat karena mampu menemukan pola asosiasi produk yang signifikan dengan tingkat efisiensi tinggi, sekaligus mengurangi beban komputasi yang tidak perlu.

Dalam penelitian ini, penentuan jumlah klaster pada algoritma K-Means dilakukan menggunakan Elbow Method yaitu dengan mengamati titik siku pada grafik hubungan antara jumlah klaster (k) dengan nilai Sum of Squared Errors (SSE). Pada algoritma FP-Growth, parameter utama yang digunakan adalah support threshold, yaitu ukuran frekuensi kemunculan suatu item dalam keseluruhan transaksi. Nilai parameter tersebut menentukan itemset yang dipertahankan dalam pembentukan aturan asosiasi, sehingga hanya item dengan tingkat kemunculan signifikan yang diperhitungkan.

2.2. Metode CRISP-DM

Penelitian ini dilakukan melalui tahapan yang terstruktur, mulai dari *data*, *data cleaning*, *data transformation* dan *model*. Tahapan ini menggunakan data mining suatu proses penambangan atau penemuan informasi baru yang dilakukan dengan cara mencari sebuah pola atau aturan tertentu dari sejumlah data yang menumpuk dan dikatakan data besar [20]. Adapun alur tahapan terstruktur pada penelitian ini dapat dilihat pada Gambar 1.

p-ISSN: 2775-4227
e-ISSN: 2775-4219



Gambar 1. Tahapan Penelitian

2.2.1 Data

Memahami data secara mendalam pada tahap awal, memastikan bahwa data yang digunakan untuk proses selanjutnya telah terverifikasi dan sesuai dengan tujuan bisnis, sehingga hasil penambangan data menjadi lebih akurat dan dapat diandalkan [21]. Tahap persiapan data melibatkan penyesuaian data agar dapat diproses sesuai dengan kebutuhan algoritma penambangan data, sehingga memungkinkan penambangan data dilakukan. Tahap persiapan data tidak hanya harus memastikan kelayakan teknis dari proses penambangan data, tetapi juga melibatkan keputusan terkait pengumpulan data. Keputusan-keputusan ini memastikan bahwa hanya data yang relevan yang dipilih, yang kemudian disimpan dalam berkas atau data mart terpisah untuk memastikan stabilitas dan efisiensi selama proses penambangan data yang bersifat iteratif [22]. Data yang digunakan berupa transaksi penjualan produk minuman selama periode April 2023–April 2024, dengan total 29.020 baris data. Data mencakup atribut faktur, kode produk, nama produk, jumlah pembelian (Qty), harga, subtotal, diskon, dan total.

2.2.2 Data Cleaning

Pada tahap ini, data cleansing atau pembersihan data merupakan proses menganalisis data dengan cara mengubah atau menghapus data yang diakibatkan oleh kesalahan input (human error) dan kesalahan pada sistem [23]. Tujuan dari data cleansing adalah untuk meningkatkan kualitas data dengan memperbaiki kesalahan, menormalisasi format, dan menghilangkan duplikat sehingga data menjadi lebih bermanfaat dan dapat diandalkan untuk analisis atau penggunaan lainnya [24]. Cakupan proses *data cleaning* dilakukan sebagai tahap awal dalam pengolahan data untuk memastikan kualitas dan konsistensi dataset. Tahap ini mencakup kegiatan seperti menghapus data duplikat, memeriksa adanya ketidakkonsistenan, serta memperbaiki kesalahan pencatatan pada data sesuai dengan tahapan dalam proses KDD. Pada penelitian ini, proses *data selection* tidak menghapus data produk yang memiliki kesamaan antara produk dan faktur, sehingga keseluruhan data tetap dipertahankan. Setelah tahap *cleaning* selesai data transaksi yang siap digunakan dalam proses klasterisasi menggunakan algoritma K-Means.

