

IMPLEMENTASI ALGORITMA ASOSIASI *FP-GROWTH* DAN KLASIFIKASI *K-MEANS* TERHADAP POLA PEMBELIAN KONSUMEN DI *MARKETPLACE SHOPEE*

Muhamad Arif Saifudin, Henni Endah Wahanani, Achmad Junaidi

Program Studi Informatika Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jawa Timur
19081010011@student.upnjatim.ac.id

ABSTRAK

Belanja online saat ini menjadi kebiasaan masyarakat, terutama dalam bidang fashion pakaian. Konsumen cenderung berbelanja produk fashion online, terutama saat bulan Ramadhan menjelang Idul Fitri, khususnya untuk busana koko atau busana muslim. Dalam menghadapi lonjakan permintaan tersebut, pemilik toko online perlu merancang strategi penjualan yang efektif. Oleh karena itu diperlukan strategi pemasaran menggunakan data mining menggunakan algoritma K-Means dan FP-Growth untuk menganalisis pola pembelian konsumen berdasarkan data histori penjualan di Shopee. K-Means digunakan untuk klusterisasi dan segmentasi pelanggan dengan model RFM, menghasilkan 3 kelompok dengan karakteristik unik. FP-Growth digunakan untuk mencari aturan asosiasi item, menghasilkan rekomendasi stok produk. Hasil menunjukkan kluster 2 dari K-Means memiliki karakteristik yang dominan dengan nama kluster “Superaktif”, dan FP-Growth mengidentifikasi beberapa item yang sering muncul dalam setiap transaksi serta item-item yang sering dibeli secara bersamaan dari dataset data transaksi penjualan.

Kata kunci : *FP-Growth, K-Means, Segmentasi Pelanggan, LRFM*

1. PENDAHULUAN

Perkembangan teknologi dan internet telah mengubah cara konsumen bertransaksi dan berbelanja, dengan belanja online menjadi opsi yang lebih mudah dan praktis. Proses belanja online memungkinkan konsumen memperoleh barang atau jasa secara langsung dan interaktif melalui Internet, menghilangkan perantara fisik. Shopee, sebagai salah satu platform e-commerce terkemuka di Indonesia, menjadi pilihan utama konsumen dengan menawarkan berbagai kemudahan dan sistem pembayaran yang beragam [1].

Bulan Ramadhan, menjelang Hari Raya Idul Fitri, menjadi periode di mana konsumen cenderung beralih ke belanja online, terutama untuk keperluan busana muslim atau baju koko. Tradisi ini menjadi bagian integral dari budaya masyarakat Indonesia, yang mencari keberagaman dalam berbelanja dan memiliki baju baru untuk merayakan Hari Raya Idul Fitri [2]. Pusatbusanamuslim1, sebagai salah satu toko baju koko di Shopee, berhadapan dengan persaingan sengit di pasar yang penuh dengan toko serupa.

Dalam menghadapi persaingan ini, pemilik bisnis perlu memahami perubahan perilaku konsumen dan merancang strategi pemasaran yang efektif [3]. Data transaksi penjualan, terutama selama periode Bulan Ramadhan, menjadi sumber informasi yang berharga untuk meningkatkan kinerja bisnis [4]. Oleh karena itu, analisis perilaku pembelian konsumen menjadi aspek yang krusial.

Penelitian ini menerapkan metode data mining, dengan fokus pada clustering dan association rule, untuk menganalisis data penjualan dari pusatbusanamuslim1 di Shopee selama Bulan

Ramadhan. Metode clustering digunakan untuk melakukan segmentasi pelanggan berdasarkan model LRFM (Length, Recency, Frequency, dan Monetary), sementara association rule digunakan untuk mengidentifikasi aturan asosiasi antar item yang sering dibeli bersama. Tujuan dari analisis ini adalah memberikan wawasan yang lebih mendalam tentang preferensi dan kebutuhan konsumen, serta merancang strategi pemasaran yang lebih efektif dan efisien. Hasil dari penelitian ini diharapkan dapat memberikan pandangan yang lebih tajam tentang perilaku konsumen dan membantu pemilik bisnis mengoptimalkan strategi pemasaran mereka selama Bulan Ramadhan. Data transaksi penjualan dapat diolah menjadi informasi yang berguna untuk meningkatkan penjualan produk, dan metode data mining menjadi alat yang efektif untuk mencapai hal ini.

Dalam menghadapi tantangan persaingan di platform e-commerce seperti Shopee, pemahaman mendalam tentang perilaku konsumen melalui analisis data mining menjadi kunci sukses [5]. Dengan demikian, penelitian ini tidak hanya berfokus pada meningkatkan penjualan selama Bulan Ramadhan tetapi juga memberikan landasan untuk strategi pemasaran yang lebih adaptif dan berkelanjutan.

Dataset yang digunakan dalam penelitian ini diambil dari data penjualan salah satu toko di e-commerce Shopee yaitu pusatbusanamuslim1 selama periode penjualan bulan ramadhan menjelang hari raya idul fitri yaitu tanggal 30 Maret 2023 – 12 April 2023, 10 April 2022 – 23 April 2022, 25 April 2021 – 8 Mei 2021, dan 3 Mei 2020 – 16 Mei 2020. Dataset yang digunakan diambil dari periode penjualan selama

bulan ramadhan dari tahun-tahun sebelumnya dan juga bulan ramadhan tahun ini karena fokus penelitian kali ini untuk meningkatkan penjualan produk selama bulan ramadhan. Setelah dataset didapatkan, dilakukan proses pre-processing untuk membersihkan data transaksi pembelian produk/barang harian karena fokus penelitian ini yaitu produk/barang harian. Hasil dari metode K-Means clustering yaitu terdapat 3 kelompok pelanggan yang terdiri dari klaster 0, klaster 1, dan klaster 2 dimana dengan hasil segmentasi pelanggan dengan LRFM klaster 2 merupakan klaster terbaik yang diberi nama "Pelanggan Superaktif". Sementara dari metode Asosiasi FP-Growth menghasilkan item yang sering muncul dalam setiap transaksi dan item yang sering dibeli secara bersamaan.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Menurut penelitian yang ditulis Panji Indra Pangestu dan rekan-rekan yang berjudul Analisis Segmentasi Pelanggan Berbasis Pecency Frequency Monetary (RFM) menggunakan algoritma K-Means pada tahun 2023 yang bertujuan untuk melakukan segmentasi pelanggan dari data transaksi penjualan Perusahaan dengan Algoritma K-Means model RFM menghasilkan 3 cluster dari 4168 pelanggan yaitu Low Loyalti, Medium Loyalti dan Highest Loyalti.

