

**ANALISIS PERILAKU KONSUMEN PADA USAHA RITEL DENGAN  
MENGUNAKAN METODE *ASSOCIATION RULE - MARKET BASKET ANALYSIS*  
DAN *CLUSTERING* SEBAGAI USULAN STRATEGI PENINGKATAN PENJUALAN  
(Studi Kasus: INTIMART GEDONGAN)**

**TUGAS AKHIR**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Strata-1  
Program Studi Teknik Industri - Fakultas Teknologi Industri  
Universitas Islam Indonesia**



Nama : Muhammad Rafly Qowi Baihaqie  
No. Mahasiswa : 19522109

**PROGRAM STUDI TEKNIK INDUSTRI  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ISLAM INDONESIA  
YOGYAKARTA  
2023**

## PERNYATAAN KEASLIAN

ii

### PERNYATAAN KEASLIAN

Saya mengakui bahwa tugas akhir ini adalah hasil karya saya sendiri kecuali kutipan dan ringkasan yang seluruhnya sudah saya jelaskan sumbernya. Jika dikemudian hari ternyata terbukti pengakuan saya ini tidak benar dan melanggar peraturan yang sah maka saya bersedia ijazah yang telah saya terima ditarik kembali oleh Universitas Islam Indonesia.

Yogyakarta, 07 Agustus 2023



(Muhammad Rafly Qowi Baihaqie)

19522109

**SURAT BUKTI PENELITIAN****Cv.Intimart**

Alamat : Gedongan Sinduadi Melati Sleman Yogyakarta  
No.Telp.0877-1313-5029,Kode Pos:55284.

**SURAT KETERANGAN**

Kami yang bertandatangan dibawah ini, menerangkan bahwa mahasiswa dengan keterangan sebagai berikut:

Nama : Muhammad Rafly Qowi Baihaqie  
No. Mhs : 19522109  
Perguruan Tinggi : Universitas Islam Indonesia  
Fakultas : Fakultas Teknologi Industri  
Jurusan : Teknik Industri

Yang bersangkutan telah selesai melakukan penelitian di Toko Intimart Gedongan selama 2 (dua) bulan, terhitung mulai tanggal 01 Juni 2023 sampai dengan 01 Agustus 2023. Untuk memperoleh data dalam rangka penyusunan Tugas Akhir yang berjudul **“ANALISIS PERILAKU KONSUMEN PADA USAHA RITEL DENGAN MENGGUNAKAN METODE ASSOCIATION RULE - MARKET BASKET ANALYSIS DAN CLUSTERING SEBAGAI USULAN STRATEGI PENINGKATAN PENJUALAN (Studi Kasus: INTIMART GEDONGAN)”**

Demikian surat keterangan ini dibuat dan diberikan kepada yang bersangkutan untuk dapat digunakan sebagaimana mestinya, atas perhatian dan kerja samanya yang baik, kami ucapkan terima kasih.

Yogyakarta, 7 Agustus 2023

Hormat Kami

  
(Sudyanto)

Gedongan, Sinduadi, Mlati, Sleman  
Daerah Istimewa Yogyakarta - 55284

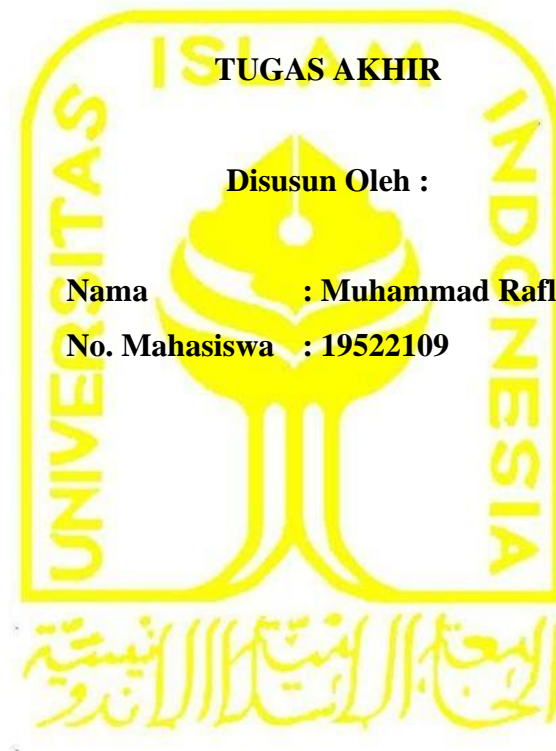
intimart.official@gmail.com

+62 856 4316 4598



**LEMBAR PENGESAHAN PEMBIMBING**

**ANALISIS PERILAKU KONSUMEN PADA USAHA RITEL DENGAN  
MENGUNAKAN METODE ASSOCIATION RULE - MARKET BASKET  
ANALYSIS DAN CLUSTERING SEBAGAI USULAN STRATEGI PENINGKATAN  
PENJUALAN (Studi Kasus: INTIMART GEDONGAN)**



**Yogyakarta, 08 Agustus 2023**

**Dosen Pembimbing**

**Annisa Uswatun Khasanah, S.T., M.B.A., M.Sc.**

**LEMBAR PENGESAHAN DOSEN PENGUJI**

**ANALISIS PERILAKU KONSUMEN PADA USAHA RITEL DENGAN  
MENGUNAKAN METODE ASSOCIATION RULE - MARKET BASKET  
ANALYSIS DAN CLUSTERING SEBAGAI USULAN STRATEGI PENINGKATAN  
PENJUALAN (Studi Kasus: INTIMART GEDONGAN)**

**TUGAS AKHIR**

**Disusun Oleh :**

**Nama : Muhammad Rafly Qowi Baihaqie**

**No. Mahasiswa : 19 522 109**

**Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk  
memperoleh gelar Sarjana Strata-1 Teknik Industri Fakultas Teknologi Industri  
Universitas Islam Indonesia**

**Yogyakarta, 07 - September - 2023**

**Tim Penguji**

Annisa Uswatun Khasanah, S.T., M.B.A., M.Sc.

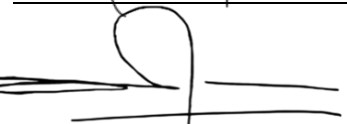
Ketua

Dr. Drs. Imam Djati Widodo, M.Eng.Sc.

Anggota I

Danang Setiawan, S.T., M.T.

Anggota II

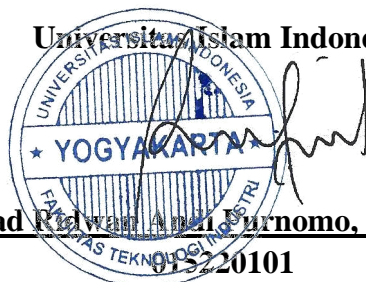




**Mengetahui,**

**Ketua Program Studi Teknik Industri Program Sarjana**

**Fakultas Teknologi Industri**

**Universitas Islam Indonesia**



**Ir. Muhammad Ridwan Anandjurnomo, ST, Msc, PhD, IPM**

**015220101**

## **HALAMAN PERSEMBAHAN**

Dengan tulus ikhlas dan penuh rasa cinta, rasa syukur, serta penghormatan yang setinggi-tingginya, skripsi ini saya persembahkan kepada kedua orang tua tercinta, Ibu Desi Arisanti dan Bapak Yan Kurniadi. Kehadiran, dukungan tanpa henti, kasih sayang, serta dorongan Ibu dan Bapak telah memberikan cahaya dan arahan dalam setiap langkah perjalanan hidup saya. Tak lupa, penghargaan dan ucapan terima kasih yang tak terhingga juga saya sampaikan kepada saudara-saudara serta teman-teman terdekat yang senantiasa memberikan semangat, dukungan, dan doa-doa tulus.

**MOTTO**

*"Janganlah kamu bersikap lemah, dan janganlah (pula) kamu bersedih hati, padahal kamulah orang-orang yang paling tinggi (derajatnya), jika kamu orang-orang yang beriman."*

(Surat Ali Imran ayat 139)

*"Maybe I'm Not Late. Maybe You Guys Are Early."*

(Miles Morales)

*"That which does not kill us makes us stronger."*

(Friedrich Nietzsche)

## KATA PENGANTAR

*Assalamu 'alaikum warahmatullahi wabarakatuh.*

*Alhamdulillah* *rabbi'l' alamin*, segala puji syukur penulis panjatkan kepada Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis senantiasa dalam keadaan sehat dan dapat menyelesaikan Tugas Akhir yang berjudul “Analisis Perilaku Konsumen Pada Usaha Ritel Dengan Menggunakan Metode *Association Rule - Market Basket Analysis* dan *Clustering* Sebagai Usulan Strategi Peningkatan Penjualan (Studi Kasus: Intimart Gedongan)”. Tak lupa penulis mengucapkan shalawat serta salam yang dipersembahkan kepada Nabi besar Muhammad SAW, utusan Allah yang telah membawa cahaya petunjuk bagi seluruh umat manusia.

Tugas Akhir merupakan salah satu prasyarat kelulusan untuk menyelesaikan program studi S-1, Jurusan Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia. Penyusunan Tugas Akhir ini tidaklah berjalan tanpa rintangan, namun di balik setiap tantangan tersebut, penulis merasakan dukungan yang luar biasa dari berbagai pihak. Penulis ingin mengungkapkan rasa terima kasih yang tulus kepada mereka yang telah turut berperan serta dalam perjalanan ini, baik secara langsung maupun tidak langsung, sehingga laporan Tugas Akhir ini dapat terselesaikan dengan baik. Untuk itu penulis ingin mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Ir. Hari Purnomo M.T selaku Dekan Fakultas Teknologi Industri, Universitas Islam Indonesia.
2. Bapak Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM. selaku Ketua Program Studi Teknik Industri Program Sarjana Universitas Islam Indonesia.
3. Ibu Annisa Uswatun Khasanah, S.T., M.Sc. selaku dosen pembimbing yang telah memberikan arahan, bimbingan, serta pengetahuan yang berharga dalam penyusunan Tugas Akhir ini. Kehadiran dan dedikasi beliau telah menjadi motivasi peneliti dalam menyusun Tugas Akhir ini.
4. Bapak Sudyanto selaku pemilik CV. Intimart sekaligus pemilik toko Intimart Gedongan yang telah memberikan izin serta membantu selama proses penelitian Tugas Akhir berlangsung.



5. Kedua orang tua saya yaitu Bapak Yan Kurniadi dan Ibu Desi Arisanti yang selalu memberikan doa, dorongan, motivasi, dan dukungan baik moral maupun materi sehingga dapat membantu saya selama menyelesaikan Tugas Akhir ini.
6. Sahabat saya yaitu Raihanah Marziqah yang telah hadir dengan memberikan bantuan, nasihat, serta kontribusi berharga dalam mengatasi berbagai kendala dan tantangan selama penyusunan Tugas Akhir.
7. Teman-teman saya yaitu Rizha, Dio, Imam, Oka, Majid, Kevin, Daffa, Tiara, Fairuz, dan Gadis yang telah menemani saya selama saya berkuliah dan juga memberikan bantuan selama penyusunan Tugas Akhir.
8. Teman teman jurusan Program Studi Teknik Industri Universitas Islam Indonesia angkatan 2019 terima kasih telah bekerjasama selama jalannya perkuliahan.
9. Serta semua pihak yang tidak dapat disebutkan satu persatu yang telah memberikan dukungan dan semangat selama penelitian dan penulisan laporan tugas akhir.

Dengan penuh kesadaran, penulis mengakui bahwa tugas akhir ini masih memiliki berbagai kekurangan dan jauh dari kata sempurna. Proses penyusunan tidak luput dari hambatan dan tantangan yang menjadikan perjalanan ini sebagai proses pembelajaran berharga. Oleh karena itu, dengan segala kerendahan hati, penulis menyambut dan menerima dengan tangan terbuka setiap kritik dan saran yang membangun. Melalui masukan berharga tersebut, diharapkan laporan tugas akhir ini dapat diperbaiki dan ditingkatkan kualitasnya.