2.2.3 Data Transformation

Transformation merupakan proses mengubah data yang dipilih untuk membuat data tersebut sesuai dalam data mining. Prosesnya sangat bergantung pada database, model informasi atau jenis yang akan dicari. Definisi lain dari transformasi adalah pencarian fitur yang berguna untuk merepresentasikan data tergantung pada tujuan yang ingin dicapai [25]. Tujuan utama dari data transformation adalah mengubah format dan representasi data agar lebih sesuai dan relevan untuk analisis lanjutan dengan beberapa teknik yang digunakan pada tahapan ini antara lain transformasi atribut, transformasi skala, pembentukan fitur baru, dan diskritisasi data [26]. Cakupan teknik data transformation dengan normalization, pemilihan attribute, dan discretization. Pada normalisasi memastikan bahwa semua klaster berada dalam skala yang sama untuk memastikan bahwa semua klaster berada dalam skala yang sama dalam rentang 0-1 dan Diskritisasi menganalisis data klaster dalam klasifikasi atau pengelompokan yang memerlukan data dalam bentuk kategori dengan nilai numerik telah dikonversi ke dalam kategori 0, 1, dan 2.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

2.2.5 Data Model

Pada tahap model, melibatkan pemilihan teknik data mining yang akan digunakan, serta penentuan tools, algoritma, dan parameter yang optimal. Dalam penelitian ini, algoritma yang diterapkan adalah K-Means Clustering, sebuah algoritma Unsupervised Learning yang sering digunakan untuk mengelompokkan data berdasarkan kesamaan fitur [27]. Penggunaan metode K-Means dipilih karena dapat membantu toko dalam menganalisis dan mengklasifikasikan data untuk menentukan persediaan produk secara cepat dan efisien dari sejumlah besar data transaksi penjualan. Oleh karena itu, penelitian ini menerapkan analisis data mining dengan teknik clustering menggunakan algoritma K-Means [7]. Association Rule Mining dengan menggunakan algoritma FP-Growth adalah salah satu teknik data mining yang relevan dalam konteks ini. Algoritma ini dirancang untuk menemukan hubungan antar produk yang sering dibeli secara bersamaan. Dengan menerapkan algoritma FP-Growth dalam Association Rule Mining, pemilik usaha dapat mengeksplorasi hubungan antara produk, meningkatkan efektivitas promosi, dan akhirnya mengoptimalkan penjualan [28]. Cakupan data model proses analisis dilakukan menggunakan Google Colabs dengan bantuan beberapa library, antara lain matplotlib.pyplot untuk visualisasi data dan KMeans dari sklearn.cluster untuk proses klasterisasi. Langkah-langkah yang dilakukan meliputi inisialisasi variabel, perhitungan nilai inertia untuk berbagai jumlah klaster, serta pembuatan grafik Elbow Method guna menentukan jumlah klaster optimal. Selanjutnya, algoritma K-Means dijalankan melalui beberapa iterasi, di mana posisi centroid diperbarui hingga distribusi data stabil. Hasil akhir dari proses ini membagi data transaksi ke dalam tiga klaster dengan karakteristik berbeda, yaitu klaster pelanggan dengan jumlah pembelian besar, klaster pelanggan dengan jumlah pembelian menengah, dan klaster pelanggan dengan jumlah pembelian kecil atau satuan. Distribusi ini kemudian divisualisasikan dalam bentuk grafik sehingga memudahkan interpretasi pola data berdasarkan hasil klasterisasi.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Klasterisasi Kmeans

Penelitian ini memproses variasi dalam suatu kelompok dengan memaksimalkan variasi antar kelompok melalui tehnik pengelompokan. Adapun proses tehnik pengelompokan melalui tahapan berikut:

Data transaksi penjualan terdapat 29.020 row untuk pengolahan data klastering K-Means pada distributor ABC. Sebelum proses klastering, dilakukan pemeriksaan dan penghapusan data duplikat. Namun, tidak ditemukan data ganda. Pengelompokan awal dilakukan terhadap 13.811 faktur, yang masing-masing mencakup produk jus, teh, dan air mineral. Kolom jumlah menunjukkan total produk yang dibeli per faktur, sebagaimana ditampilkan pada tabel 1 data transaksi penjualan.