Menurut penelitian yang ditulis Neni Purwati dan rekan-rekan yang berjudul Komparasi Metode Apriori dan FP-Growth Data Mining Untuk Mengetahui Pola Penjualan pada tahun 2023 menggunakan algoritma asosiasi FP-Growth yang bertujuan untuk meningkatkan penjualan produk parfum yang menghasilkan The Most Tree Items dengan minimum support 20% yaitu Jo Malone 82.49%.

Menurut penelitian yang ditulis Apriansa dan rekan-rekan yang berjudul Sistem Rekomendasi Penentuan Poin Produk Menggunakan Algoritma FP-Growth Dan K-Means Clustering pada tahun 2022 menggunakan algoritma K-Means dan FP-Growth dengan tujuan perbaikan pendataan produk untuk mempermudah menemukan barang yang menghasilkan 27 kombinasi item produk dari FP-Growth dan tiga cluster dari K-Means.

Menurut penelitian yang ditulis Wiga Maulana Baihaqi dan rekan-rekan yang berjudul Penerapan Teknik Clustering Sebagai Strategi Pemasaran pada Penjualan Buku Di Tokopedia dan Shopee menggunakan algoritma K-Means pada tahun 2019 dengan tujuan meningkatkan strategi penjualan ditengah ketatnya persaingan menghasilkan tiga kategori yaitu banyak, sedang, dan sedikit yang menggambarkan jumlah buku yang harus disediakan.

2.2. Perilaku Konsumen

Perilaku konsumen merupakan studi tentang bagaimana para individu, kelompok maupun organisasi dalam memilih, membeli, menggunakan barang, jasa, ide atau pengalaman guna memuaskan

keinginan dan kebutuhan konsumen tersebut. Konsumen dapat dipengaruhi dengan berbagai faktor antara lain faktor budaya, sosial, pribadi, dan psikologis.

Faktor budaya merupakan pengaruh yang paling luas dan paling dalam pada konsumen. Budaya meliputi pengamatan yang menyeluruh terhadap sifat – sifat masyarakat secara utuh termasuk bahasa, pengetahuan, hukum, agama, kebiasaan makan, musik, kesenian, teknologi, pola kerja, produk, dan benda – benda lain yang menunjukkan sesuatu yang khas tentang masyarakat yang bersangkutan [6].

2.3. E-Commerce Shopee

E-commerce atau electronic commerce merupakan sebuah platform digital yang berhubungan dengan proses membeli, menjual, serta memperdagangkan barang, jasa melalui media internet. Menurut Hoffman dan Fodor, e-commerce dapat berjalan dengan baik apabila dijalankan berdasarkan prinsip 4C ini: connection (koneksi), creation (penciptaan), consumption (konsumsi) dan control (pengendalian).

Dengan semakin pesatnya pertumbuhan e-commerce di Indonesia lantas turut membuat perusahaan e-commerce Shopee ikut mengambil posisi dan menjadi salah satu ecommerce yang cukup banyak digandrungi konsumen, khususnya mahasiswa. Shopee adalah sebuah aplikasi yang bergerak dibidang jual beli secara online dan dapat diakses secara mudah dengan menggunakan smartphone. Shopee hadir dalam bentuk aplikasi yang memudahkan penggunaanya dalam melakukan kegiatan berbelanja secara online tanpa harus ribet menggunakan perangkat komputer.

2.4. Data Mining

Data Mining adalah proses ekstraksi informasi dari kumpulan data melalui penggunaan algoritma dan teknik yang melibatkan bidang ilmu statistik, mesin pembelajaran, dan sistem manajemen database. Data Mining digunakan untuk ekstraksi informasi penting yang tersembunyi dari dataset yang besar. Dengan adanya data mining maka akan didapatkan suatu permata berupa pengetahuan di dalam kumpulan data-data yang banyak jumlahnya [7].

2.5. Clustering K-Means

Algoritma K-Means Clustering adalah suatu metode penganalisaan data atau metode data mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. K-Means merupakan algoritma clustering yang berulang-ulang. Algoritma K-Means menetapkan nilai-nilai cluster (k) secara acak, dimana nilai tersebut menjadi pusat dari cluster atau disebut sebagai centroid, mean, atau means [8]. Tahapan melakukan clustering atau pengelompokan dengan metode K-Means dapat dijelaskan sebagai berikut:

- 1) Pilih jumlah cluster k
- 2) Inisialisasi k pusat cluster ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat cluster diberi nilai awal dengan angka-angka random.
- 3) Tempatkan semua data/ objek ke cluster terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke cluster tertentu ditentukan jarak antara data dengan pusat cluster. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat cluster. Jarak paling antara satu data dengan satu cluster tertentu akan menentukan suatu data masuk dalam cluster mana. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut:

$$D(i, j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2} \quad (1)$$

Dimana :

$D(i, j)$ = Jarak data ke i ke pusat cluster j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j ada atribut ke k