*Wassalamualaikum Warahmatullahi Wabarakatuh.*

Yogyakarta, 08 Agustus 2023

Penulis



Muhammad Rafly Qowi Baihaqie

## ABSTRAK

Dalam industri ritel yang sangat kompetitif, perusahaan dituntut untuk terus berinovasi dan mencari strategi bisnis yang unik guna meningkatkan performa penjualan mereka. Intimart Gedongan merupakan salah satu toko minimarket yang berada di bawah naungan perusahaan CV Intimart. CV Intimart sendiri telah berdiri sejak tahun 2018 dan mengoperasikan lebih dari 10 cabang atau minimarket yang tersebar di kota Yogyakarta dan Solo. Dalam pelaksanaannya, Intimart Gedongan masih memiliki penjualan yang naik turun selama 1 tahun sehingga sales penjualan tidak memenuhi target secara konsisten. Tujuan dari penelitian ini adalah untuk menganalisis pola perilaku pembelian konsumen di toko Intimart Gedongan guna memberikan rekomendasi strategi yang dapat efektif dalam meningkatkan penjualan. Data yang digunakan pada penelitian ini adalah data transaksi penjualan dari Toko Intimart Gedongan bulan November 2022, Desember 2022, dan Januari 2023. Metode yang digunakan pada penelitian ini adalah *Association Rule – Market Basket Analysis (AR-MBA)* algoritma *FP-Growth* dan *Clustering Analysis* algoritma *K-Means*. Berdasarkan hasil pengolahan data, pada Proses *Clustering* dengan *K-means* peneliti menggunakan *software VSCode* dan memanfaatkan *library scikit-learn* dalam bahasa pemrograman *Python*, didapatkan 4 cluster yang terbentuk yaitu, *Mid-Morning Moderates*, *Diverse Afternoon Buyers*, *Evening Moderates*, *High-Value Customers*. Pada metode *AR- MBA* algoritma *FP-Growth* dengan bantuan *library mlxtend* dalam bahasa pemrograman *Python*, peneliti menggunakan data dari kluster 2, karena kluster ini memiliki nilai transaksi yang berada dalam kisaran cukup tinggi dan pelanggan cenderung membeli berbagai variasi produk dimana kluster ini dapat berpotensi menjadi kluster *High-Value Customers*. Dari penelitian ini dihasilkan aturan asosiasi sebanyak 104 rules. Hasil analisis yang diperoleh dapat dijadikan pertimbangan dalam melakukan strategi pemasaran untuk meningkatkan penjualan diantaranya *product bundling* dan program poin kepada pelanggan.

**Kata kunci:** Pola Perilaku Konsumen, *AR-MBA*, *Clustering*, *K-Means*, *FP-Growth*, Strategi Meningkatkan Penjualan

## DAFTAR ISI

<b>PERNYATAAN KEASLIAN.....</b>	<b>ii</b>
<b>SURAT BUKTI PENELITIAN .....</b>	<b>iii</b>
<b>LEMBAR PENGESAHAN PEMBIMBING .....</b>	<b>iv</b>
<b>LEMBAR PENGESAHAN DOSEN PENGUJI.....</b>	<b>v</b>
<b>HALAMAN PERSEMBAHAN .....</b>	<b>vi</b>
<b>MOTTO .....</b>	<b>vii</b>
<b>KATA PENGANTAR.....</b>	<b>viii</b>
<b>ABSTRAK .....</b>	<b>x</b>
<b>DAFTAR ISI.....</b>	<b>xi</b>
<b>DAFTAR TABEL .....</b>	<b>xiv</b>
<b>DAFTAR GAMBAR.....</b>	<b>xv</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Tujuan Penelitian .....	5
1.4 Batasan Masalah .....	5
1.5 Manfaat Penelitian .....	5
1.6 Sistematika Penulisan .....	6
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>8</b>
2.1 Kajian Literatur .....	8
2.1.1 <i>Clustering</i> .....	8
2.1.2 Association Rule – Market Basket Analysis .....	12
2.1.3 Integrasi <i>Clustering</i> dan Ar .....	15
2.1.4 Posisi Penelitian.....	20
2.2 Landasan Teori .....	20
2.2.1 Ritel .....	20
2.2.2 Data mining .....	22
2.2.3 Association Rule.....	25
2.2.4 Market Basket Analysis.....	26
2.2.5 Clustering.....	27
2.2.6 Algoritma <i>FP-Growth</i> .....	28
2.2.7 Algoritma <i>K-means</i> .....	30
2.2.8 Metode <i>Elbow</i> .....	32
2.2.9 <i>Silhouette Score</i> .....	33

2.2.10 Python .....	34
<b>BAB III METODE PENELITIAN .....</b>	<b>37</b>
3.1 Objek Penelitian .....	37
3.2 Metode Pengumpulan Data .....	37
3.3 Jenis Data .....	37
3.4 Alur Penelitian .....	38
<b>BAB IV PENGUMPULAN DAN PENGOLAHAN DATA .....</b>	<b>45</b>
4.1 Pengumpulan Data .....	45
4.2 Pengolahan Data .....	46
4.2.1 Data Transaksi .....	46
4.2.2 <i>Preprocessing</i> Dataset untuk Melakukan <i>Clustering</i> .....	47
4.2.2.1 Data Cleaning pada <i>Preprocessing</i> Metode <i>Clustering</i> .....	47
4.2.2.2 Data Transformation pada <i>Preprocessing</i> Metode <i>Clustering</i> .....	48
4.2.2.3 Data Selection pada <i>Preprocessing</i> Metode <i>Clustering</i> .....	50
4.2.2.4 Data Normalization pada <i>Preprocessing</i> Metode <i>Clustering</i> .....	52
4.2.3 Proses <i>Clustering</i> dengan K-means .....	52
4.2.3.1 Penentuan jumlah Cluster menggunakan Elbow Method dan Silhouette Score .....	52
4.2.3.2 Proses <i>Clustering</i> menggunakan Algoritma <i>K-Means</i> .....	54
4.2.4 Hasil <i>Clustering</i> K-Means .....	54
4.2.5 <i>Preprocessing</i> Dataset untuk Melakukan <i>AR-MBA</i> .....	58
4.2.5.1 Data Selection pada Metode <i>AR-MBA</i> .....	58
4.2.5.2 Data Cleaning pada Metode <i>AR-MBA</i> .....	59
4.2.5.3 Data Transformation pada Metode <i>AR-MBA</i> .....	60
4.2.6 Proses <i>AR-MBA</i> dengan <i>FP-Growth</i> .....	61
<b>BAB V ANALISIS DAN PEMBAHASAN .....</b>	<b>67</b>
5.1 Analisis <i>Clustering</i> .....	67
5.1.1 Analisis Deskripsi Data .....	67
5.1.2 Proporsi Anggota Kluster .....	68
5.1.3 Analisis Karakteristik Kluster .....	69
5.2 Analisis Association Rule Market Basket Analysis .....	72
5.2.1 Analisis Perbandingan <i>Association Rule</i> .....	72
5.2.2 Analisis Frekuensi Pembelian .....	73
5.2.3 <i>Output Association Rule</i> .....	75
5.3 Rekomendasi Hasil Penelitian .....	78
<b>BAB VI KESIMPULAN DAN SARAN .....</b>	<b>82</b>
6.1 Kesimpulan .....	82

6.2 Saran .....	87
<b>DAFTAR PUSTAKA.....</b>	<b>88</b>
<b>LAMPIRAN.....</b>	<b>94</b>

## DAFTAR TABEL

Tabel 2. 1 Kajian Induktif <i>Clustering</i> .....	10
Tabel 2. 2 Kajian Induktif <i>AR-MBA</i> .....	14
Tabel 2. 3 Kajian Induktif Integrasi <i>Fp-Growth</i> dan <i>K-Means</i> .....	19
Tabel 4. 1 Nama Kolom dan Penjelasannya .....	45
Tabel 4. 2 Data Transaksi Toko Intimart Gedongan.....	46
Tabel 4. 3 Tampilan Data yang Memiliki <i>Missing Value</i> .....	47
Tabel 4. 4 Data Transformasi Kode Faktur .....	48
Tabel 4. 5 Data Transformasi Nama Barang dan Jam Transaksi.....	49
Tabel 4. 6 Deskripsi Variabel <i>Clustering</i> .....	50
Tabel 4. 7 Tampilan Data <i>Clustering</i> .....	51
Tabel 4. 8 Tampilan Data Setelah dilakukan Normalisasi.....	52
Tabel 4. 9 Tampilan Hasil <i>Cluster</i> pada Data.....	54
Tabel 4. 10 Pusat <i>Cluster</i> dalam <i>Z-Score</i> .....	54
Tabel 4. 11 Deskripsi Statistik dari Tiap <i>Cluster</i> .....	55
Tabel 4. 12 Rentang Persebaran Variabel dari Tiap <i>Cluster</i> .....	56
Tabel 4. 13 Tabel <i>Data Selection</i> pada <i>AR-MBA</i> .....	58
Tabel 4. 14 Tampilan Data Memiliki 1 Barang .....	59
Tabel 4. 15 Sebelum dan Sesudah Pembersihan Data .....	60
Tabel 4. 16 Hasil Transformasi Data Menjadi <i>Boolean</i> .....	60
Tabel 4. 17 Perbandingan Nilai <i>Minimum Support</i> .....	61
Tabel 4. 18 Deskripsi Aturan Asosiasi Tiap <i>Cluster</i> .....	62
Tabel 4. 19 Aturan Asosiasi pada <i>Cluster 1</i> .....	63
Tabel 4. 20 Aturan Asosiasi pada <i>Cluster 2</i> .....	64
Tabel 4. 21 Aturan Asosiasi pada <i>Cluster 3</i> .....	65
Tabel 4. 22 Aturan Asosiasi pada <i>Cluster 4</i> .....	65
Tabel 5. 1 Deskripsi Data Metode <i>Clustering</i> .....	67
Tabel 5. 2 Proporsi Jumlah <i>Cluster</i> .....	68
Tabel 5. 3 Tabel Rata-Rata Pusat <i>Cluster</i> .....	69
Tabel 5. 4 Tabel Persebaran Data Tiap <i>Cluster</i> .....	69
Tabel 5. 5 Perbandingan <i>Minimum Support</i> .....	73
Tabel 5. 6 Frekuensi Pembelian Kluster 2 .....	74
Tabel 5. 7 Aturan Asosiasi dengan <i>Support</i> Tertinggi.....	76

**DAFTAR GAMBAR**

Gambar 1. 1 Data Sales Toko Intimart Gedongan Tahun 2022.....	2
Gambar 3. 1 Alur Penelitian .....	39
Gambar 4. 1 Jumlah <i>Missing Value</i> pada Data.....	47
Gambar 4. 2 Grafik <i>Elbow Method</i> .....	53
Gambar 4. 3 Grafik <i>Silhouette Score</i> .....	53
Gambar 5. 1 Frekuensi Kemunculan Produk pada Kluster2.....	75

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Dalam industri ritel yang sangat kompetitif ini, perusahaan dituntut untuk terus berinovasi dan mencari strategi bisnis yang unik guna meningkatkan performa penjualan mereka. Hal ini menjadi suatu keharusan agar perusahaan dapat mempertahankan keunggulan kompetitif di pasar yang terus berubah (Achmad et al., 2023). Persaingan dalam industri ritel dapat dibuktikan dengan banyaknya toko ritel yang ada di Indonesia. Menurut data dari Euromonitor, total gerai ritel di Indonesia mencapai 3,61 juta pada tahun 2021 (Rizaty, 2022). Pertumbuhan ritel di Indonesia dapat dilihat dari total belanja yang dikeluarkan oleh masyarakat. Pada tahun 2020 sendiri, orang Indonesia membelanjakan 19% dari total pengeluaran rumah tangga mereka untuk membeli produk FMCG (Nuraini, 2022). Namun, dalam konteks pertumbuhan ritel yang semakin positif tersebut, banyak toko ritel yang mengalami kesulitan dan bahkan mengalami kegagalan. Menurut Budihardjo Iduansjah, Ketua Umum Himpunan Peritel dan Penyewa Pusat Perbelanjaan Indonesia (Hippindo), salah satu faktor penyebabnya adalah perubahan perilaku konsumen dalam berbelanja. Survei menunjukkan bahwa masyarakat lebih memilih untuk berbelanja di toko ritel kecil untuk menghemat waktu (Fitriyani, 2023). Menurut penelitian yang dibuat oleh (Nisa, 2020) yang meneliti terkait perilaku konsumen dalam memilih tempat belanja antara minimarket dan toko kelontong. Masyarakat lebih memilih untuk berbelanja produk dan barang di minimarket karena mereka percaya bahwa kualitasnya lebih terjamin dan penataan barang di minimarket lebih baik. Selain itu, minimarket juga seringkali menawarkan berbagai promosi menarik kepada konsumen. Meskipun demikian, konsumen juga tetap memilih untuk berbelanja di toko kelontong, terutama untuk keperluan dapur, karena di sana harga barang lebih murah.