Tabel 1. Data Transaksi Penjualan

No	Faktur	Produk	Jumlah
1	10000934	Prima Pet 600 ML (Karton 24 SHORT NECK), Stee Pet 390 ML (Karton 12 – PCS), Sosro 220 ML	3
2	10001392	Stee Pet 390 ML (Karton 12 – PCS)	1
3	10001484	Sosro Sparkling LL 300 ML (Karton 12)	1
4	10001621	Fruit Tea Apel 350 ML (Karton 12 – SHRINK)	1
…	…	…	…
13808	9998063	Stee Pet 390 ML (Karton 12 – PCS), Sosro 350 ML (Karton 12 Aseptic Less Sugar), Fruit Tea Apel 350 ML (Karton 12 – SHRINK)	3
13809	9998141	Fruit Tea Apel 200 ML (Karton 24), Sosro Kotak 330 ML (Karton 24)	2
13810	9998689	Prima Air Mineral 330 ML (Short Neck)	1
13811	9999059	Sosro 220 ML, Sosro 350 ML (Karton 12 Aseptic Less Sugar)	2

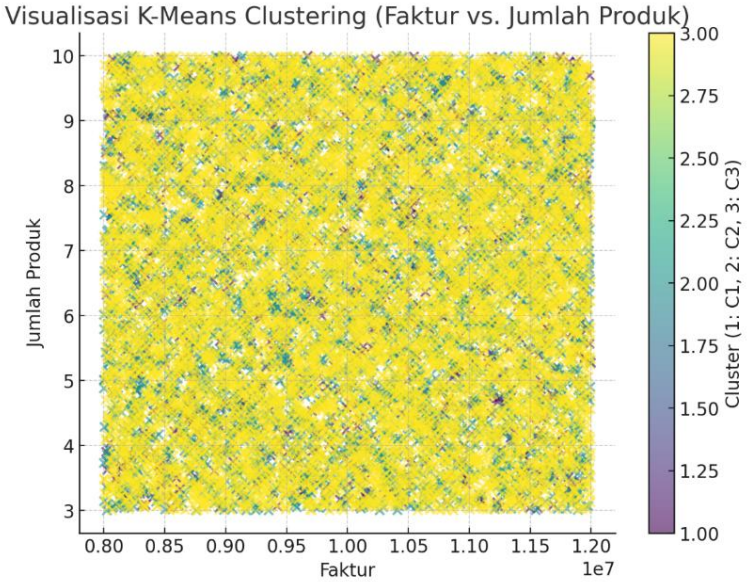
Merupakan operasi dasar yang dilakukan seperti penghapusan noise. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data Tahapan Proses KDD. Untuk penelitian ini data selection tidak ada penghapusan data Produk yang berisi data yang sama antara Produk dan Faktur yang artinya pada proses klastering K-Means dengan total data yang siap diolah sampel produk yang dibeli sebanyak 13.811 data.

Transformasi data dilakukan melalui normalisasi dan diskritisasi untuk mempersiapkan data dalam proses data mining. Normalisasi dilakukan pada atribut seperti 'Struk', 'Teh', 'Jus', dan 'Air Mineral' dalam rentang 0–1 agar berada dalam skala yang sama. Selanjutnya, data didiskritisasi ke dalam tiga kategori (0, 1, 2) untuk mendukung analisis klaster dan klasifikasi.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Penelitian ini menerapkan metode K-Means untuk mengelompokkan data transaksi berdasarkan kemiripan karakteristik. Visualisasi Elbow Method menunjukkan bahwa dua hingga tiga klaster merupakan jumlah optimal berdasarkan nilai SSE. Namun, eksperimen dilanjutkan dengan tiga klaster (C1, C2, dan C3) untuk eksplorasi lebih lanjut. Melalui tiga iterasi, dilakukan pembaruan jarak ke centroid dan klasifikasi ulang. Awalnya, klaster C2 mendominasi, namun setelah iterasi kedua dan ketiga, klaster C3 tumbuh signifikan hingga menjadi klaster terbesar, menunjukkan penyesuaian karakteristik data terhadap centroid baru. Sementara itu, klaster C1 tetap stabil dengan jumlah anggota yang lebih kecil. Distribusi data menunjukkan bahwa: C1 cenderung berisi faktur dengan jumlah produk lebih banyak, kemungkinan dari pelanggan besar atau distributor. C2 berisi transaksi dengan jumlah produk sedang, mewakili toko ritel kecil atau menengah. C3 terdiri dari faktur dengan satu produk, yang mengindikasikan pembelian individual untuk konsumsi pribadi. Visualisasi akhir menunjukkan dominasi klaster C3 (9.010 faktur), diikuti oleh C2 (3.848 faktur) dan C1 (953 faktur). Proses ini menghasilkan pemetaan yang lebih akurat terhadap karakteristik pelanggan berdasarkan pola pembelian.