2.6. Model LRFM

Model LRFM merupakan metode yang sering dipadukan dengan metode clustering. Metode clustering dan model LRFM memiliki hubungan erat dalam mendapatkan hasil kelompok pelanggan berdasarkan karakteristik dari nilai atribut LRFM. Dalam proses clustering, model LRFM sangat diperlukan untuk mengetahui nilai loyalitas para pelanggan yang dilihat dari nilai atribut LRFM seperti melihat waktu terakhir transaksi yang dilakukan oleh pelanggan (recency), seberapa sering pelanggan tersebut sering melakukan transaksi (frequency), seberapa banyak nilai jumlah dalam satuan rupiah yang dikeluarkan oleh pelanggan saat melakukan transaksi (monetary), dan seberapa lama hubungan yang terjalin antara pelanggan dengan perusahaan (length) yang dihasilkan pada masing-masing pelanggan. Oleh karena itu jika proses clustering dilakukan tanpa menggunakan model LRFM maka perolehan hasil dalam menentukan kelompok pelanggan yang memiliki loyalitas dan berpotensi menguntungkan bagi perusahaan tidak akan berhasil.

$$1) L_{(i)} = TL_{(i)} - TF_{(i)} \quad (2)$$

Dimana :

$L_{(i)}$: length pada pelanggan i

$TL_{(i)}$: tanggal terakhir pelanggan i melakukan transaksi

$TF_{(i)}$: tanggal pertama kali pelanggan i melakukan Transaksi

$$2) R_{(i)} = TR - TL_{(i)} \quad (3)$$

Dimana :

$R_{(i)}$: recency pada pelanggan i

TR : tanggal penelitian

$TL_{(i)}$: tanggal terakhir pelanggan i melakukan transaksi

$$3) F_{(i)} = \sum_{t=TA}^{t=TR} fr_{(t)} \quad (4)$$

Dimana :

$F_{(i)}$: frequency pada pelanggan i

t : waktu dalam tanggal

TA : tanggal awal yang ditentukan

TR : tanggal penelitian

$fr_{(t)}$: jumlah transaksi yang dilakukan pelanggan i pada setiap waktu t

$$4) M_{(i)} = \sum_{t=TA}^{t=TR} mr_{(t)} \quad (5)$$

Dimana :

$M_{(i)}$: monetary pada pelanggan i

t : waktu dalam tanggal

TA : tanggal awal yang ditentukan

TR : tanggal penelitian

$mr_{(t)}$: jumlah uang yang dibayarkan pelanggan i untuk setiap transaksi pada waktu t

2.7. Association Rules FP-Growth

Association rules merupakan salah satu metode yang bertujuan mencari pola yang sering muncul diantara banyak transaksi, dimana setiap transaksi terdiri dari beberapa item. Sehingga metode ini akan menemukan hubungan antar item [9].

FP-Growth adalah algoritma yang dapat menemukan pola yang sering muncul dari sekumpulan data dan struktur data yang digunakan adalah FP-Tree pada algoritma FP-Growth. Dengan menggunakan FP-Tree, algoritma FP-Growth dapat langsung mengekstrak frequent itemset dari FP-Tree yang telah dibentuk dengan menggunakan prinsip divided and conquer [7].

Support (nilai penunjang) adalah persentase kombinasi item tersebut dalam database. Confidence (nilai kepastian) adalah kuatnya hubungan antar item dalam aturan asosiasi. Sebuah aturan asosiasi dikatakan interesting jika nilai support adalah lebih besar dari minimum support dan juga nilai confidence adalah lebih besar dari minimum confidence. Sementara untuk menguji nilai kevalid-an hubungan antar item menggunakan lift ratio. Adapun formula yang digunakan untuk menghitung nilai support, confidence, dan lift ratio pada persamaan (6), (7), dan (8) sebagai berikut:

$$\text{Support (A,B)} = \frac{\text{Jumlah transaksi A dan B}}{\Sigma \text{transaksi}} \times 100\% \quad (6)$$

$$\text{Confidence (A,B)} = \frac{\text{Jumlah transaksi A dan B}}{\Sigma \text{transaksi A}} \times 100\% \quad (7)$$

$$\text{Lift Rasio} = \frac{\text{support (A,B)}}{\text{support A} \times \text{support B}} \quad (8)$$

3. METODE PENELITIAN

3.1. Pengumpulan Data

Data yang dibutuhkan dalam penelitian ini adalah data histori penjualan dari salah satu toko yang ada di marketplace Shopee toko pusatbusanamuslim1 dengan periode penjualan tertentu yaitu periode penjualan Bulan Ramadhan menjelang Hari Raya Idul Fitri yaitu tanggal 30 Maret 2023 – 12 April 2023, 10 April 2022 – 23 April 2022, 25 April 2021 – 8 Mei 2021, dan 3 Mei 2020 – 16 Mei 2020.

3.2. Pengolahan Data

Data histori penjualan yang digunakan sebagai dataset terlebih dahulu melalui tahap preprocessing untuk memudahkan dan mengoptimalkan pengolahan data.

Tabel 1. Tahap *Preprocessing*

Data sebelum preprocessing	Setelah preprocessing
NIBRAS SARIMBIT LAIKA BLACK, NIBRAS NBC 20 GAMIS PUTIH DAN HITAM POLOS	NIBRAS SARIMBIT LAIKA BLACK NIBRAS NBC 20 GAMIS PUTIH DAN HITAM POLOS
NIBRAS SARIMBIT LAIKA BLACK, NIBRAS KOKO NK 101	NIBRAS SARIMBIT LAIKA BLACK NIBRAS KOKO NK 101
NIBRAS KOKO NSK 89 DAN GAMIS NB A89, NIBRAS KOOKO NK 109	NIBRAS KOKO NSK 89 DAN GAMIS NB A89 NIBRAS KOOKO NK 109

Tabel di atas merupakan perbandingan sebelum dan sesudah preprocessing. Dimana tujuan dari preprocessing yaitu mengubah data teks berupa nama barang menjadi unit yang lebih kecil.