Industri ritel yang sangat kompetitif mendorong pemilik perusahaan toko minimarket Intimart Gedongan untuk terus melakukan perubahan demi menghadapi tantangan yang ada. Intimart Gedongan merupakan salah satu toko minimarket yang berada di bawah naungan perusahaan CV Intimart. CV Intimart sendiri telah berdiri sejak tahun 2018 dan mengoperasikan lebih dari 10 cabang atau minimarket yang tersebar di kota Yogyakarta dan Solo. Meskipun telah mengalami perkembangan yang signifikan, Toko Intimart Gedongan tidak terlepas dari berbagai permasalahan yang dihadapi. Ketika melakukan wawancara dengan pemilik toko atau minimarket Intimart Gedongan. Pemilik mengaku bahwa sales hanya



mencapai target ketika bulan-bulan tertentu seperti bulan Idul Fitri atau Ramadan. Namun, pada bulan-bulan lainnya, penjualan masih belum memenuhi harapan. Oleh karena itu, Intimart Gedongan perlu mencari strategi promosi yang lebih tepat sasaran untuk menarik minat pelanggan dan mendorong mereka untuk membeli produk. Berikut merupakan data penjualan toko Intimart Gedongan pada bulan Januari tahun 2022 sampai Desember tahun 2022:

Gambar 1. 1 Data Sales Toko Intimart Gedongan Tahun 2022

Bulan	Tahun	Sales	Target Penjualan	Presentase Target
Januari	2022	265,000,000	300,000,000	88.33%
Februari	2022	246,660,000	300,000,000	82.22%
Maret	2022	350,010,000	300,000,000	116.67%
April	2022	375,000,000	300,000,000	125.00%
Mei	2022	366,660,000	300,000,000	122.22%
Juni	2022	266,670,000	300,000,000	88.89%
Juli	2022	241,680,000	300,000,000	80.56%
Agustus	2022	291,660,000	300,000,000	97.22%
September	2022	266,670,000	300,000,000	88.89%
Oktober	2022	241,680,000	300,000,000	80.56%
November	2022	275,000,000	300,000,000	91.67%
Desember	2022	294,000,000	300,000,000	98.00%

Berdasarkan data yang tertera pada tabel di atas, dapat diketahui bahwa toko Intimart Gedongan mengalami fluktuasi penjualan selama satu tahun. Dari data di atas toko Intimart Gedongan memiliki penjualan yang naik turun selama 1 tahun. Toko tersebut baru mencapai target penjualan ketika berada pada bulan Maret, April, Mei, dan hampir menyentuh pada bulan Desember. Faktor-faktor musiman seperti menjelang bulan Idul Fitri atau Ramadhan dan menjelang akhir tahun memainkan peran penting dalam mencapai target tersebut.

Dari data di atas maka didapatkan permasalahan pada sales penjualan tidak memenuhi target secara konsisten. Oleh karena itu, perlu adanya perbaikan masalah guna meningkatkan serta memenuhi target penjualan. Penjualan yang tidak memenuhi target penjualan pada minimarket Intimart Gedongan merupakan permasalahan yang saat ini dihadapi oleh minimarket ini. Salah satu faktor penyebab tidak tercapainya target penjualan adalah kurangnya efektivitas dan ketepatan sasaran dalam melakukan promosi. Dalam rangka meningkatkan penjualan dan memperkuat posisi toko di pasar, strategi pemasaran yang lebih tepat sasaran dan efektif perlu dikembangkan. Salah satu strategi yang dapat dilakukan adalah dengan melakukan segmentasi pelanggan dan analisis perilaku pelanggan menggunakan data internal perusahaan. Toko Intimart Gedongan memiliki banyak data yang tersimpan dalam database perusahaannya. Dari

banyaknya data yang dimiliki oleh toko Intimart Gedongan, salah satu data yang belum dimanfaatkan secara maksimal adalah data transaksi. Padahal data ini bisa memberikan informasi tambahan tentang perilaku pelanggan, tren penjualan, preferensi produk, dan kebutuhan bisnis. Penting bagi perusahaan untuk memanfaatkan data transaksi guna meningkatkan efektivitas dan keuntungan bisnis. Untuk menciptakan strategi penjualan yang efektif, teknologi informasi dapat ditingkatkan penggunaannya dengan memanfaatkan data transaksi yang sudah ada (Takdirillah, 2020).

Dengan melakukan analisis segmentasi menggunakan data transaksi penjualan, kita dapat memperoleh pemahaman tentang karakteristik pelanggan yang berbeda dalam beberapa kelompok. Pelanggan akan dikelompokkan berdasarkan perilaku mereka saat melakukan pembelian sehingga dapat mempermudah dalam menjalankan strategi pemasaran yang tepat sasaran (Auliasari & Kertaningtyas, 2019). Salah satu metode yang dapat digunakan untuk mengelompokkan konsumen atau segmentasi adalah metode *clustering*. *Clustering* adalah metode untuk mengelompokkan data berdasarkan kesamaan karakteristik antara satu data dengan data yang lain (Shaliha et al., 2021). Salah satu algoritma pada metode *Clustering* akan menggunakan algoritma *k-means*. Algoritma ini bekerja dengan tahapan yaitu menentukan jumlah *cluster* yang diinginkan, melakukan alokasi data ke dalam jumlah *cluster* yang telah ditentukan, menentukan nilai centroid pada tiap-tiap *cluster*, menghitung jarak terdekat antara setiap titik data dengan menggunakan rumus Euclidean, dan menampilkan hasil perhitungan berdasarkan jarak (Sani, 2018). Algoritma *K-means* memiliki kelebihan berupa kemampuan mengelompokkan data besar dengan sangat cepat. Ini dibuktikan pada penelitian yang dilakukan oleh Kamila et al. (2019) yang berjudul “*Perbandingan Algoritma K-means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau*” saat melakukan pengolahan data, algoritma *K-means* hanya membutuhkan waktu 1 detik sedangkan ketika menggunakan algoritma *k-medoids* membutuhkan waktu lebih dari 1 menit. Hal ini menunjukkan bahwa algoritma *K-means* memiliki keunggulan dalam efisiensi waktu saat melakukan iterasi dengan jumlah data yang besar dibandingkan dengan metode *k-medoids*. Dengan demikian, penggunaan algoritma *K-means* dapat memberikan keuntungan yang signifikan dalam pengelompokan data yang lebih cepat dan efisien.

Selain menggunakan metode *Clustering*, metode lain yang digunakan pada penelitian ini adalah metode *Market Basket Analysis* (MBA). *Market Basket Analysis* (MBA) adalah teknik analisis data yang digunakan untuk menemukan pola pembelian produk menggunakan data transaksi penjualan. Tujuan utama MBA dalam pemasaran adalah memberikan informasi kepada pemilik toko untuk memahami perilaku pembelian pembeli, yang dapat membantu

dalam pengambilan keputusan yang tepat (Kaur & Kang, 2016). Salah satu algoritma yang dapat digunakan dalam metode Market Basket Analysis (MBA) adalah algoritma *FP-Growth*. *FP-Growth* adalah sebuah algoritma yang digunakan untuk mengidentifikasi himpunan data yang sering muncul (frequent itemset) dalam kumpulan data tanpa menggunakan kandidat generation (Bunda, 2020). Algoritma *FP-Growth* menggunakan struktur data tree dalam proses pencarian frequent itemset, sehingga mampu melakukan pencarian dengan lebih efisien dan cepat dibandingkan dengan metode apriori (Setiawan & Gita Anugrah, 2019). Dengan menggunakan pendekatan ini, algoritma *FP-Growth* dapat menghindari proses yang disebut candidate generation, yang dapat memakan waktu dan menghabiskan sumber daya komputasi yang besar. Sebagai hasilnya, algoritma *FP-Growth* menjadi salah satu pilihan populer untuk analisis pola asosiasi dalam big data. Hal ini dibuktikan pada penelitian yang dilakukan oleh (Erwin, 2009) berjudul “*Analisis Market Basket Dengan Algoritma Apriori dan FP-Growth*” algoritma apriori memiliki kekurangan dalam hal waktu pengolahan data dan alokasi memori yang besar. Ini disebabkan karena pemindai data yang dilakukan berulang kali. Disisi lain algoritma *FP-Growth* menggunakan struktur data *FP-Tree* yang dapat mengompres data transaksi dengan item yang sama, sehingga penggunaan memori yang lebih efisien dan proses pencarian frequent itemset menjadi lebih cepat.

Dengan melihat permasalahan yang ada pada toko Intimart Gedongan, penelitian ini dilakukan untuk memberikan usulan strategi pemasaran yang berguna agar toko tersebut dapat mencapai target penjualan. Usulan strategi tersebut dapat diperoleh dengan menggunakan metode *Association Rule - Market Basket Analysis (AR-MBA)* dan metode *Clustering*. Metode *Association Rule - Market Basket Analysis* akan mendukung toko dalam mengidentifikasi pola pembelian antar item dalam transaksi pelanggan, sehingga dapat memberikan rekomendasi penjualan yang relevan. Sementara itu, metode *Clustering* dapat memberikan pengelompokan/segmentasi pelanggan berdasarkan data transaksi dengan membantu dalam mengklasifikasikan pola distribusi secara keseluruhan dan menemukan keterkaitan menarik antara atribut data. Dengan menerapkan kedua pendekatan ini, Intimart Gedongan dapat mengembangkan strategi pemasaran yang lebih spesifik dan meningkatkan personalisasi dalam rekomendasi produk kepada pelanggan.

## **1.2 Rumusan Masalah**

Berikut merupakan rumusan masalah pada penelitian ini:

1. Apa saja hasil perbandingan *cluster* pada minimarket Intimart Gedongan?

2. Bagaimana pola perilaku pembelian konsumen pada toko Intimart Gedongan berdasarkan metode Association Rules – Market Basket Analysis?
3. Bagaimana rekomendasi yang dapat diberikan kepada toko Intimart Gedongan menggunakan metode *clustering* dan *AR-MBA*?

### 1.3 Tujuan Penelitian

Tujuan dari penelitian yang dilakukan adalah sebagai berikut.

1. Mengetahui pola perilaku pembelian konsumen pada toko Intimart Gedongan berdasarkan metode *Association Rules – Market Basket Analysis*.
2. Mengetahui hasil *clustering* dan analisisnya berdasarkan data transaksi yang ada pada toko Intimart Gedongan.
3. Menghasilkan usulan rekomendasi strategi peningkatan penjualan berdasarkan hasil analisis metode *Association Rules – Market Basket Analysis* dan *Clustering*.

### 1.4 Batasan Masalah

Pada Penelitian ini diperlukan adanya batasan dalam melakukan penelitian, berikut merupakan batasan masalah dari penelitian ini.

1. Penelitian ini dilakukan pada toko Intimart Gedongan yang berlokasi di Jl. Selokan Mataram, Gedongan, Sinduadi, Kec. Mlati, Kabupaten Sleman, Daerah Istimewa Yogyakarta.
2. Objek penelitian ini adalah transaksi belanja yang dilakukan di toko Intimart Gedongan.
3. Data yang akan dianalisis merupakan data transaksi penjualan pada toko Intimart Gedongan dari tanggal 1 November 2022 sampai 31 Januari 2023
4. Metode yang digunakan dalam penelitian ini adalah metode Association Rule - Market Basket (*AR-MBA*) dan *Clustering*.
5. Hasil analisis yang akan diberikan adalah usulan strategi pemasaran guna meningkatkan dan memenuhi target penjualan.
6. Hasil yang diberikan dalam penelitian ini dapat menjadi saran dan rekomendasi untuk toko Intimart Gedongan.

### 1.5 Manfaat Penelitian

Adapun manfaat dari penelitian yang dilakukan adalah sebagai berikut:

1. Bagi Perusahaan

Penelitian ini memberikan solusi bagi perusahaan untuk mengatasi permasalahan yang dihadapi dalam meningkatkan penjualan dan mencapai target penjualan. Penggunaan metode Association Rules - Market Basket Analysis dan *Clustering* dapat membantu perusahaan dalam mengidentifikasi pola pembelian, memberikan rekomendasi penjualan yang relevan, dan mengklasifikasikan pola distribusi secara keseluruhan. Dengan menerapkan metode ini, perusahaan dapat mengembangkan strategi pemasaran yang lebih efektif dan meningkatkan kepuasan pelanggan.

## 2. Bagi Peneliti

Penelitian ini berkontribusi dalam memperkaya literatur akademis dan praktis mengenai penggunaan metode *Association Rules – Market Basket Analysis* dan *Clustering* dalam konteks industri ritel. Hasil penelitian ini dapat menjadi referensi bagi pembaca atau peneliti yang tertarik dengan topik ini, sehingga dapat membantu mereka dalam mengembangkan solusi-solusi menggunakan metode ini.

### **1.6 Sistematika Penulisan**

Adanya sistematika penulisan adalah sebagai acuan dalam pembuatan laporan penelitian ini, berikut merupakan sistematika penulisan:

#### **BAB I: PENDAHULUAN**

Pada bab ini memuat latar belakang, rumusan masalah dan pertanyaan penelitian, tujuan penelitian, manfaat penelitian, batasan permasalahan, dan sistematika penulisan laporan TA.

#### **BAB II: KAJIAN LITERATUR**

Pada bab ini memuat kajian literatur atau landasan teori yang menjelaskan terkait teori-teori yang relevan dalam penelitian ini, dengan menggunakan referensi dari jurnal dan buku terkait. Tujuan dari bab ini adalah untuk mendapatkan pemahaman yang mendalam tentang konsep-konsep yang akan digunakan dalam penelitian dan mengacu pada sumber-sumber yang dapat dipercaya.