Pada gambar 2 dibawah menunjukkan klaster yang berbeda, dengan legenda warna di sebelah kanan yang menjelaskan kategori klaster (1: C1, 2: C2, 3: C3). Dari visualisasi ini, terlihat bahwa klaster C3 (kuning) mendominasi jumlah data, yang sesuai dengan jumlah anggotanya yang terbesar (9.010 faktur). Sementara itu, C2 (hijau) memiliki jumlah anggota sedang, yaitu 3.848 faktur, dan C1 (ungu) adalah klaster terkecil dengan 953 faktur. Distribusi ini memperlihatkan bagaimana data dikelompokkan berdasarkan pola tertentu, di mana klaster C3 menjadi pusat utama dari data yang dianalisis, dapat dilihat pada gambar 2.



Gambar 2. Hasil Clustering dengan K-Means

Pada tabel 2 hasil K-Means dibawah menunjukkan hasil akhir dari proses clustering berdasarkan data faktur, jenis produk yang dipesan, jumlah produk, dan klaster hasil pengelompokan (C1, C2, dan C3). Dari tabel ini, terlihat bahwa setiap faktur memiliki daftar produk tertentu dengan jumlah yang bervariasi. Faktur yang masuk ke dalam C1 cenderung memiliki jumlah produk yang lebih besar, misalnya faktur 10002194 dengan 6 produk, dan faktur 10012646 juga dengan 6 produk. Ini menunjukkan bahwa C1 kemungkinan besar terdiri dari transaksi dengan jumlah produk yang lebih banyak, yang bisa mengindikasikan pelanggan dengan pembelian dalam skala lebih besar, seperti distributor atau toko besar. Sementara itu, klaster C2 terdiri dari faktur dengan jumlah produk yang sedikit lebih rendah dibandingkan C1, tetapi masih dalam jumlah yang cukup banyak. Misalnya, faktur 10000934 memiliki 3 produk, dan faktur 10003130 memiliki 4 produk. Ini menunjukkan bahwa C2 kemungkinan besar berisi pelanggan menengah atau toko ritel kecil yang membeli dalam jumlah sedang. Produk yang dipesan dalam klaster ini bervariasi, tetapi terlihat ada pola tertentu, seperti pemesanan minuman dalam kemasan karton atau shrink pack yang umum ditemukan di toko atau minimarket. Berbeda dengan C1 dan C2, klaster C3 tampaknya berisi faktur dengan jumlah produk yang sangat sedikit, bahkan hanya 1 produk per transaksi, seperti faktur 9997732, 9997733, dan 9998689. Ini menunjukkan bahwa klaster C3 mungkin terdiri dari pelanggan pengklasteran pembelian satuan, misalnya konsumen yang membeli dalam jumlah kecil untuk konsumsi pribadi, dapat dilihat pada tabel 2.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Tabel 2. Hasil K-Means

Klaster	Jumlah Faktur	Karakteristik Utama
C1	953 (6,9%)	Pembelian besar, pelanggan grosir/distributor
C2	3.848 (27,8%)	Pembelian menengah, toko ritel kecil/menengah
C3	9.010 (65,3%)	Pembelian kecil, pelanggan individu atau eceran

Fakta bahwa Cluster 3 (C3) mencakup lebih dari setengah total transaksi menunjukkan bahwa sebagian besar pelanggan Distributor Randy Jaya Berkah merupakan pembeli eceran dengan jumlah produk sedikit. Pola ini menunjukkan besarnya potensi pasar di segmen konsumen akhir yang melakukan pembelian rutin dalam volume kecil.