3.3. Normalisasi Data LRFM

Pada proses ini dilakukan pengelompokkan konsumen berdasarkan perilaku pembelian mereka yang dapat membantu dalam memahami dan mengidentifikasi kelompok-kelompok konsumen yang memiliki karakteristik serupa dalam hal frekuensi pembelian, jumlah uang yang dikeluarkan, dan waktu sejak pembelian terakhir. Pada proses ini akan dilakukan normalisasi LRFM dengan dataset yang telah dilakukan preprocessing.

Tabel 2. *Data Transformation*

Username	Recency	Frequency	Amount
009kamal	710	2	460000
2uv9jxkjgs	355	1	150000
5ckqbsnqf3	366	4	920000
alhaq	7	4	920000
_byduri87	365	4	920000
_s789xl57h	1067	1	150000

Data di atas merupakan data transformation yang akan dilakukan proses normalisasi karena besarnya selisih antara atribut Monetary, Frequency, dan Recency. Proses normalisasi dilakukan dengan perhitungan seperti di bawah.

Untuk *username* 009kamal

Untuk mencari *Recency* dari *username* 009 kamal dapat dilakukan perhitungan :

$$\begin{aligned}\text{Normalized_R} &= (R - R_{\min}) / (R_{\max} - R_{\min}) \\ &= (710 - 7) / (1067 - 7) \\ &= 703 / 1060 \\ &= 0.66\end{aligned}$$

Untuk mencari *Frequency* dari *username* 009 kamal dapat dilakukan perhitungan :

$$\begin{aligned}\text{Normalized_F} &= (F - F_{\min}) / (F_{\max} - F_{\min}) \\ &= (2 - 1) / (4 - 1) \\ &= 1 / 3 \\ &= 0.25\end{aligned}$$

Untuk mencari *Monetary* dari *username* 009 kamal dapat dilakukan perhitungan :

$$\begin{aligned}\text{Normalized_M} &= (M - M_{\min}) / (M_{\max} - M_{\min}) \\ &= (460000 - 150000) / (920000 - 150000) \\ &= 0.31\end{aligned}$$

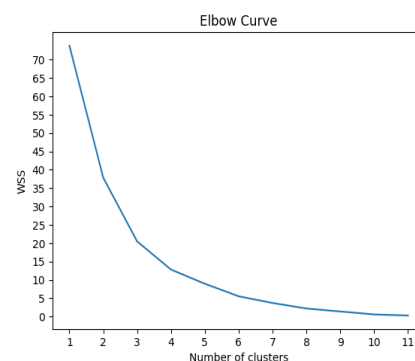
Tabel 3. Normalisasi *LRFM*

Username	Amount	Frequency	Recency
009kamal	0.31	0.25	0.66
2uv9jxkjgs	0.00	0.00	0.33
5ckqbsnqf3	0.77	0.75	0.34
alhaq	0.77	0.75	0.006
_byduri87	0.77	0.75	0.33
_s789xl57h	0.00	0.00	0.99

Tabel di atas merupakan tabel hasil normalisasi dengan tujuan untuk mengubah nilai antara rentang 0 hingga 1.

3.4. Elbow Curve

Metode Elbow merupakan teknik yang digunakan untuk mendapatkan informasi dalam menentukan jumlah kluster yang optimal dengan cara memeriksa perbandingan hasil antara jumlah kluster yang membentuk siku pada suatu titik [10].



Gambar 1. *Output Metode Elbow*

Gambar di atas merupakan hasil dari metode Elbow Dimana dapat dilihat ada beberapa nilai WSS dari K yang mengalami penurunan paling besar selanjutnya akan turun secara perlahan sampai nilai WSS dari K stabil. Terlihat penurunan drastic membentuk siku pada titik K=3, sehingga nilai K paling ideal yang digunakan dalam proses pengujian clustering yaitu K=3.

3.5. Clustering

```
[0 1 2 2 0 0 0 0 0 0 0 1 1 1 1 0 0 0 1 1 0 0 0 1 0 0 0 0 1 1 0 2 2 0 0
0 2 0 0 0 0 0 1 0 0 0 0 1 2 0 0 0 0 1 0 1 0 0 1 0 1 1 0 1 0 1 0 0 0 1
1 2 0 1 1 0 1 0 1 0 0 0 0 1 0 1 1 1 2 0 1 0 0 0 0 0 0 1 0 0 0 2 0 1 1 0 1
1 0 0 0 0 1 0 0 0 0 0 1 0 0 1 2 0 0 1 0 0 0 0 1 0 1 1 1 1 0 2 0 0 1 0 1
1 1 0 1 1 0 1 0 1 0 0 1 0 0 1 0 0 0 1 1 1 0 0 0 1 1 2 0 1 1 0 1 2 0 2 0 0
0 1 1 0 0 0 1 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 2 0 1 1 0 0 0 0 0 0 0 1 1 0
1 0 1 0 2 1 1 0 0 0 0 0 0 0 0 0 1 0 0 1 0 1 0 0 1 0 2 0 2 0 0 1 1 1 0 1
0 2 2 1 0 2 0 1 1 1 0 1 0 0 0 0 2 1 0 0 1 1 0 1 2 1 1 0 1 1 0 1 0 0 0 0 0
2 0 2 2 1 0 0 1 0 2 0 0 0 0 0 2 1 1 0 1 2 0 1 1 0 1 0 0 1 1 1 1 0 0 0 0 1
1 0 1 1 0 0 0 0 0 1 0 1 0]
```

Silhouette Score: 0.5641087778885784

Gambar 2. Output Clustering

Gambar di atas merupakan hasil dari proses model K-Means dengan penetapan parameter yang digunakan yaitu 3 klaster, 10 insialisasi acak dan maksimal 50 iterasi untuk setiap inisialisasi.