#### **BAB III: METODE PENELITIAN**

Pada bab ini memuat obyek penelitian, data yang digunakan dan tahapan yang telah dilakukan dalam penelitian secara ringkas dan jelas. Metode ini dapat meliputi metode pengumpulan data, alat bantu analisis data, pembangunan model, desain dan prototyping. Metode penelitian

menginformasikan What, How, dan Why. Jika penelitian menggunakan rumus tertentu maka rumus dan teori yang digunakan dimasukkan ke bagian metode penelitian.

#### **BAB IV: PENGUMPULAN DAN PENGOLAHAN DATA**

Pada bab ini memuat proses pengolahan data dengan prosedur tertentu, termasuk gambar dan grafik yang diperoleh dari hasil penelitian

#### **BAB V: HASIL DAN PEMBAHASAN**

Bab ini mencakup proses pengumpulan data observasi atau pengukuran secara sistematis untuk penelitian. Data yang terkumpul kemudian akan diolah sesuai dengan metode yang relevan untuk menganalisis masalah yang sedang diteliti.

#### **BAB VI: KESIMPULAN DAN SARAN**

Bab ini merupakan penutup dari penelitian yang telah dilakukan, di mana bab-bab sebelumnya telah menjelaskan isi penelitian secara detail. Pada bagian kesimpulan, disajikan secara singkat hasil penelitian yang telah dilaksanakan dan diberikan saran-saran yang dapat digunakan sebagai referensi.

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Kajian Literatur

Dalam kajian induktif, penelitian sebelumnya digunakan sebagai pembanding dan sebagai landasan bagi penelitian yang akan dilakukan. Dengan mengacu pada penelitian sebelumnya, penelitian baru ini berusaha untuk memperluas pemahaman dan memberikan kontribusi baru dalam bidang yang sama.

##### 2.1.1 Clustering

Pada penelitian yang dilakukan oleh Auliasari & Kertaningtyas (2019) yang berjudul “Penerapan Algoritma *K-means* untuk Segmentasi Konsumen Menggunakan R”. Dalam penelitian ini, metode *K-means* digunakan untuk membantu perusahaan memahami hubungan antara kecenderungan pembelian produk oleh konsumen yang sebelumnya mungkin tidak diperhatikan. Penelitian ini bertujuan untuk menargetkan konsumen yang memberikan keuntungan dengan menggunakan analisis kluster dan algoritma *K-means* yang diimplementasikan menggunakan bahasa pemrograman R. Hasil pengelompokan dataset konsumen PT. Super Sukses Niaga pada tahun 2018 menunjukkan bahwa konsumen yang memiliki potensi paling tinggi terdapat dalam kluster 2, dengan rata-rata total belanja setahun sebesar 4.198.6656. Kluster 2 didominasi oleh jenis usaha PT, dengan nilai rata-rata jenis usaha mendekati 2 (label untuk jenis usaha PT). Hal serupa terlihat pada nilai rata-rata lokasi yang mendekati 2 (label untuk lokasi di pulau Kalimantan) pada kluster 2. Oleh karena itu, dapat disimpulkan bahwa konsumen yang memiliki tingkat transaksi tinggi pada kluster 2 adalah PT. BIS INDUSTRI, PT. REA KALTIM, dan CV. MEDCO MINING.

Pada Penelitian berjudul “Penerapan Algoritma Association Rules Dalam Penentuan Pola Pembelian Berdasarkan Hasil Clustering” yang disusun oleh (Octavia et al., 2023). Penelitian ini dilakukan pada Toko Buku Zanafa menggunakan data transaksi bulan Januari hingga Desember 2022. Penelitian ini bertujuan menggunakan *FP-Growth* dan *K-Medoids* untuk memberikan informasi terkait penempatan tata letak buku yang tepat agar strategi penjualan lebih efektif. Hasil dari penelitian ini menunjukkan terbentuk 6 kluster dengan nilai DBI sebesar 0.639 dan 7 kluster pada data bulan juli dengan nilai DBI sebesar 0.349. pada metode *AR-MBA* menggunakan *rules 30% support* dan *70% confidence*.

Pada penelitian berjudul “*The application of data mining and RFM model in market segmentation of a veterinary hospital*” oleh Wei et al. (2019). Penelitian ini mengadopsi teknik pengelompokan dua tahap, kombinasi peta pengorganisasian mandiri dan metode *K-means*, dan model RFM (recency, frequency, dan monetary) untuk mengkategorikan pelanggan di rumah sakit hewan di Kota Taichung, Taiwan, untuk membuat strategi pemasaran yang efektif di pasar yang kompetitif ini. Berdasarkan 1.784 pelanggan yang hanya berfokus pada kucing pada tahun 2014, dua belas klaster terbentuk. Secara khusus, enam dari dua belas klaster dapat diklasifikasikan sebagai pelanggan terbaik dan paling setia. Tiga klaster tidak pasti dan kehilangan pelanggan. Satu klaster dianggap sebagai yang terbaik tetapi telah kehilangan pelanggan. Terakhir, berdasarkan model RFM dan berdasarkan model RFM, dua klaster, yang memiliki nilai recency yang relatif lebih tinggi daripada nilai rata-rata, termasuk dalam pelanggan yang tidak pasti tetapi baru.

Pada Penelitian berjudul “*RFM Analysis Using K-means Clustering to Improve Revenue and Customer Retention*” yang ditulis oleh Dawane et al. (2021). Dalam penelitian ini, peneliti melakukan segmentasi pelanggan menggunakan metode RFM Analysis, yang merupakan salah satu metode kunci dalam analisis pemasaran, untuk membagi pelanggan berdasarkan kriteria perilaku yang mereka tunjukkan sebelumnya. RFM merupakan singkatan dari *Recency*, *Frequency*, dan *Monetary*. Dengan semakin banyaknya data transaksi masa lalu yang tersedia, analisis RFM dapat digunakan secara efektif untuk membagi pelanggan dan mengambil tindakan bisnis selanjutnya. Penelitian ini melakukan segmentasi pelanggan berdasarkan data transaksi masa lalu menggunakan algoritma *clustering K-means* dalam bahasa pemrograman Python, dan berdasarkan segmen yang terbentuk, saran tindakan bisnis yang direkomendasikan disampaikan.

Pada penelitian berjudul “PERANCANGAN SEGMENTASI PELANGGAN DENGAN METODE *CLUSTERING K-MEANS* DAN MODEL RFM PADA KLINIK KECANTIKAN SEOUL SECRET” yang ditulis oleh Setiawan et al. (2021). Dalam penelitian ini, peneliti mempelajari segmentasi pelanggan Seoul Secret, sebuah klinik kecantikan yang terletak di Jakarta Selatan. Seoul Secret berkomitmen untuk memberikan pelayanan dan perawatan yang optimal guna memenuhi kebutuhan pelanggan. Pada awal tahun 2018, Seoul Secret memperkenalkan program paket *treatment*. Namun, pendapatan dari pengguna paket mengalami fluktuasi pada tahun 2019 dan 2020, bahkan tidak mencapai target yang ditetapkan. Meskipun jumlah pelanggan meningkat, tidak sebanding dengan jumlah pelanggan yang tetap setia. Selama tiga tahun terakhir, jumlah pelanggan yang hilang lebih tinggi daripada jumlah pelanggan yang tetap setia. Oleh karena itu, pihak Seoul Secret menyadari pentingnya



mengelola data pelanggan guna memberikan informasi yang lebih baik tentang para pelanggan. Tujuan penelitian ini adalah melakukan segmentasi terhadap pelanggan yang menggunakan paket *treatment* dan memberikan usulan strategi pemasaran yang tepat. Penelitian ini menggunakan metode *Clustering K-means* dan menerapkan model RFM. Metode *Clustering K-means* digunakan untuk mengelompokkan pelanggan berdasarkan karakteristik yang serupa ke dalam sejumlah K atau *cluster* yang telah ditentukan. Sementara itu, model RFM digunakan untuk mengelompokkan data berdasarkan atribut *Recency*, *Frequency*, dan *Monetary*. Dengan menggabungkan model RFM dan metode *Clustering K-means*, penelitian ini dapat menggambarkan segmentasi pelanggan berdasarkan perilaku mereka. Selain itu, penelitian ini juga menghasilkan strategi pemasaran yang dapat diterapkan oleh Seoul Secret, antara lain memberikan *cashback*, memberikan *voucher*, menyediakan program keanggotaan, menerapkan sistem poin, menawarkan paket *treatment* dengan layanan pijat gratis, memberikan bonus *treatment*, menggunakan sistem *appointment*, menambah jumlah dokter kecantikan, dan menyediakan layanan prioritas.

Pada penelitian berjudul “Sistem Rekomendasi Penawaran Produk Pada Online Shop Menggunakan K-Means Clustering” yang ditulis oleh Naufal et al. (2022). Dalam penelitian ini, peneliti akan membuat sistem rekomendasi penawaran produk pada *online shop* menggunakan algoritma K-Means. Dalam penelitian ini, data transaksi selama satu tahun dijadikan dasar untuk mengelompokkan pelanggan ke dalam tiga kluster yang berbeda. Hasil klustering tersebut mengungkap karakteristik masing-masing kluster penjualan, di mana *cluster 1* menunjukkan tingkat penjualan yang sedang dengan rentang usia pembeli antara 36 hingga 50 tahun, *cluster 2* menunjukkan penjualan paling tinggi dengan rentang usia pembeli antara 18 hingga 26 tahun, dan *cluster 3* menunjukkan penjualan yang lebih rendah dengan rentang usia pembeli antara 27 hingga 35 tahun. Berdasarkan temuan ini, peneliti merekomendasikan untuk menawarkan produk terpopuler dari setiap kluster tersebut kepada pelanggan guna meningkatkan efisiensi dan efektivitas rekomendasi produk pada *platform online shop*.

Berikut merupakan tabel penelitian terdahulu mengenai metode *Clustering* yang berisikan penulis, objek, metode, dan hasil dari penelitian terdahulu:

Tabel 2. 1 Kajian Induktif *Clustering*

No	Penulis	Objek	Variabel	Metode	Hasil
----	---------	-------	----------	--------	-------

1	(Auliasari & Kertaningtyas, 2019)	Perusahaan X	Customer ID, Nama Pelanggan, Jenis, Lokasi, Total Belanja Setahun.	<i>Clustering</i> menggunakan algoritma <i>K-Means</i>	Fokus layanan dan pemasaran pada konsumen <i>cluster</i> ke-2.
2	(Octavia et al., 2023)	Toko Buku Zanafa	Item dibeli dan Item tidak dibeli	<i>Clustering</i> menggunakan algoritma <i>K-Medoids</i>	Terbentuk 6 kluster pada data keseluruhan dengan nilai DBI 0.639.
3	(Wei et al., 2019)	Rumah sakit hewan di Kota Taichung	<i>Recency</i> (Jumlah hari terakhir transaksi), <i>Frequency</i> (Jumlah pelanggan berkunjung), <i>Monetary</i> (Jumlah uang yang di tranksaksikan)	<i>Clustering</i> menggunakan algoritma <i>K-Means</i>	Pelanggan dengan simbol <i>RFM</i> dianggap setia, <i>FM</i> adalah pelanggan yang hilang, <i>R</i> adalah pelanggan baru, dan <i>none</i> adalah pelanggan yang hilang.
4	(Dawane et al., 2021)	Perusahaan ritel produk FMCG	<i>Gender</i> , <i>Recency</i> (Jumlah hari terakhir transaksi), <i>Frequency</i> (Jumlah transaksi yang dilakukan), <i>Monetary</i> (Jumlah uang yang dari transaksi)	<i>Clustering</i> menggunakan algoritma <i>K-Means</i>	<i>Cluster</i> 2 memiliki pelanggan terbaik untuk bisnis dengan nilai moneter tinggi, frekuensi pemesanan tinggi, dan rentang waktu transaksi terakhir rendah.
5	(Setiawan et al., 2021)	Klinik Kecantikan Seoul Secret	Kode Pelanggan, Nama Pelanggan,	<i>Clustering</i> menggunakan algoritma <i>K-Means</i>	Segmentasi pelanggan klinik kecantikan Seoul Secret menggunakan <i>RFM</i> dan metode <i>K-means</i> menghasilkan 4 <i>cluster</i> .

---

			Jumlah Pemasukan, Tanggal Transaksi		
6	(Naufal et al., 2022)	<i>Online Shop</i>	Tanggal, Bulan, <i>Customer</i> , Umur, Merek, Type, Harga, Beli, Margin	<i>Clustering</i> menggunakan algoritma <i>K-</i> <i>Means</i>	Nilai <i>sillhouette coeficient</i> didapatkan <i>cluster</i> dengan jumlah 3 karena memiliki nilai paling mendekati $S_i = 1$ .