Sebaliknya, Cluster 1 (C1) merepresentasikan pelanggan besar (distributor sekunder atau toko besar) yang membeli dalam jumlah banyak (>5 produk per transaksi). Meskipun jumlahnya sedikit, segmen ini memberikan kontribusi besar terhadap total volume penjualan. Cluster 2 (C2) berfungsi sebagai kelompok transisi yang mewakili toko kecil atau pelanggan bisnis menengah dengan frekuensi pembelian sedang.

Hasil ini sejalan dengan penelitian oleh Rahmawati et al. [12], yang menunjukkan bahwa dalam konteks penjualan ritel, segmen dengan jumlah transaksi kecil cenderung mendominasi, sedangkan pelanggan besar memberikan kontribusi signifikan terhadap total nilai penjualan. Dengan demikian, manajemen dapat menyusun strategi promosi dan distribusi yang berbeda untuk setiap klaster: C1 difokuskan pada diskon volume, C2 pada insentif pembelian rutin, dan C3 pada program loyalitas pelanggan.

3.2. Hasil Association Rule FP-Growth

Penelitian ini mengidentifikasi pola tersembunyi antar kluster melalui analisis nilai dukungan dari kelompok item yang mencakup berbagai kategori:

FP-Growth memastikan bahwa data yang digunakan sudah sesuai dari hasil klastering K-Means yang dihasilkan. Itemset dianalisis berdasarkan nilai support dari tingkat kemunculan suatu produk dalam transaksi dibandingkan dengan total jumlah transaksi menunjukkan Sosro Kotak 200 ML memiliki nilai support tertinggi (0.194), diikuti oleh Stee Pet 390 ML (0.190) dan Sosro 350 ML (0.180), menandakan frekuensi kemunculan yang tinggi dalam transaksi. Sementara itu, Prima Pet 600 ML memiliki support terendah (0.176), menunjukkan frekuensi yang lebih rendah dibandingkan produk lainnya.

Pola transaksi yang sering terjadi berdasarkan algoritma FP-Growth, ROOT mewakili titik awal dari seluruh transaksi dan setiap cabang menunjukkan jalur yang mewakili kombinasi produk yang sering dibeli seperti Sosro Kotak 200 ML, Stee Pet 390 ML, dan Prima Pet 600 ML sering muncul dalam banyak transaksi. Frekuensi tinggi pada beberapa node menandakan bahwa produk-produk ini populer dan memiliki asosiasi kuat dengan barang lainnya.

Pola frequent itemset dalam transaksi, FP-Tree ini dimulai dari ROOT yang tersusun secara hierarkis: Prima Pet 600 ML (Karton 24 SHORT NECK), Stee Pet 390 ML (Karton 12 - PCS), dan Sosro 220 ML.

Pola pembelian dalam dataset transaksi, FP-Tree ini memiliki ROOT sebagai node awal, yang kemudian diikuti oleh satu node yang merepresentasikan Stee Pet 390 ML (Karton 12 - P).

Pola pembelian dalam transaksi tertentu, FP-Tree ini memiliki ROOT sebagai node awal, yang kemudian diikuti oleh satu node yang merepresentasikan produk Sosro Sparkling LL 300 ML (Karton).