3.6. Pencarian Minimum Support dan Lift Ratio

Pada proses penerapan algoritma FP-Growth pada penelitian ini setelah data dilakukan preprocessing, kemudian dicari kemunculan setiap item untuk diproses agar menghasilkan nilai support dari setiap item.

Tabel 4. Frekuensi Tiap Item

Kode	Item	Frekuensi
A	NIBRAS SARIMBIT 70 COKLAT	86
B	NIBRAS ATASAN PRIA NK 120	51
C	NIBRAS KOKO NK 52	40
D	NIBRAS SARIMBIT FAIDA PLUM	30
E	NIBRAS SARIMBIT CARYNA ARMY COUPLE	29

Tabel di atas merupakan tabel kemunculan setiap item Dimana dari informasi tersebut dapat dicari nilai dari masing-masing minimum support, minimum confidence, dan lift ratio.

$$\begin{aligned} \text{Minimum Support (A)} &= \frac{\text{Jumlah transaksi A}}{\Sigma \text{transaksi}} \times 100\% \\ &= \frac{86}{399} \times 100\% \\ &= 21.55\% \end{aligned}$$

$$\begin{aligned} \text{Minimum Confidence (A,B)} &= \frac{\text{Jumlah transaksi A dan B}}{\Sigma \text{transaksi A}} \times 100\% \\ &= \frac{151}{86} \times 100\% \\ &= 175.85\% \end{aligned}$$

$$\begin{aligned} \text{Lift Ratio} &= \frac{\text{support (A,B)}}{\text{support A} \times \text{support B}} \\ &= \frac{37.84}{275.44} \\ &= 0.1375 \end{aligned}$$

3.7. Pembentukan FP-Tree

Pada proses pembentukan FP-Tree kali ini langkah awal yang dilakukan yaitu mencari item yang muncul disetiap transaksi dalam transaksi.

4. HASIL DAN PEMBAHASAN

Pada proses implementasi data mining dalam penelitian kali ini memakai software Visual StudioCode menggunakan bahasa Python dengan algoritma yang digunakan yaitu K-Means dan FP-Growth dengan tujuan untuk segmentasi pelanggan dan penentuan rekomendasi stok produk. Hasil dari implementasi dua algoritma sebagai berikut.

4.1. Segmentasi Pelanggan

Pada proses ini dilakukan pencarian agregasi dari hasil normalisasi LRFM dengan mencari mean, median, serta standart deviasi untuk setiap atribut LRFM.

Cluster_Id		mean	median	std
0	Amount	354600.000000	460000.0	147218.225876
	Frequency	1.660000	2.0	0.474898
	Recency	887.465000	889.0	178.304998
1	Amount	312327.586207	300000.0	130918.878026
	Frequency	1.594828	2.0	0.493055
	Recency	202.422414	354.0	175.253563
2	Amount	917333.333333	920000.0	71628.030216
	Frequency	4.000000	4.0	0.262613
	Recency	197.500000	360.5	182.008857

Gambar 3. Output Agregasi Segmentasi Pelanggan

Dari gambar di atas dapat disimpulkan bahwa :

- Cluster 0: "Pelanggan Konsisten" Pelanggan pada klaster ini memiliki karakteristik waktu terakhir pembelian yang relatif lama, frekuensi pembelian yang tidak terlalu tinggi, dan jumlah pembelian yang menengah. Meskipun tidak sering berbelanja, mereka tetap konsisten dalam pembelian.
- Cluster 1: "Pelanggan Aktif" Pelanggan pada klaster ini memiliki karakteristik waktu terakhir pembelian yang relatif lama, frekuensi pembelian yang tidak terlalu tinggi, dan jumlah pembelian yang menengah. Meskipun tidak sering berbelanja, mereka tetap konsisten dalam pembelian.
- Cluster 2: "Pelanggan Superaktif" Pelanggan pada klaster ini memiliki karakteristik waktu terakhir pembelian yang tidak terlalu rendah, frekuensi pembelian yang tinggi, dan jumlah pembelian yang besar. Pelanggan dalam klaster ini cenderung sering berbelanja dalam jumlah yang signifikan. Kemudian dilakukan pelabelan hasil cluster terhadap hasil normalisasi dengan model LRFM.

Tabel 5. Pelabelan Cluster Hasil LRFM

Username	Recency	Frequency	Amount	Cluster ID
009kamal	0.661080	0.25	0.31	1
2uv9jxkjgs	0.330540	0.00	0.00	2
5ckqbsnqf3	0.340782	0.75	0.77	0
alhaq	0.006518	0.75	0.77	0
_byduri87	0.339851	0.75	0.77	0
yyumtlvuvk	0.993482	0.00	0.00	1
zahraserlika	0.004655	0.00	0.00	2

Tabel di atas merupakan pelabelan hasil dari proses LRFM dengan hasil Cluster. Dimana menunjukkan Cluster dengan rentang 0 sampai 1 dari hasil LRFM segmentasi pelanggan.

4.2. Proses Pencarian Frequent Itemset

Pada proses pencarian frekuensi tiap item dilakukan untuk mempresentasikan setiap item dalam setiap transaksi dan menghitung kemunculan setiap item.

	Items	incident_count
0	NIBRAS SARIMBIT 70 COKLAT	86
1	NIBRAS ATASAN PRIA NK 120	51
2	NIBRAS KOKO NK 52	40
3	NIBRAS SARIMBIT FAIDA PLUM	30
4	NIBRAS SARIMBIT CARYNA ARMY COUPLE	29
5	NIBRAS SARIMBIT LAIKA BLACK	29
6	NIBRAS SARIMBIT SYARI	25
7	NIBRAS SARIMBIT CHESA GREY COUPLE	23
8	NIBRAS KOKO POLOS NK B3	23
9	NIBRAS SARIMBIT FAIDAA GREY LILAC	22
10	NIBRAS SARIMBIT ESTELA NAVY	22
11	NIBRAS GAMIS NB B108	21
12	NIBRAS NBC 20 GAMIS PUTIH DAN HITAM POLOS	20
13	NIBRAS SARIMBIT DAANIA BURGUNDY COUPLE	19
14	NIBRAS SARIMBIT DAANIA TAUPE GREY COUPLE	16
15	NIBRAS KOKO NSK 95 DAN GAMIS NB B95 COUPLE	15
16	NIBRAS KOKO NSK 89 DAN GAMIS NB A89	15
17	NIBRAS SARIMBIT KHAWLA VIOLET	15
18	NIBRAS KOKO NK 101	14
19	NIBRAS KOKO NK 109	14

Gambar 4. Output Pencarian Frequent Itemset

Gambar di atas merupakan hasil dari proses pencarian frekuensi setiap item dalam semua data transaksi penjualan. Pada gambar tersebut terdapat nama barang dan jumlah kemunculan item.