---

### 2.1.2 Association Rule – Market Basket Analysis

Pada penelitian yang dilakukan oleh Bunda (2020) yang berjudul “ALGORITMA *FP-GROWTH* UNTUK MENGANALISA POLA PEMBELIAN OLEH-OLEH (STUDI KASUS DI PUSAT OLEH-OLEH UMMI AUFA HAKIM)”. Penelitian ini bertujuan untuk meningkatkan penjualan di Pusat Oleh-Oleh Ummi Aufa Hakim. Data penjualan akan diolah menggunakan metode *Association Rule* dengan menggunakan Algoritma *FP-Growth*. Hasil tersebut diuji menggunakan aplikasi yang dikembangkan dengan menggunakan bahasa pemrograman PHP dan MySQL. Hasil pengujian menunjukkan adanya asosiasi produk yang dibeli bersamaan, diantaranya adalah pembelian produk Cincang Kuning 250gr dan Sj. Balado Merah 250gr dengan nilai *support* sebesar 20% dan *confidence* sebesar 100%, pembelian Ganepo 250gr dan Sj. Balado Merah dengan nilai *support* 20% dan *confidence* 100%, dan pembelian Krk. Kaling Kecil 250gr dan Sj. Balado Merah 250gr dengan nilai *support* 20% dan *confidence* 100%. Dari hasil pengujian tersebut, dapat disimpulkan bahwa penerapan data mining dengan menggunakan algoritma *FP-Growth* dapat digunakan untuk menganalisa pola belanja konsumen dan menjadi rekomendasi dalam tata letak produk di rak persiapan. Dengan cara ini, produk yang sering dibeli dalam waktu yang bersamaan dapat diletakkan berdekatan di dalam toko.

Pada penelitian berjudul “*Marketing strategy for the determination of staple consumer products using FP-growth and apriori algorithm*” yang disusun oleh Ariestya et al. (2019). Tujuan penelitian ini adalah untuk mengidentifikasi strategi pemasaran yang tepat guna meningkatkan penjualan produk. Strategi pemasaran ditentukan melalui analisis permintaan konsumen terhadap kebutuhan dasar. Algoritma yang digunakan dalam penelitian ini adalah *FP-Growth* dan *Apriori*. Hasil dari penelitian tersebut adalah penggunaan nilai kepercayaan minimum sebesar 0.6 dan dukungan minimum sebesar 0.45. Algoritma *FP-Growth* menghasilkan 5 aturan, sementara Algoritma *Apriori* menghasilkan 3 aturan. Dari hasil

penelitian, dapat disimpulkan bahwa Algoritma FP-Growth mengambil keputusan aturan yang lebih baik dibandingkan Algoritma Apriori dalam menentukan strategi pemasaran, karena mampu menghasilkan lebih banyak keputusan terkait cara produk dijual.

Pada penelitian berjudul “ANALISIS DAN IMPLEMENTASI ALGORITMA *FP-GROWTH* PADA APLIKASI SMART UNTUK MENENTUKAN *MARKET BASKET ANALYSIS* PADA USAHA RETAIL (STUDI KASUS : PT.X)” oleh Larasati et al. (2015). Penelitian ini membahas tentang cara kerja sistem market basket analysis dengan menggunakan algoritma *FP-Growth*, perancangan sistem dengan menggunakan algoritma *FP-Growth*, dan penentuan frequent item set dan rule. Hasil pengujian menunjukkan bahwa dengan menggunakan nilai *minimum support* sebesar 0.006 dan *minimum confidence* sebesar 0.6, penelitian ini tidak menghasilkan rule. Meskipun demikian, aplikasi SMART memiliki potensi untuk membantu peritel dalam menganalisis pola pembelian pelanggan dan mengoptimalkan penjualan produk yang sesuai dengan kebutuhan pelanggan.

Pada penelitian berjudul “Perbandingan Algoritma *Apriori* Dan *Fp Growth* Terhadap Market Basket Analysis Pada Data Penjualan *Bakery*” oleh Fathurrahman et al. (2023). Penelitian ini menemukan bahwa kompetisi yang semakin sengit di industri roti telah mendorong para pelaku bisnis untuk mengembangkan strategi yang lebih baik. Salah satu masalah yang dihadapi adalah sulitnya memahami pola pembelian konsumen karena keterbatasan analisis data transaksional. Oleh karena itu, metode market basket analysis digunakan untuk mengungkap pola pembelian konsumen. Hasil dari penelitian ini menunjukkan bahwa dengan menggunakan metode asosiasi dengan algoritma Apriori, pola kombinasi dengan *confidence* tertinggi adalah Alfajores dibeli juga dengan Kopi dengan *confidence* 54,06%, dan ketika membeli Cake juga membeli Kopi dengan *confidence* 52,69%. Sedangkan dengan algoritma FP growth ditemukan bahwa ketika Anda membeli Pastry Anda juga membeli Kopi dengan *confidence* 55.21% dan ketika Anda membeli Biskuit Anda juga membeli Kopi dengan *confidence* 51.84%.

Pada penelitian berjudul “*Market basket analysis with association rules*” yang ditulis oleh Ünvan (2021). Penelitian ini dilakukan untuk melakukan analisis keranjang pasar dengan menggunakan aturan asosiasi. Data yang digunakan dalam penelitian ini adalah data penjualan setiap supermarket yang diperoleh dari situs web Vancouver Island University. Data tersebut dianalisis dalam program Weka dengan menggunakan dataset yang berisi 225 produk yang berbeda. Apriori dan FP Growth, yang merupakan algoritma aturan asosiasi, dicoba secara berurutan. Karena kumpulan data bersifat kategorikal, algoritma Apriori tidak memberikan hasil apa pun. Oleh karena itu, algoritma FP Growth digunakan dan 10 aturan teratas diberikan

sesuai dengan nilai keyakinan. Aturan terbaiknya adalah: pelanggan yang membeli susu, manisan, dan pizza pepperoni (beku) juga akan mendapatkan telur. Aturan terbaik dengan nilai keyakinan 21.06 dan kepercayaan 1 (100%) adalah aturan ini. 24 pelanggan yang menerima 3 produk ini dalam dataset menerima telur. Demikian pula dengan aturan-aturan lain yang diinterpretasikan dalam penelitian ini. Hasilnya, penempatan produk di supermarket dapat dilakukan sesuai dengan aturan-aturan tersebut. Hal ini akan meningkatkan penjualan produk tersebut dan secara langsung meningkatkan pendapatan supermarket.

Pada penelitian berjudul “*Study of application of data mining market basket analysis for knowing sales pattern (association of items) at the O! Fish restaurant using apriori algorithm*” yang ditulis oleh Kurnia et al. (2019). Penelitian ini bertujuan untuk mengolah data transaksi yang dimiliki oleh Restoran O! Fish menggunakan metode aturan asosiasi. Teknik data mining dengan algoritma Apriori digunakan untuk menghasilkan aturan asosiasi, yang membantu restoran merancang promosi yang lebih efektif. Dalam pengujian dengan 150 data transaksi, penulis menetapkan nilai minimum support 4% dan confidence 60% untuk menentukan aturan asosiasi yang kuat. Hasilnya adalah 7 aturan asosiasi yang mengungkap kombinasi item-menu yang sering dibeli bersamaan. Dengan penerapan aturan ini, restoran dapat mengembangkan strategi promosi yang lebih tepat sasaran dan meningkatkan penjualan dibandingkan dengan promosi manual.

Pada penelitian berjudul “*Comparison of Apriori, Apriori-TID and FP-Growth Algorithms in Market Basket Analysis at Grocery Stores*” yang ditulis oleh Idris et al. (2022). Penelitian ini bertujuan untuk membandingkan metode *Apriori*, *Apriori-TID*, dan *FP-Growth* dalam menentukan perilaku transaksi konsumen dan menghitung jumlah transaksi konsumen dalam beberapa musim berdasarkan data yang diperoleh dari database Analisis Keranjang Belanja. Data yang digunakan pada penelitian ini merupakan data transaksi dari toko grosir. Hasil dari penelitian ini menemukan bahwa algoritma *Apriori* memiliki waktu komputasi paling lama dibandingkan dua algoritma lainnya, karena harus memindai basis data setiap kali iterasi, sehingga waktu yang dibutuhkan meningkat setiap kali iterasi. Sedangkan algoritma *FP-Growth* memiliki waktu komputasi yang lebih cepat dibandingkan kedua algoritma tersebut.

Berikut merupakan tabel penelitian terdahulu mengenai metode *AR-MBA* yang berisikan penulis, objek, metode, dan hasil dari penelitian terdahulu:

Tabel 2. 2 Kajian Induktif *AR-MBA*

No	Penulis	Objek	Metode	Hasil
----	---------	-------	--------	-------

1	(Bunda, 2020)	Data Transaksi dari Pusat Oleh-oleh Umami Afa Hakim	AR-MBA menggunakan algoritma <i>FP-Growth</i>	Dalam penelitian ini, hasil perhitungan <i>FP-Growth</i> menghasilkan aturan asosiasi dengan <i>minimum support</i> sebesar 3 dan <i>minimum confidence</i> sebesar 0.5.
2	(Ariestya et al., 2019)	Data transaksi	AR-MBA menggunakan algoritma <i>FP-Growth</i> dan <i>Apriori</i>	Algoritma <i>FP-Growth</i> menghasilkan 5 aturan (rule) sementara Algoritma <i>Apriori</i> menghasilkan 3 aturan (rule).
3	(Fathurrahman et al., 2023)	Data transaksi dari Jenis perusahaan <i>bakery</i>	AR-MBA menggunakan algoritma <i>FP-Growth</i> dan <i>Apriori</i>	Algoritma <i>Apriori</i> menghasilkan pola kombinasi dengan <i>confidence</i> tertinggi "Jika membeli Alfajores, maka juga membeli Coffe" dan "Jika membeli Cake, maka juga membeli Coffe".
4	(Ünvan, 2021)	Data transaksi dari supermarket pada situs web Vancouver Island University	AR-MBA menggunakan algoritma <i>FP-Growth</i> dan <i>Apriori</i>	Dalam penelitian ini, ditemukan bahwa 64% pelanggan yang membeli Asuransi Casco juga memperoleh Asuransi Trafik.
5	(Kurnia et al., 2019)	Data Transaksi dari Restoran O! Fish	AR-MBA menggunakan algoritma <i>Apriori</i>	Penulis menetapkan nilai minimum support 4% dan confidence 60% dan mendapatkan sebanyak 7 aturan asosiasi.
6	(Idris et al., 2022)	Data transaksi dari Toko Grosir	AR-MBA menggunakan algoritma <i>Apriori</i> , <i>Apriori-TID</i> , dan <i>FP-Growth</i>	Berdasarkan waktu komputasi yang digunakan, algoritma <i>FP-Growth</i> memiliki kinerja terbaik di antara dua algoritma lainnya

### 2.1.3 Integrasi Clustering dan Ar

Pada penelitian berjudul “Penerapan Algoritma *FP-Growth* dan *K-means* pada Data Transaksi Minimarket” yang ditulis oleh Mamahit & Qoiriah (2019). Dalam penelitian ini, tantangan yang dihadapi oleh pembeli dalam menemukan kebutuhan sehari-hari di minimarket

dibahas. Prosedur penataan produk yang acak saat ini tidak sesuai dengan pola belanja pembeli, dan ketiadaan paket produk semakin memperumit pengalaman berbelanja pembeli. Untuk mengatasi masalah ini, diusulkan pendekatan yang lebih efektif dengan menganalisis data transaksi di minimarket untuk mengungkap pola hubungan. Metode association rule, bersama dengan algoritma *FP-Growth* dan *K-means*, digunakan untuk penelitian ini. Dataset dibagi menjadi 10 kelompok dengan menggunakan proses *K-means*, sedangkan proses *FP-Growth* menggabungkan kriteria *minimum support*, *minimum confidence*, dan *lift ratio* dengan *minimum support* 20% dan *minimum confidence* 50%. Temuan penelitian ini memberikan rekomendasi untuk penataan 10 rak di minimarket dan menyarankan 21 paket/bundel untuk meningkatkan pengalaman berbelanja dan memenuhi kebutuhan pembeli.

Pada penelitian berjudul “*Determination of Student Subjects in Higher Education Using Hybrid Data Mining Method with the K-means Algorithm and FP Growth*” yang ditulis oleh Rani et al. (2021). Penelitian ini bertujuan untuk mengelompokkan data nilai mahasiswa dan melihat pola kemunculan data berdasarkan nilai yang telah dikelompokkan, sehingga mahasiswa dapat menggunakan hasil pemodelan tersebut sebagai panduan dalam memilih mata kuliah spesialisasi pada semester berikutnya. Metode yang digunakan dalam penelitian ini adalah metode *K-means* dan *FP-Growth*. Hasil penelitian ini memberikan masukan kepada mahasiswa atau pembimbing akademik dalam menyusun kartu rencana studi mahasiswa. Dosen dan mahasiswa dapat menganalisis mata kuliah spesialisasi yang tepat dengan mengikuti pola yang diberikan. Dalam penelitian ini, terdapat tiga kelompok mata kuliah, yaitu mata kuliah spesialisasi, mata kuliah pilihan 1, dan mata kuliah pilihan 2. Berdasarkan hasil keluaran, terlihat bahwa 6 mahasiswa memilih mata kuliah pilihan 1, 5 mahasiswa memilih mata kuliah pilihan 2, dan 5 mahasiswa memilih mata kuliah pilihan 3.