Hasil pembentukan FP-Tree menunjukkan bahwa beberapa produk sering muncul bersamaan dalam satu transaksi, seperti:

{Sosro Kotak 200 ML} → {Stee Pet 390 ML}
{Prima Pet 600 ML} → {Stee Pet 390 ML, Sosro Kotak 200 ML}

Hal ini menandakan adanya pola pembelian berulang (co-occurrence pattern) antarproduk yang dapat dimanfaatkan dalam strategi cross-selling dan bundling. Sebagai contoh, pelanggan yang membeli Sosro Kotak 200 ML berpotensi besar untuk membeli Stee Pet 390 ML pada transaksi yang sama.

Integrasi hasil dari K-Means dan FP-Growth memberikan wawasan strategis bagi manajemen Distributor Randy Jaya Berkah:

1. Optimasi Persediaan: Produk dengan nilai support tinggi seperti Sosro Kotak 200 ML dan Stee Pet 390 ML perlu mendapatkan prioritas dalam stok agar mengurangi risiko stockout.
2. Strategi Promosi: Hubungan kuat antarproduk dapat dimanfaatkan untuk promosi bundling (paket minuman campuran) guna meningkatkan nilai transaksi rata-rata.
3. Segmentasi Pelanggan: Klasterisasi dapat digunakan untuk menyesuaikan strategi pemasaran per segmen, misalnya promosi grosir untuk C1 dan diskon eceran untuk C3.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

4. KESIMPULAN

Penelitian ini membuktikan bahwa kombinasi algoritma K-Means dan FP-Growth efektif dalam menganalisis pola transaksi penjualan pada Distributor Randy Jaya Berkah. Pendekatan ini memberikan pemahaman menyeluruh terhadap segmentasi pelanggan dan pola keterkaitan antarproduk yang sering dibeli bersamaan.

Hasil klasterisasi menunjukkan adanya perbedaan perilaku pembelian antara pelanggan besar, ritel menengah, dan individu, sedangkan analisis asosiasi mengidentifikasi produk inti dengan frekuensi tinggi yang dapat dijadikan acuan dalam strategi promosi dan pengelolaan persediaan. Hal ini memperlihatkan bahwa penerapan data mining mampu mengubah data transaksi menjadi wawasan strategis untuk mendukung keputusan bisnis berbasis data.

Secara praktis, penelitian ini memberikan kontribusi bagi manajemen distribusi dalam merancang kebijakan stok, strategi cross-selling, dan perencanaan pemasaran yang lebih efisien.

Secara akademik, penelitian ini memperkuat literatur penerapan data mining di sektor distribusi dengan menggabungkan analisis klaster dan asosiasi untuk menemukan pola pembelian yang bermakna.

Untuk penelitian selanjutnya, disarankan:

1. Mengintegrasikan data real-time agar pola pembelian dapat dipantau dan dianalisis secara dinamis.
2. Melakukan perbandingan dengan algoritma lain seperti DBSCAN untuk klasterisasi non-linear dan Apriori untuk asosiasi berbasis confidence rule, guna menilai efektivitas relatif dari setiap pendekatan.

Memperluas analisis dengan memasukkan variabel perilaku pelanggan seperti waktu pembelian atau frekuensi transaksi, sehingga hasilnya lebih komprehensif.

DAFTAR PUSTAKA

[1] T. Yuwono, A. Suroso, and W. Novandari, "Information and communication technology in SMEs: a systematic literature review," *J. Innov. Entrep.*, vol. 13, no. 1, 2024, doi: 10.1186/s13731-024-00392-6.

[2] A. Rosid, Suparji, and F. Fuad, "Perjanjian Distributor dengan Pedagang Kelontong (Studi Kasus Perjanjian Distributor Pemasok Snack ke Pedagang Kelontong di Pasar Cipete Utara Jakarta Selatan)," *UNES Law Rev.*, vol. 6, no. 4, pp. 10932–10943, 2024, [Online]. Available: doi: 10.31933/unesrev.v6i4

[3] V. F. Dharmawan and M. Megawati, "Analisis Pengaruh Harga, Dan Lokasi Terhadap Kepuasan Konsumen Pada Toko Bangunan Pelita Mas Di Kota Palembang," *Publ. Ris. Mhs. Manaj.*, vol. 5, no. 2, pp. 203–208, 2024, doi: 10.35957/prmm.v5i2.7496.