4.3. Visualisasi dengan Treemap Chart

Pada proses dilakukan visualisasi hasil analisis dengan treemap chart. Tujuan dari proses ini untuk memberikan visualisasi yang lebih jelas dan informatif mengenai 50 item teratas berdasarkan frekuensi kemunculan setiap item.



Gambar 5. Visualisasi Treemap Chart

Gambar di atas menunjukkan hasil visualisasi dengan treemap chart. Gambar tersebut dapat dilihat item yang paling sering muncul dengan informasi tiap incident count tiap item. Item dengan jumlah paling banyak akan berwarna biru tua dengan penurunan warna sesuai dengan penurunan jumlah item.

4.4. Implementasi Association Rules FP-Growth

Selanjutnya akan dilakukan proses untuk mengidentifikasi pola asosiasi dalam dataset dengan menetapkan parameter minimum support sebesar 0.001. Tujuannya untuk mencari item yang sering muncul dalam setiap transaksi berdasarkan nilai supportnya.

support	itemsets
0 0.0725	(NIBRAS SARIMBIT LAIKA BLACK)
1 0.0550	(NIBRAS SARIMBIT ESTELA NAVY)
2 0.0500	(NIBRAS NBC 20 GAMIS PUTIH DAN HITAM POLOS)
3 0.0350	(NIBRAS KOKO NK 101)
4 0.0375	(NIBRAS KOKO NSK 89 DAN GAMIS NB A89)
...	...
138 0.0025	(SARIMBIT KHAWLA VIOLET, NIBRAS KOKO NK 101 NI...
139 0.0025	(NIBRAS GAMIS NB B108 NIBRAS, SARIMBIT DAANIA ...
140 0.0050	(NIBRAS SARIMBIT CHESA BROWN COUPLE, NIBRAS S...
141 0.0025	(NIBRAS SARIMBIT 70 COKLAT, NIBRAS SARIMBIT FA...
142 0.0025	(NIBRAS SARIMBIT 70 COKLAT, NIBRAS KOKO NK 109)

43 rows × 2 columns

Gambar 6. Proses Association Rules

Gambar di atas merupakan hasil dari proses implementasi Association Rules dengan Algoritma FP-Growth. Dimana dalam hasil tersebut terdapat nama barang dan juga nilai supportnya. Dimana terlihat Nibras Sarimbit Lika Black memiliki nilai support terbesar 0.00725 yang berarti item tersebut muncul sebanyak 0.725 % dalam setiap transaksi.

4.5. Proses Pencarian Aturan Asosiasi Item

Pada proses kali ini akan dicari aturan asosiasi tiap item. Proses ini bertujuan untuk menemukan aturan keterkaitan antar item satu dengan lainnya dengan menggunakan parameter minimum confidence 0.5.

	antecedents	consequents	antecedent support	consequent support	support confidence	lift	leverage	conviction	change metric	
0	(NIBRAS KOKO NSK 89 DAN GAMIS NB A89)	(NIBRAS COUPLE GAMIS NB B94 DAN KOKO NSK 95)	0.0025	0.0025	0.0025	1.0	35.769231	0.002419	inf	0.809925
1	(NIBRAS KOKO NSK 89 DAN GAMIS NB A89)	(NIBRAS NB B93 DAN KOKO NSK 95 COUPLE)	0.0025	0.0025	0.0025	1.0	35.769231	0.002401	inf	0.874937
2	(NIBRAS KOKO NSK 89 DAN GAMIS NB A89)	(NIBRAS KOKO NSK 95 DAN GAMIS NB B95 COUPLE)	0.0025	0.0025	0.0025	1.0	25.666667	0.002406	inf	0.844912
3	(NIBRAS COUPLE GAMIS NB B94 DAN KOKO NSK 95)	(NIBRAS KOKO NK 110)	0.0025	0.0025	0.0025	1.0	40.000000	0.002407	inf	0.977444
4	(NIBRAS COUPLE GAMIS NB B94 DAN KOKO NSK 95)	(NIBRAS KOKO NK 110)	0.0025	0.0025	0.0025	1.0	35.769231	0.002419	inf	0.809925
5	(NIBRAS KOKO NK 101 NIBRAS)	(SARIMBIT KHAWLA VIOLET)	0.0025	0.0025	0.0025	1.0	40.000000	0.002404	inf	1.000000
6	(SARIMBIT KHAWLA VIOLET)	(NIBRAS KOKO NK 101 NIBRAS)	0.0025	0.0025	0.0025	1.0	40.000000	0.002404	inf	1.000000
7	(SARIMBIT DAANIA BURGUNDY COUPLE)	(NIBRAS GAMIS NB B108 NIBRAS)	0.0025	0.0025	0.0025	1.0	40.000000	0.002404	inf	1.000000
8	(NIBRAS GAMIS NB B108 NIBRAS)	(SARIMBIT DAANIA BURGUNDY COUPLE)	0.0025	0.0025	0.0025	1.0	40.000000	0.002404	inf	1.000000
9	(NIBRAS SARIMBIT CHESA BROWN COUPLE)	(NIBRAS SARIMBIT 70 COKLAT)	0.0025	0.2150	0.0025	1.0	4.831163	0.002625	inf	0.780945
10	(NIBRAS SARIMBIT FAIDA PLUM)	(NIBRAS SARIMBIT 70 COKLAT)	0.0025	0.2150	0.0025	1.0	4.831163	0.002625	inf	0.780945
11	(NIBRAS KOKO NK 109)	(NIBRAS SARIMBIT 70 COKLAT)	0.0025	0.2150	0.0025	1.0	4.831163	0.002625	inf	0.780945