Pada penelitian yang dilakukan oleh Liew (2018) berjudul “*Dietary habits and physical activity: Results from cluster analysis and market basket analysis*”. Penelitian ini bertujuan untuk mengidentifikasi pola kebiasaan makan yang berhubungan dengan aktivitas fisik serta persepsi siswa terhadap kualitas makanan yang tersedia di kantin sekolah dan mesin penjual otomatis. Penelitian ini didasarkan pada evaluasi *Healthy School Programme* (HSP) tahun 2011 yang mengumpulkan data tentang karakteristik demografis, kebiasaan makan, dan pola aktivitas fisik dari sampel siswa sekolah dasar, menengah, dan atas yang mewakili populasi di Amerika Serikat. Hasil penelitian menunjukkan bahwa siswa yang termasuk dalam kelompok yang berbeda memiliki kebiasaan makan, pola latihan fisik, status berat badan, pengelolaan berat badan, dan pendapat tentang kualitas makanan di kantin sekolah dan mesin penjual otomatis yang berbeda pula. Terdapat perbedaan yang signifikan dalam profil makanan dan

perilaku gaya hidup siswa yang menganggap diri mereka *overweight* atau tidak yakin tentang status berat badan mereka. Temuan dari penelitian ini dapat digunakan sebagai dasar untuk intervensi di masa depan yang bertujuan untuk mendorong pilihan makanan sehat di kalangan siswa selama mereka berada di sekolah. Inisiatif promosi kesehatan harus secara khusus ditujukan kepada siswa dengan profil makanan yang tidak sehat secara konsisten, dan dapat membantu mengurangi masalah obesitas pada anak-anak.

Pada penelitian berjudul “*USING K-MEANS ALGORITHM AND FP-GROWTH BASE ON FP-TREE STRUCTURE FOR RECOMMENDATION CUSTOMER SME*” yang ditulis oleh Syakur et al. (2018). Dalam penelitian ini, Market Basket telah menemukan pola pembelian pelanggan di usaha kecil dan menengah (UKM). Pola pembelian ini dapat digunakan untuk membuat rekomendasi produk dan promosi. Penelitian ini menggunakan algoritma *K-means* untuk mengelompokkan data penjualan dan algoritma *FP-Growth* untuk menentukan hubungan antar kelompok. *K-means* digunakan untuk mengelompokkan pelanggan berdasarkan atribut yang sama dan kemudian *FP-Growth* digunakan untuk menentukan hubungan antar pola pada setiap kelompok. *K-means* digunakan untuk mensegmentasi pelanggan berdasarkan latar belakang, karakteristik pelanggan dan tingkat daya beli. Untuk memudahkan analisis hubungan pelanggan dengan produk yang dibeli, maka setiap kelompok pelanggan diolah menggunakan *FP-Growth* untuk mengetahui relevansi barang yang dibeli. Penelitian ini juga menyajikan perbandingan kompleksitas waktu antara algoritma *FP-Growth* dan algoritma Apriori. Penelitian ini mengembangkan dan mengimplementasikan penggunaan struktur data *FP-Tree* (Frequent Pattern Tree) yang merupakan perluasan dari penggunaan tree pada struktur data. *FP-Tree* digunakan bersama dengan algoritma *FP-Growth* untuk menentukan kumpulan item yang sering muncul dalam database, sebagai lawan dari paradigma apriori yang melakukan pemindaian ulang database untuk menentukan kumpulan item yang sering muncul. Dalam penelitian ini, jumlah transaksi multi-item dan daya beli konsumen bervariasi dan pertama-tama dikelompokkan menggunakan algoritma *K-means*, menghasilkan beberapa kelompok pelanggan, termasuk lima kelompok pelanggan berdasarkan profil pelanggan. Hasil pengujian rata-rata adalah  $\text{minsupp} = 60$  dan  $\text{minconf} = 40$ , memberikan waktu pemrosesan rata-rata 957ms.

Pada penelitian berjudul “*Implementation of K-Medoids and FP-Growth Algorithms for Grouping and Product Offering Recommendations*” yang ditulis oleh Syukra et al. (2019). Dalam penelitian ini, fokus penelitian dilakukan pada ritel 212 Mart yang berlokasi di Jalan Rambutan, Kota Pekanbaru. Dalam rangka memenuhi kebutuhan konsumen dan untuk mengambil keputusan yang tepat dalam merancang strategi penjualan, keberadaan perusahaan



ini memegang peranan penting. Salah satu pendekatan untuk memahami kondisi pasar adalah dengan melakukan analisis data transaksi penjualan dengan menggunakan teknik data mining. Dalam konteks market basket analysis, metode yang umum digunakan adalah association rules. Aturan asosiasi dapat digunakan untuk merekomendasikan produk dan promosi yang tepat, membuat strategi pemasaran lebih tepat sasaran dan memastikan bahwa barang yang dipromosikan sesuai dengan kebutuhan pelanggan. Di 212 Mart, promosi produk ditentukan dengan menganalisis laporan data transaksi penjualan berdasarkan produk terlaris dan tanggal kadaluarsa. Namun, sering kali produk yang dipromosikan tidak selalu sesuai dengan kebutuhan pelanggan. Oleh karena itu, tujuan dari penelitian ini adalah menerapkan algoritma K-Medoids untuk melakukan *clustering* pada *FP-Growth* dengan harapan dapat menghasilkan aturan rekomendasi produk yang lebih baik dalam menghadapi dataset yang besar. Dengan demikian, penelitian ini diharapkan dapat memberikan rekomendasi teknis atau metode baru kepada 212 Mart dalam menentukan strategi promosi produk. Hasil percobaan menunjukkan bahwa dengan menggunakan jumlah *cluster* antara 3 sampai 9, 3 *cluster* dianggap optimal berdasarkan uji validitas Davies Bouldin Index dengan nilai 0.678. Dengan menggunakan nilai *minimum support* sebesar 5%-9% dan nilai *minimum confidence* sebesar 50%, didapatkan bahwa aturan asosiasi hanya terdapat pada *cluster* 3 dengan 5 aturan yang relevan.

Pada penelitian berjudul “*Using RFM model and market basket analysis for segmenting customers and assigning marketing strategies to resulted segments*” yang ditulis oleh Maraghi et al. (2020). Penelitian ini membahas implementasi *Customer Relationship Management* (CRM) dalam supermarket dengan tujuan membangun hubungan kuat dengan pelanggan dan menghasilkan keuntungan maksimal. Manajemen supermarket memahami bahwa pelanggan memiliki beragam kebutuhan, gaya, dan harapan yang berbeda, sehingga mereka melakukan segmentasi pelanggan untuk merespons permintaan yang berbeda dengan tepat. Selain itu, penting bagi manajemen supermarket untuk mengidentifikasi pelanggan yang menguntungkan, yang menyumbang keuntungan utama perusahaan. Dalam penelitian ini, dataset transaksi dibuat dengan menggunakan metrik *recency*, *frequency*, dan *monetary* (RFM), kemudian dilakukan pengelompokan pelanggan dengan algoritma K-means, menghasilkan enam segmen pelanggan. Setelah itu, dilakukan analisis mendalam terhadap masing-masing segmen dan merumuskan strategi pemasaran yang sesuai. Data transaksi masing-masing segmen dipisahkan dan aturan asosiasi diekstraksi menggunakan analisis keranjang belanja pasar dan algoritma Apriori. Hasilnya adalah model CRM yang mencakup pemahaman pelanggan hingga pemberian strategi pemasaran yang tepat berdasarkan pemahaman yang akurat tentang semua pelanggan, termasuk identifikasi pelanggan yang menguntungkan.

Tabel 2. 3 Kajian Induktif Integrasi *Fp-Growth* dan *K-Means*

No	Penulis	Objek	Metode	Hasil
1	(Mamahit & Qoiriah, 2019)	Minimarket X	Integrasi algoritma <i>FP-Growth</i> dan <i>K-means</i>	Hasil dari penelitian ini, penggunaan algoritma <i>FP-Growth</i> dan <i>K-means</i> berhasil menghasilkan rekomendasi penataan barang di 10 rak minimarket serta 21 paket atau bundle.
2	(Rani et al., 2021)	Universitas Putra Indonesia	Integrasi algoritma <i>FP-Growth</i> dan <i>K-means</i>	Hasil dari penelitian ini, terlihat bahwa 6 mahasiswa memilih mata kuliah pilihan 1, 5 mahasiswa memilih mata kuliah pilihan 2, dan 5 mahasiswa memilih mata kuliah pilihan 3.
3	(Liew, 2018)	<i>Healty School Program</i> (HSP) Di Amerika Serikat	Integrasi metode MBA dan <i>Clustering</i>	Perbedaan dalam kebiasaan makan, pola latihan, status berat badan, pengelolaan berat badan, dan pendapat tentang makanan di kantin sekolah dan mesin penjual otomatis pada siswa dalam <i>cluster-cluster</i> yang berbeda.
4	(Syakur et al., 2018)	Minimarket X	Integrasi metode MBA dan <i>Clustering</i>	Dari data tersebut terdapat empat <i>cluster</i> dengan nilai MSE terendah, dapat disimpulkan bahwa <i>cluster</i> 4 memiliki tingkat homogenitas yang paling tinggi dari <i>cluster</i> yang ada. Dari hasil MSE juga dapat dilihat bahwa <i>cluster</i> 4 merupakan <i>cluster</i> yang paling optimal.
5	(Syukra et al., 2019)	212 Mart	Integrasi metode MBA dan <i>Clustering</i>	Dengan menggunakan nilai <i>minimum support</i> sebesar 5%-9% dan nilai <i>minimum confidence</i> sebesar 50%, didapatkan bahwa aturan asosiasi hanya terdapat pada <i>cluster</i> 3 dengan 5 aturan yang relevan.
6	(Maraghi et al., 2020)	Supermarket X	Integrasi metode MBA dan <i>Clustering</i>	Model CRM yang mencakup pemahaman pelanggan hingga pemberian strategi pemasaran

yang tepat berdasarkan pemahaman yang akurat tentang semua pelanggan, termasuk identifikasi pelanggan yang menguntungkan

---

#### 2.1.4 Posisi Penelitian

Dalam penelitian ini, peneliti mengacu pada kajian induktif yang telah dilakukan sebelumnya untuk membangun landasan yang kuat. Berdasarkan kajian tersebut, peneliti memilih metode *AR-MBA (Association Rule-Market Basket Analysis)* dengan menggunakan algoritma *FP-Growth*, serta metode *Clustering* dengan menggunakan algoritma *K-means*. Keputusan ini didasarkan pada pengalaman dan keberhasilan penggunaan algoritma-algoritma tersebut oleh peneliti-peneliti sebelumnya. Selain itu, algoritma *FP-Growth* dan *K-means* dipilih karena sesuai dengan permasalahan yang dihadapi oleh Toko Intimart Gedongan. Penelitian ini berfokus pada menggali wawasan yang mendalam tentang preferensi pembelian pelanggan dan mengidentifikasi kelompok pelanggan yang serupa berdasarkan pola pembelian mereka. Hasil dari penelitian ini akan memberikan rekomendasi strategi pemasaran yang tepat sasaran untuk meningkatkan kinerja penjualan di Toko Intimart Gedongan.

## 2.2 Landasan Teori

### 2.2.1 Ritel

Ritel menurut definisi Nurcahyo (2019) adalah aktivitas pemasaran produk, baik barang maupun jasa, yang dilakukan dengan cara menjualnya secara eceran atau satuan langsung kepada konsumen akhir untuk digunakan oleh rumah tangga atau individu, bukan untuk tujuan penjualan kembali. Bisnis ritel adalah kegiatan ekonomi yang melibatkan penjualan produk kebutuhan sehari-hari kepada masyarakat dalam bentuk eceran, kegiatan tersebut dapat dalam bentuk tradisional maupun modern (Chaniago, 2021). Dari definisi diatas ritel adalah aktivitas memasarkan produk yang dilakukan dengan menjualannya langsung kepada konsumen akhir dalam bentuk eceran atau satuan.

Bisnis ritel di Indonesia telah terjadi sejak lama. Menurut Asosiasi Pengusaha Ritel Indonesia (APRINDO), bisnis ritel atau usaha eceran di Indonesia mulai berkembang pada kisaran tahun 1980an seiring dengan mulai dikembangkannya perekonomian Indonesia. Pertumbuhan bisnis ritel ini dipicu oleh masyarakat kelas menengah yang semakin bertambah

seiring dengan waktu. Hal ini menyebabkan banyaknya permintaan untuk membangun *supermarket* atau *minimarket* di wilayah perkotaan (Soliha, 2008).