[4] N. A. Kamilah, T. Rohana, R. Rahmat, and A. Fauzi, "Implementasi Algoritma K-Means dan K-Medoids Dalam Klasterisasi Kasus Kekerasan Terhadap Perempuan," *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 810, 2024, doi: 10.30865/mib.v8i2.7558.

[5] X. Liu, X. Liu, A. Shi, and C. Li, "Agricultural Products' Bundled Pricing Based on Consumers' Organic Preferences," *Sustain.*, vol. 15, no. 17, 2023, doi: 10.3390/su151713256.

[6] E. Saputri, "Teknik dan aplikasi data mining di Indonesia: tinjauan literatur satu dekade (2015-2024)," *IT-EXPLORE J. Penerapan Teknol. Inf. dan Komun.*, vol. 4, no. 2, pp. 138–149, 2025, doi: 10.24246/itexplore.v4i2.2025.pp138-149.

[7] R. Rahmawati, W. Prihartono, and K. Cirebon, "OPTIMASI STOK DENGAN CLUSTERING DATA MENGGUNAKAN," vol. 13, no. 2, 2025, [Online]. Available: http://dx.doi.org/10.23960/jitet.v13i2.6302

[8] I. D. Hunyadi, N. Constantinescu, and O. A. Țicleanu, "Efficient Discovery of Association Rules in E-Commerce: Comparing Candidate Generation and Pattern Growth Techniques," *Appl. Sci.*, vol. 15, no. 10, 2025, doi: 10.3390/app15105498.

[9] Fildzah Zia Ghassani, Asep Jamaludin, and Agung Susilo Yuda Irawan, "Market Basket Analysis Menggunakan Algoritma FP-Growth dalam Menentukan Cross-Selling," *J. Inform. Polinema*, no. 12, pp. 49–54, 2021, [Online]. Available: https://doi.org/10.33795/jip.v7i4.508

[10] R. Fauzi, A. W. Aranski, N. Nopriadi, and E. Hutabri, "Implementasi Data Mining Pada Penjualan Pakaian dengan Algoritma FP-Growth," *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 2, p. 436, 2023, doi: 10.30865/jurikom.v10i2.5795.

[11] H. W. Prayogo Putra Tjaya, Rino, "Implementasi Metode Clustering K-Means Untuk," vol. 0577, no. I, 2021, [Online]. Available: doi: 10.31253/algor.v3i1

[12] I. P. Mulyadi, "Klasterisasi Menggunakan Metode Algoritma K-Means dalam Meningkatkan Penjualan Tupperware," *J. Inform. Ekon. Bisnis*, vol. 4, pp. 172–179, 2022, doi: 10.37034/infeb.v4i4.164.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

[13] N. P. Gantara and I. Ali, "Penerapan Metode K-Means Clustering Pada Penjualan Barang Di Sports Station," *E-Link J. Tek. Elektro dan Inform.*, vol. 18, no. 1, p. 28, 2023, doi: 10.30587/e-link.v18i1.5339.

[14] R. B. Ardi, F. Ely Nastiti, and S. Sumarlinda, "Algoritma K-Means Clustering Untuk Segmentasi Pelanggan (Studi Kasus : Fashion Viral Solo)," *INFOTECH J.*, vol. 9, no. 1, pp. 124–131, 2023, doi: 10.31949/infotech.v9i1.5214.

[15] K. Kharomiyah, N. Rahaningsih, and R. D. Dana, "Analisis Keterkaitan Penjualan Obat melalui Penerapan Algoritma FP-Growth guna Optimalisasi Strategi Pemasaran," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 23, no. 1, p. 57, 2024, doi: 10.53513/jis.v23i1.9495.

[16] F. Febrian and Z. Fatah, "Optimasi Penentuan Paket Hemat Menggunakan Algoritma FP-Growth untuk Meningkatkan Strategi Pemasaran," *JUSIFOR J. Sist. Inf. dan Inform.*, vol. 3, no. 2, pp. 172–179, 2024, doi: 10.70609/jusifor.v3i2.5818.