Gambar 7. Proses Pencarian Aturan Asosiasi

Gambar di atas merupakan hasil dari proses pencarian aturan asosiasi antar item. Dari gambar di atas terdapat item dengan nilai support, confidence, dan lift ratio.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan dengan mengimplementasikan algoritma klasifikasi K-Means dan asosiasi FP-Grwoth dengan data uji yaitu dataset laporan penjualan produk dengan total 349 transaksi diperoleh hasil dengan algoritma K-Means untuk segmentasi pelanggan dengan model LRFM yaitu pelanggan paling baik berada di klaster 2 dengan nama "Pelanggan Superaktif" dengan karakteristik waktu terakhir pembelian yang tidak terlalu rendah, frekuensi pembelian yang tinggi, dan jumlah pembelian yang besar. Pelanggan dalam klaster ini cenderung sering berbelanja dalam jumlah yang signifikan. Sementara hasil dari implementasi algoritma FP-Grwoth menghasilkan item yang sering muncul dalam setiap transaksi yaitu Nibras Sarimbit Laika Black dengan nilai support 0.0725 dan item yang sering dibeli secara bersamaan yaitu Nibras Sarimbit Chesa Brown Couple dengan Nibras Sarimbit 70 Coklat dengan Tingkat kevaliid-an sebesar 4.65.

DAFTAR PUSTAKA

- [1] A. Elmayanti, "Fenomena Perilaku Belanja Online Sebagai Alternatif Pilihan Konsumsi Di Kalangan Remaja," *Jurnal Dinamika Sosial Budaya*, p. 9, 2023.
- [2] A. N. Harahap, "Consumptive and Productive Behavior and Nature of New Students of Business Management at The Shipbuilding Institute of Polytechnic Surabaya," *IJSED*, p. 10, 2022.
- [3] W. M. Baihaqi, "Penerapan Teknik Clustering Sebagai Strategi Pemasaran pada Penjualan Buku Di Tokopedia dan Shopee," *PARADIGMA (Jurnal Informatika dan Komputer)*, p. 6, 2019.
- [4] R. D. Lestari, "PENERAPAN DATA MINING UNTUK MENGETAHUI CUSTOMER BEHAVIOUR TOKO ARTPEDIA DENGAN MENGGUNAKAN ALGORITMA APRIORI," *E-Proceeding of Engineering*, p. 7, 2020.
- [5] D. Mahendra, "SISTEM ANALISIS POLA PEMBELIAN KONSUMEN MENGGUNAKAN ALGORITMA FP-GROWTH," *JTIK (JURNAL ILMIAH TEKNOLOGI INFORMASI DAN KOMUNIKASI)*, p. 9, 2023.
- [6] M. S. Giantara, "PENGARUH BUDAYA, SUB BUDAYA, KELAS SOSIAL, DAN PERSEPSI KUALITAS TERHADAP PERILAKU KEPUTUSAN PEMBELIAN KUE TRADISIONAL OLEH MAHASISWA DI SURABAYA," *NELITI*, p. 17, 2020.
- [7] D. Winarti, "Penerapan Data Mining untuk Analisa Tingkat Kriminalitas Dengan Algoritma Association Rule Metode FP-Growth," *Jurnal SIMTIKA*, p. 15, 2021.
- [8] S. Handoko, "IMPLEMENTASI DATA MINING UNTUK MENENTUKAN TINGKAT PENJUALAN PAKET DATA TELKOMSEL MENGGUNAKAN METODE K-MEANS CLUSTERING," *Jurnal Ilmiah Teknologi dan Rekayasa*, p. 13, 2020.
- [9] S. P. Tamba, "PENERAPAN DATA MINING ALGORITMA APRIORI DALAM MENENTUKAN STOK BAHAN BAKU PADA RESTORAN NELAYAN MENGGUNAKAN METODE ASSOCIATION RULE," *JUSIKOM PRIMA (Jurnal Sistem Informasi dan Ilmu Komputer Prima)*, p. 6, 2022.
- [10] N. E. Merliana, "ANALISA PENENTUAN JUMLAH CLUSTER TERBAIK PADA METODE K-MEANS CLUSTERING," *SENDI*, p. 6, 2023.
- [11] S. Aisyah, "Implementasi Algoritma FP-Growth Terhadap Transaksi Penjualan Produk Kopi," *KOPERTIP*, p. 6, 2020.
- [12] R. Apriansa, "Sistem Rekomendasi Penentuan Poin Produk Menggunakan Algoritma FP-Growth Dan K-Means Clstering,," *Seminar Nasional Inovasi Teknologi*, p. 6, 2022.
- [13] F. H. Azhra, "Marketing strategy palnning at alfamart lodadi stores using the clustering, ahp, and ar-mba," *International Journal of Industrial Optimization*, p. 14, 2021.
- [14] A. Damayanti, "PEMANTAUAN PENGENDALIAN PERSEDIAAN BAHAN PRODUKSI BAJU KOKO PADA KONVEKSI MANFAAT," *Jurnal Mahasiswa Manajemen UNITA*, p. 8, 2022.
- [15] R. Damayanti, "Menu Package Recommendation Using Combination of KMeans and FP-Growth Algorithms at Bakery Stores," *Jurnal Mantik*, p. 6, 2020.
- [16] Feryanto, "PENERAPAN DATA MINING UNTUK MENENTUKAN PENJUALAN SPAREPART TOYOTA DENGAN METODE K-MEANS CLUSTERING," *Jurnal Sistem Informasi Ilmu Komputer Prima (JUSIKOM PRIMA)*, p. 6, 2019.
- [17] M. S. Giantara, "PENGARUH BUDAYA, SUB BUDAYA, KELAS SOSIAL, DAN PERSEPSI KUALITAS TERHADAP PERILAKU KEPUTUSAN PEMBELIAN KUE TRADISIONAL OLEH MAHASISWA DI SURABAYA," *NELITI*, p. 17, 2020.