Ritel sendiri memiliki peran penting dalam menjalankan bisnisnya. Menurut (Nurchayyo, 2019) dalam bukunya yang berjudul “Pengelolaan Bisnis Ritel”, industri ritel mempunyai peran penting dalam 2 pihak, yaitu:

1. Peran penting bisnis ritel dari sisi produsen.

Bisnis ritel merupakan pihak yang memiliki keahlian dalam penjualan produk perusahaan dan berperan sebagai ujung tombak yang sangat mempengaruhi kesuksesan penjualan produk. Bisnis ritel juga menjadi sumber informasi berharga mengenai produk yang mereka jual. Selain itu, produsen dapat menggunakan strategi pemasaran seperti pemasangan iklan, pengadakan undian, atau pemberian hadiah kepada konsumen melalui toko-toko ritel. Melalui kerjasama dengan bisnis ritel, produsen dapat mencapai target pasar yang lebih luas dan memperkuat citra produk mereka.

2. Peran penting bisnis ritel dari sisi konsumen.

Bisnis ritel memberikan manfaat bagi konsumen dan masyarakat secara umum. Pertama, lokasi usaha ritel yang dekat dengan rumah pelanggan memungkinkan pelanggan untuk dengan cepat mendapatkan produk yang mereka butuhkan tanpa harus menunggu lama. Kedua, usaha ritel memudahkan pelanggan dalam memilih dan membandingkan berbagai jenis produk, baik dalam hal bentuk, kualitas, maupun harga. Para peritel berusaha menciptakan suasana belanja yang nyaman untuk memuaskan pelanggan. Ketiga, usaha ritel berperan dalam menjaga harga jual tetap rendah agar dapat bersaing dalam memenuhi kebutuhan pelanggan. Keempat, melalui promosi dan penjualan produk, usaha ritel membantu meningkatkan standar hidup masyarakat. Terakhir, kehadiran usaha ritel memungkinkan dilakukannya produksi massal yang efektif dalam mendistribusikan produk kepada pelanggan. Dengan demikian, usaha ritel berperan penting dalam memenuhi kebutuhan pelanggan, meningkatkan standar hidup, dan mendorong pertumbuhan ekonomi.

Dalam buku “*MANAJEMEN RITEL & IMPLEMENTASINYA*” yang ditulis oleh Chaniago (2021) membagi ritel dalam beberapa sisi yaitu dari ukurannya, teknologi, dan pemasaran. Berikut merupakan pembagian rinci mengenai 3 sisi tersebut:

1. Dari sisi ukurannya (besar atau kecilnya lahan dan barang yang diperjual belikan) dibagi atas:

- a. Hypermarket dengan luas lahan lebih dari 5000 m persegi.
- b. Supermarket dengan luas lahan diantara 1000 m sampai 5000 m persegi.

- c. Minimarket dengan luas lahan diantara 100 m hingga 1000 m persegi.
  - d. Nanostore dengan luas lahan dibawah 50 m persegi.
2. Dari sisi teknologi dan pelayanan pada konsumen, toko ritel dibagi menjadi 2 yaitu:
- a. Ritel Modern
 

Ritel modern merupakan bentuk bisnis yang memiliki karakteristik dan ciri-ciri khusus. Dengan lokasi yang strategis, pelayanan yang baik, fasilitas lengkap, dan pengelolaan yang terencana, ritel modern berusaha memberikan pengalaman berbelanja yang optimal bagi konsumen. Selain itu, kepemilikan oleh perorangan atau kelompok perusahaan besar juga mempengaruhi sistem operasional dan pengembangan bisnis ritel modern.
  - b. Ritel Tradisional
 

Ritel tradisional memiliki ciri-ciri yang berbeda dengan ritel modern. Terdapat keterbatasan dalam ketersediaan barang dagang, penggunaan teknologi, manajemen keuangan, evaluasi bisnis, dan pengembangan bisnis yang belum terencana secara baik. Meskipun demikian, nanostore masih merupakan salah satu saluran distribusi yang paling banyak ditemui dan tersebar di berbagai daerah di Indonesia.
3. Dari sisi pemasarannya, toko ritel dibagi menjadi 3 yaitu:
- a. Ritel *offline*. Ritel jenis ini akan melayani konsumen atau pelanggan secara langsung.
  - b. Ritel *online*. Ritel jenis ini menggunakan model bisnis *online* dengan memanfaatkan teknologi IT sebagai tempat untuk menjual produk. Transaksi dilakukan secara virtual dimana pembeli melakukan pembayaran secara *online* melalui internet. Interaksi penjual dan pembeli dalam ritel jenis ini dilakukan menggunakan *platform online* seperti website, sosial media, dan email.
  - c. Ritel off-on (*offline* dan *online*). Ritel jenis ini merupakan ritel dengan menggabungkan model ritel *offline* dengan ritel *online*. Dengan menggunakan strategi ini, ritel dapat memanfaatkan keuntungan dari kedua model penjualan.

### 2.2.2 Data mining

*Data mining* menurut definisi oleh Kantardzic (2020) adalah proses iteratif yang melibatkan eksplorasi dan penemuan pola, hubungan, dan wawasan dalam sebuah kumpulan data, yang dapat dicapai melalui penggunaan algoritma otomatis maupun metode manual. *Data mining* menurut Muslim et al. (2019) merupakan langkah ekstraksi sebuah data untuk mendapatkan informasi yang sebelumnya tidak diketahui atau bersifat implisit. *Data mining*

disini juga berkaitan dengan bidang lain seperti statistik, *machine learning*, *pattern recognition*, teknologi *database*, dan komputasi berkinerja tinggi. Menurut Han et al. (2011), *Data mining* adalah metodologi yang digunakan untuk menemukan pola-pola menarik dalam volume data yang besar. Proses ini melibatkan beberapa tahap, seperti pembersihan data, integrasi data, pemilihan data, transformasi data, identifikasi pola, evaluasi pola, dan penyajian pengetahuan.

Proses *data mining* adalah sebuah rangkaian langkah yang melibatkan beberapa fase. Menurut Aggarwal (2015) ada tiga fase dalam proses *data mining*, yaitu:

### 1. *Data collection*

Pengumpulan data dapat membutuhkan penggunaan perangkat khusus seperti jaringan sensor, tenaga kerja manual seperti pengumpulan survei pengguna, atau perangkat lunak seperti mesin penjelajahan dokumen Web untuk mengumpulkan dokumen. Meskipun tahap ini sangat spesifik tergantung pada aplikasi dan sering diluar cakupan *data mining*, tahap ini sangat penting karena pilihan yang baik pada tahap ini dapat berdampak signifikan pada proses *data mining*. Setelah tahap pengumpulan, data sering disimpan dalam *database*, atau secara umum, gudang data (*data warehouse*) untuk diproses.

### 2. *Feature extraction* dan *data cleaning*

Setelah data dikumpulkan, seringkali data tersebut tidak dalam bentuk yang sesuai untuk diproses. Misalnya, data dapat terkode dalam log yang kompleks atau *free-form document*. Dalam banyak kasus, berbagai jenis data dapat bercampur secara acak dalam *free-form document*. Untuk membuat data cocok untuk diproses, penting untuk mentransformasikannya ke dalam format yang ramah terhadap algoritma *data mining*, seperti format multidimensional, *time series*, atau semistruktur. Format multidimensional adalah yang paling umum digunakan, di mana bidang-bidang data sesuai dengan berbagai properti yang diukur yang disebut sebagai fitur, atribut, atau dimensi. Penting untuk mengekstraksi fitur yang relevan untuk proses *mining*. Tahap ekstraksi fitur sering dilakukan secara paralel dengan pembersihan data, di mana bagian-bagian data yang hilang atau salah diperkirakan atau diperbaiki. Dalam banyak kasus, data dapat diekstraksi dari beberapa sumber dan perlu diintegrasikan ke dalam format yang terpadu untuk diproses. Hasil akhir dari prosedur ini adalah kumpulan data yang terstruktur dengan baik, yang dapat digunakan secara efektif oleh program komputer. Setelah tahap ekstraksi fitur, data dapat kembali disimpan dalam *database* untuk diproses.

### 3. *Analytical processing and algorithms*

Setelah data melewati 2 fase diatas, proses terakhir dari *data mining* adalah merancang metode analisis yang efektif dari data yang telah diproses.

Menurut Kantardzic (2020) *data mining* memiliki tujuan secara umum yang dapat dibagi menjadi 2 kategori, yaitu:

1. Prediktif

Tujuan *data mining* adalah untuk memperkirakan atau memprediksi nilai atribut tertentu berdasarkan nilai atribut lainnya. Secara umum, variabel yang diprediksi disebut sebagai variabel target atau variabel tak bebas. Di sisi lain, atribut yang digunakan untuk membuat prediksi disebut sebagai *explanatory* variabel atau variabel bebas.

2. Deskriptif

Tujuan *data mining* adalah untuk menemukan pola (korelasi, tren, *cluster*, trayektori, dan anomali) yang menggambarkan hubungan penting dalam data. Tugas deskriptif *data mining* biasanya memerlukan penyelidikan dan teknik *postprocessing* seringkali diperlukan untuk validasi dan penjelasan hasil.

Pada buku berjudul “*Data mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*” yang ditulis oleh Han et al. (2011). dinyatakan bahwa *data mining* memiliki berbagai aplikasi yang relevan, seperti pada bidang *business intelligence*, *Websearch*, *bioinformatics*, *health informatics*, *finance*, *digital libraries*, dan *digital governments*. Berikut ini adalah dua contoh di mana *data mining* sering digunakan:

1. *Business Intelligence*

Teknologi *Business Intelligence*, atau intelegensi bisnis, memberikan informasi tentang operasi bisnis, termasuk data saat ini, data historis, dan perkiraan. Pembuatan laporan, pengolahan analisis online, manajemen kinerja bisnis, informasi kompetitif, perbandingan, dan analitika prediktif adalah beberapa contohnya. Banyak perusahaan mungkin tidak dapat melakukan analisis pasar yang efektif, membandingkan umpan balik pelanggan terhadap product serupa, menemukan kekuatan dan kelemahan pesaing, mempertahankan pelanggan berharga, dan membuat keputusan bisnis yang cerdas jika mereka tidak using teknik *data mining*.

2. *Web Search Engines*

*Web Search Engines* atau mesin pencari adalah sebuah server komputer yang dirancang untuk mencari data di internet. Seringkali, hasil pencarian pengguna dikembalikan dalam bentuk daftar. Hasilnya dapat berupa halaman web, foto, atau jenis file lainnya. Selain itu,

beberapa mesin pencari mencari dan mengembalikan data yang tersedia di direktori terbuka atau base data publik.

### 2.2.3 Association Rule

*Association rule* adalah sebuah teknik dalam data mining yang bertujuan untuk menemukan pola asosiasi dalam suatu kejadian Muslim et al. (2019). Menurut Baskoro et al. (2013), *Association Rules* merupakan sebuah metodologi yang digunakan untuk mengidentifikasi hubungan yang menarik dan tersembunyi dalam himpunan data besar. Salah satu aplikasi utama dari metode ini adalah pada Analisis Keranjang Belanja (*Market Basket Analysis*). Penambangan pola asosiasi awalnya diusulkan dalam konteks penambangan aturan asosiasi, di mana langkah tambahan dimasukkan berdasarkan ukuran yang dikenal sebagai tingkat kepercayaan (*confidence*) dari aturan tersebut (Wanto et al., 2020). Tahap awal adalah mencari frequent itemset, yaitu kombinasi item yang sering muncul dan memenuhi persyaratan *minimum support*. Pencarian dilakukan berdasarkan nilai *support* dalam database (Amelia & Utomo, 2019).

Dalam buku berjudul “Belajar *Data Mining* dengan *Rapid Miner*” yang ditulis Baskoro et al. (2013) dalam penentuan *Association Rule*, terdapat sebuah metrik yang disebut sebagai ukuran ketertarikan (*interestingness measure*). Ukuran ini diperoleh melalui proses pengolahan data dengan menggunakan perhitungan khusus. Terdapat dua jenis ukuran yang umum digunakan, yaitu:

1. *Support*: Bagian transaksi yang mengandung kedua X dan Y.

$$Support(A) = \frac{Jumlah\ transaksi\ mengandung\ A}{Total\ Transaksi} \quad (2.1)$$

2. *Confidence*: Seberapa sering item dalam Y muncul di transaksi yang mengandung X.

$$Confidence = P(B|A) = \frac{Jumlah\ transaksi\ mengandung\ A\ dan\ B}{Jumlah\ transaksi\ mengandung\ A} \quad (2.2)$$

Aturan asosiasi (*association rule*) memiliki manfaat yang signifikan bagi pemilik toko. Beberapa manfaatnya antara lain, pertama, dapat membantu menentukan penempatan barang-barang secara tepat dalam layout toko. Sebagai contoh, susu dapat diletakkan berdekatan dengan diapers karena memiliki asosiasi yang kuat. Kedua, aturan asosiasi juga digunakan untuk promosi produk, di mana produk yang sering dibeli bersamaan dapat dipromosikan



bersama untuk meningkatkan penjualan. Ketiga, aturan asosiasi dapat membantu dalam segmentasi pembeli, yaitu mengelompokkan pelanggan berdasarkan pola pembelian mereka untuk strategi pemasaran yang lebih efektif. Keempat, aturan asosiasi juga dapat digunakan dalam pembuatan katalog produk yang relevan dan menarik bagi pelanggan. Kelima, dengan menganalisis pola kecenderungan pola belanja pelanggan, pemilik toko dapat mendapatkan wawasan yang berharga untuk pengambilan keputusan bisnis yang lebih cerdas (Wanto et al., 2020).