[17] M. Hosseinzadeh *et al.*, "Data cleansing mechanisms and approaches for big data analytics: a systematic study," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 1, pp. 99–111, 2023, doi: 10.1007/s12652-021-03590-2.

[18] I. Juwita and I. Ali, "Penerapan Pola Penjualan Dengan Menggunakan Metode Algoritma Asosiasi Fp-Growth Bertujuan Untuk Meningkatkan Penjualan Kopi Di Point Coffee," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1600–1607, 2024, doi: 10.36040/jati.v8i2.9025.

[19] R. H. Bemthuis, R. R. Govers, and A. Asadi, "A CRISP-DM-based methodology for assessing agent-based simulation models using process mining," *J. Simul.*, 2025, doi: 10.1080/17477778.2025.2508245.

[20] R. R. Nasution, D. Saripurna, and T. Haramaini, "Penerapan Data Mining dalam Menentukan Pola Peminjaman Perlengkapan Outdoor RR Adventure dengan Menggunakan Algoritma Apriori," *Blend Sains J. Tek.*, vol. 3, no. 1, pp. 63–73, 2024, doi: 10.56211/blendsains.v3i1.546.

[21] H. Asyraf and M. E. Prasetya, "Implementasi Metode CRISP DM dan Algoritma Decision Tree Untuk Strategi Produksi Kerajinan Tangan pada UMKM A," *J. Media Inform. Budidarma*, vol. 8, no. 1, p. 94, 2024, doi: 10.30865/mib.v8i1.7050.

[22] F. Hochkamp, A. A. Scheidler, and M. Rabe, "Review of Maturity Models for Data Mining and Proposal of a Data Preparation Maturity Model Prototype for Data Mining," *Computers*, vol. 14, no. 4, 2025, doi: 10.3390/computers14040146.

[23] C. N. Prabiantissa, "Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika," *SNESTIK Semin. Nas. Tek. Elektro, Sist. Inf. dan Tek. Inform.*, vol. 4, pp. 219–224, 2021, [Online]. Available: https://doi.org/10.31284/p.snestik.2021.1818

[24] S. Wijanarko and A. Santoso, "VOL. X NO. 1 FEBRUARI 2024 JURNAL TEKNIK INFORMATIKA STMIK ANTAR BANGSA Penerapan Fungsi Mid Dan Find pada Pembersihan Data Alamat," vol. X, no. 1, pp. 14–18, 2024, [Online]. Available: doi: https://doi.org/10.51998/jti.v10i1.557

[25] M. Atalya Angelus Leza, N. Widya Utami, and P. Anugrah Cahya Dewi, "Prediksi Prestasi Siswa Smas Katolik Santo Yoseph Denpasar Berdasarkan Kedisiplinan Dan Tingkat Ekonomi Orang Tua Menggunakan Metode Knowledge Discovery in Database Dan Algoritma Regresi Linier Berganda," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 373–379, 2024, doi: 10.36040/jati.v8i1.8754.

[26] U. Suriani, "Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5," *Journalcisa*, vol. 3, no. 2, pp. 55–66, 2023, [Online]. Available: doi: 10.51519/journalcisa.v4i2.393

[27] R. Kusumawati, "Analisis Geospasial Klaster Inovasi Mendakan Mawar Desa," *J. Mnemon.*, vol. 8, no. 1, pp. 83–91, 2025, [Online]. Available: doi: 10.36040/mnemonic.v8i1.13100

[28] A. S. Amer, N. Suarna, I. Ali, and D. I. Efendi, "PENGOPTIMALAN MODEL ASOSIASI PENJUALAN PRODUK DI KEDAI MINUMAN MENGGUNAKAN METODE ALGORITMA FP-GROWTH," vol. 13, no. 1, 2025, [Online]. Available: doi: 10.23960/jitet.v13i1.5881