- [18] Habriyanto, "Analisis Perbandingan Pola Konsumsi Masyarakat Kota Jambi Sebelum, Sesaat an Sesudah Bulan Ramadhan," *Al-Tijary Jurnal Ekonomi dan Bisnis Islam*, p. 13, 2019.
- [19] A. F. Hasanah, "ANALISIS PERILAKU KONSUMEN MASYARAKAT PONOROGO SESAAT DAN SESUDAH DATANGNYA BULAN," *IJOIS: Indonesian Journal of Islamic Studies*, p. 12, 2020.
- [20] B. F. Hasugian, "Penerapan Metode Association Rule Untuk Menganalisa Pola Pemakaian Bahan Kimia Di Laboratorium Menggunakan Algoritma FP-Growth (Studi Kasus di Laboratorium KIMIA PT. PLN (Persero) Sektor Pembangkitan Belawan Medan,," *ALGORITMA: Jurnal Ilmu Komputer dan Informatika*, p. 14, 2019.
- [21] R. Hidayat, "Analisis Clustering Dalam Pengelompokan Penjualan Menggunakan Algoritma," *CESS (Journal of Computing Engineering, System and Science)*, p. 15, 2022.
- [22] D. Melati, "ASSOCIAION RULEDALAM MENENTUKAN CROSSSELLINGPRODUK MENGGUNAKAN ALGORITMA FP-GROWTH," *VOTEKNIKA*, p. 10, 2019.
- [23] A. R. Mulyawan, "MARKETING MAPS PADA LEMBAGA AMIL ZAKAT MENGGUNAKAN ALGORITMA CLUSTERING DAN ASSOCIATION RULES," *Jurnal Sistem Informasi*, p. 15, 2020.
- [24] Mustakim, "STRATEGI PROMOSI PERGURUAN TINGGI MENGGUNAKAN MODEL . CLUSTERING DAN ASOSIATION," p. 6, 2022.
- [25] J. Nasir, "PENERAPAN DATA MINING CLUSTERING DALAM MENGELOMPOKAN BUKU DENGAN METODE K-MEANS," *SIMETRIS*, p. 13, 2020.
- [26] F. Nuraeni, "OPTIMIZATION OF MARKET BASKET ANALYSIS USING CENTROID-BASED CLUSTERING ALGORITHM AND FPGROWTH ALGORITHM," *Jurnal Teknik Informatika (JUTIF)*, p. 10, 2022.
- [27] Octavia, "Penerapan Algoritma Association Rules Dalam Penentuan Pola Pembelian Berdasarkan Hasil Clustering," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, p. 10, 2023.
- [28] P. I. Pangestu, "ANALISIS SEGMENTASI PELANGGAN BERBASIS REGENCY FREQUENCY MONETARY (RFM) MENGGUNAKAN ALGORITMA K-MEANS," *JATI (Jurnal Mahasiswa Teknik Informatika)*, p. 7, 2023.
- [29] D. Pranitasari, "Analisis Kepuasan Pelanggan Elektronik Shopee Menggunakan Metode E-Service Quality dan Kartesius," *Jurnal Akuntansi dan Manajemen (JAM)*, p. 20, 2021.
- [30] N. Purwati, "Komparasi Metode Apriori dan FP-Growth Data Mining Untuk Mengetahui Pola Penjualan," *Jurnal Pengembangan IT (JPIT)*, p. 7, 2023.
- [31] U. D. Puspita, "SEGMENTASI CUSTOMER LIFETIME VALUE PADA MODEL LRFM MENGGUNAKAN METODE K-MEANS EUCLIDEAN DISTANCE," *Bimaster*, p. 10, 2023.
- [32] A. Rahmadsyah, "Analisa Association Rule Pada Algoritma Apriori Untuk Minat Pembelian Alat Kesehatan," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, p. 7, 2021.
- [33] A. Rahman, "Analisis Data Produk Elektronik Di E-Commerce Dengan Metode Algoritma K-Means Menggunakan Python," *JAIIT (Journal Of Advance in Information and Industrial Technology)*, p. 8, 2021.
- [34] A. Saepuddin, "ANALISA PERILAKU KONSUMEN: STRATEGI MEMENANGKAN PERSAINGAN BISNIS DI ERA EKONOMI DIGITAL," *G-Tech Jurnal Teknologi Terapan*, p. 11, 2020.
- [35] I. Syukra, "Implementation of K-Medoids and FP-Growth Algorithms for Grouping and Product Offering Recommendations," *Indonesian Journal of Artifical Intellegence and Data Mining*, p. 10, 2019.
- [36] Warisa, "STUDI KASUS PENJUALAN KOSMETIK MENGGUNAKAN METODE ASSOCIATION RULE (APRIORI)," *Technologia*, p. 5, 2021.
- [37] K. Wijaya, "Penerapan Algoritma FP-Growth Untuk Analisis Data Transaksi Penjualan Di Internet Learning Cafe Kaliurang," *Jurnal Nasional Komputasi dan Teknologi Informasi*, p. 10, 2022.
- [38] D. Winarti, "Penerapan Data Mining untuk Analisa Tingkat Kriminalitas Dengan Algoritma Association Rule Metode FP-Growth," *Jurnal SIMTIKA*, p. 15, 2021.