#### 2.2.4 Market Basket Analysis

Menurut Nurwahdania et al. (2021), Analisis Keranjang Belanja (*Market Basket Analysis*) adalah proses untuk menganalisis pola pembelian konsumen dengan menemukan hubungan antara item-item yang berbeda yang ditempatkan dalam keranjang belanja mereka. Istilah Analisis Keranjang Belanja (*Market Basket Analysis*) berasal dari kejadian yang sering terjadi di pasar swalayan, di mana konsumen menaruh semua barang yang ingin mereka beli ke dalam keranjang yang biasanya disediakan oleh swalayan tersebut (Baskoro et al., 2013). *Market Basket Analysis* didasarkan pada asosiasi antara keputusan pembelian konsumen. Saat berbelanja di *supermarket*, konsumen cenderung membeli lebih dari satu produk, bahkan seluruh keranjang produk, terutama dari kategori produk yang berbeda (Elisa, 2018).

Menurut Han et al. (2011) dalam bukunya menjelaskan MBA atau *Market Basket Analysis* dapat membantu pemilik toko untuk memahami kebiasaan pembelian pelanggan. Dengan menganalisis data transaksi pelanggan, pemilik toko dapat mengetahui kelompok atau set barang yang cenderung dibeli dalam satu kunjungan ke toko. Hasil analisis ini dapat digunakan untuk merencanakan strategi pemasaran, periklanan, atau dalam desain katalog baru. Ada dua cara untuk mempresentasikan data dalam Analisis Keranjang Belanja (*Market Basket Analysis*) (Larose & Larose, 2015):

1. Menggunakan format data transaksional.

Format data transaksional membutuhkan hanya dua atribut, yaitu atribut ID transaksi dan atribut konten atau isi belanja dari transaksi tersebut.

2. Menggunakan format data tabular.

Dalam format data tabular, setiap catatan mewakili satu transaksi terpisah dengan sejumlah atribut penanda 0/1 sesuai dengan jumlah item yang ada.

### 2.2.5 Clustering

*Clustering* adalah proses pengelompokan sekumpulan objek data menjadi beberapa kelompok atau *cluster*, sehingga objek-objek dalam satu *cluster* memiliki kemiripan tinggi, namun sangat berbeda dengan objek-objek dalam kelompok lainnya (Kantardzic, 2020). *Clustering* berbeda dari klasifikasi karena tidak ada variabel target dalam *clustering*. Tugas *clustering* tidak mencoba mengklasifikasikan, memperkirakan, atau memprediksi nilai dari variabel target (Larose & Larose, 2015). Metode ini dapat digunakan untuk memberikan label pada data yang belum diketahui, karena *clustering* sering dikategorikan sebagai metode pembelajaran tanpa pengawasan (*unsupervised learning*) (Muslim et al., 2019).

Keberhasilan klasifikasi dalam *clustering* tergantung pada bentuk distribusi vektor karakteristik pola. Jika titik-titik vektor terkelompokkan dengan baik, dengan titik-titik sampel dalam kelompok yang sama terkonsentrasi dan titik-titik sampel dalam kelompok yang berbeda berjarak jauh, maka penggunaan fungsi jarak untuk mengklasifikasikan titik-titik tersebut akan lebih mudah (Li & Wu, 2012). Algoritma yang ada pada *clustering* didasari oleh 2 pendekatan (Kantardzic, 2020), yaitu:

1. *Hierarchical Clustering*: Teknik pengelompokan hirarkis mengorganisasi data dalam urutan kelompok yang bersarang. Hasilnya dapat ditampilkan dalam bentuk dendrogram atau struktur pohon. Algoritma ini membagi data menjadi kelompok yang semakin kecil atau menggabungkan kelompok yang semakin besar secara hierarkis.
2. *Terative Square-Error Partitional Clustering*: Algoritma ini berusaha mendapatkan partisi yang meminimalkan penyebaran dalam kluster atau memaksimalkan penyebaran antar kluster. Tujuannya adalah untuk mengelompokkan data menjadi kluster yang memiliki kesamaan tertentu.

Dalam melakukan *clustering* terdapat langkah-langkah dalam proses pembuatan kelompok atau *cluster*, yaitu (Halkidi et al., 2001):

1. Seleksi Fitur.

Tujuannya adalah memilih dengan tepat fitur-fitur yang akan digunakan dalam *clustering* agar dapat mengkodekan sebanyak mungkin informasi yang berkaitan dengan tugas yang menjadi minat kita. Oleh karena itu, pra-pemrosesan data mungkin diperlukan sebelum digunakan dalam tugas *clustering*.

2. Algoritma *Clustering*.

Langkah ini merujuk pada pemilihan algoritma yang menghasilkan definisi skema *clustering* yang baik untuk satu set data.

3. Validasi hasil.

Kevalidan hasil algoritma *clustering* diperiksa menggunakan kriteria dan teknik yang sesuai.

#### 4. Interpretasi hasil.

Penerjemahan hasil *clustering* menjadi pemahaman yang bermakna dalam konteks aplikasi, serta memberikan wawasan dan rekomendasi berdasarkan analisis *clustering* kepada pemangku kepentingan yang terkait.

Beberapa contoh dalam penerapan *clustering* dapat dilihat pada poin dibawah ini (Aggarwal, 2015):

##### 1. Segmentasi pelanggan.

Dalam banyak aplikasi, penting untuk menentukan pelanggan yang serupa satu sama lain dalam konteks berbagai tugas promosi produk. Tahap segmentasi memainkan peran penting dalam proses ini.

##### 2. Ringkasan data.

Karena kelompok dalam *clustering* dapat dianggap sebagai kelompok yang serupa dari catatan (*records*), kelompok serupa ini dapat digunakan untuk membuat ringkasan data.

##### 3. Aplikasi pada masalah data mining lainnya.

Karena *clustering* dianggap sebagai versi *unsupervised* dari klasifikasi, sering digunakan sebagai blok bangunan untuk memecahkan masalah klasifikasi. Selain itu, masalah ini juga digunakan dalam konteks masalah analisis outlier.

#### 2.2.6 Algoritma FP-Growth

*FP-Growth* adalah sebuah algoritma yang digunakan untuk mengidentifikasi himpunan data yang sering muncul (*frequent itemset*) dalam kumpulan data. *FP-Growth* merupakan metode yang efisien untuk menambang himpunan item yang sering muncul dalam basis data besar. Algoritma ini dapat menemukan himpunan item yang sering tanpa perlu melakukan proses pembentukan kandidat (*candidat generation*) yang memakan waktu seperti yang dilakukan oleh algoritma Apriori (Kantardzic, 2020). Menurut Fitria et al. (2017) *FP-Growth* adalah sebuah pengembangan dari algoritma apriori yang mengatasi kekurangan algoritma tersebut dengan menghilangkan tahap pembangkitan kandidat, karena *FP-Growth* menggunakan pendekatan berbasis pohon (*tree-based*) dalam pencarian itemset yang sering muncul. Dalam menerapkan metode *FP-Growth*, metode ini dapat dibagi menjadi tiga tahap utama (Larasati et al., 2015), yaitu:

1. Tahap pembentukan basis pola kondisional.
2. Tahap pembuatan *FP-Tree* kondisional.

3. Tahap pencarian itemset yang sering muncul (*frequent itemset*).

*FP-Tree* dibangun dengan menggunakan 2 kali pemindaian pada himpunan data (Kavitha & Selvi, 2016). Berikut merupakan penjelasannya.

1. Pemindaian pertama, data dipindai dan kemudian dilakukan perhitungan *support* untuk setiap item. Kemudian, *itemset* yang tidak sering muncul diabaikan, dan *itemset* yang sering muncul diurutkan berdasarkan *support* mereka secara menurun.
2. Pemindaian kedua, membentuk *node* yang sesuai dengan *itemset* dan memiliki *counter*.

Dalam *FP-Growth*, *lift ratio* dan *confidence* digunakan sebagai metrik evaluasi untuk mengevaluasi kekuatan dan relevansi dari aturan asosiasi yang dihasilkan. *Lift ratio* digunakan untuk mengevaluasi kepentingan aturan yang telah terbentuk berdasarkan nilai *support* dan *confidence*. *Lift ratio* diperoleh dengan membandingkan nilai *confidence* dengan nilai *benchmark confidence*. *Benchmark confidence* sendiri didapatkan dengan membandingkan jumlah keseluruhan item yang merupakan konsekuensi aturan dengan total jumlah transaksi (Fitria et al., 2017). Perhitungan *lift* dapat dilihat pada rumus dibawah ini (Anggrawan et al., 2021).

$$\text{Lift Rasio} = \frac{\text{confidence}}{\text{benchmark confidence}} \quad (2.3)$$

Sedangkan untuk perhitungan *benchmark confidence* dapat dilihat pada rumus dibawah ini.

$$\text{Confidence Benchmark} = \frac{\sum \text{transaksi dengan item dalam consequent}}{\sum \text{transaksi dalam database}} \quad (2.4)$$

Aturan yang memiliki *lift rasio* lebih besar dari satu dianggap bermanfaat karena menunjukkan tingkat kekuatan asosiasi yang lebih besar. Semakin tinggi nilai *lift rasio*, semakin kuat hubungan asosiasi antara item-item dalam aturan tersebut.

Pada penelitian berjudul “Comparative Study on Apriori Algorithm and Fp Growth Algorithm with Pros and Cons“ oleh Kavitha & Selvi (2016). Dalam penelitian ini, disajikan sebuah studi komparatif antara dua algoritma, yaitu algoritma Apriori dan algoritma FP-Growth. Algoritma Apriori bekerja dengan menemukan set item yang sering muncul dan memverifikasi frekuensi subsetnya melalui pembuatan set kandidat dan pengujian. Sementara itu, algoritma FP-Growth menggunakan pendekatan yang berbeda dengan menghindari pembuatan set kandidat. Algoritma FP-Growth secara efisien menambang pola yang sering muncul dari basis data yang besar dengan memanfaatkan pertumbuhan fragmen pola dan struktur pohon awalan yang diperluas. Dengan mengeliminasi langkah pembuatan kandidat

itemset, algoritma FP-Growth dapat mencapai efisiensi yang lebih baik dibandingkan dengan algoritma Apriori.

### 2.2.7 Algoritma *K-means*

Algoritma *K-means* termasuk dalam kategori pembelajaran tanpa pengawasan (*unsupervised learning*) dan digunakan untuk mengelompokkan data ke dalam beberapa kelompok dengan pendekatan partisi (Wanto et al., 2020). Algoritma ini menggunakan square-error dan kriteria error sebagai dasar untuk meminimalkan indeks kinerja kluster. Dalam mencari hasil optimal, algoritma ini mencoba menemukan pembagian K yang memenuhi suatu kriteria tertentu (Li & Wu, 2012). Algoritma *K-means* sendiri merupakan algoritma yang sederhana dan efektif untuk menemukan kelompok atau *cluster* dalam data (Larose & Larose, 2015). Tujuan pengelompokan data adalah mengoptimalkan fungsi objektif yang ditetapkan dalam proses pengelompokan, dengan fokus pada pengurangan variasi dalam kelompok yang sama dan peningkatan variasi antara kelompok (Sulistiyawati & Supriyanto, 2021). Berikut merupakan langkah-langkah dalam menggunakan algoritma *K-means* (Sulistiyawati & Supriyanto, 2021).

1. Mengajukan pertanyaan kepada pengguna untuk menentukan jumlah kelompok (k) yang akan digunakan dalam partisi data.
2. Menetapkan secara acak k rekaman sebagai lokasi awal pusat kelompok.
3. Dalam algoritma *K-means*, langkah penting adalah menghitung jarak antara setiap data input dengan setiap *centroid* menggunakan rumus jarak Euclidean (Euclidean Distance). Proses ini dilakukan untuk mencari *centroid* yang paling dekat dengan setiap data. Dalam Euclidean Distance, kita mengukur jarak antara dua titik dengan menghitung akar kuadrat dari jumlah kuadrat perbedaan koordinatnya. Dengan menggunakan persamaan Euclidean Distance, kita dapat menentukan *centroid* terdekat untuk setiap data input.

$$De = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (2.5)$$

Keterangan:

De = Euclidean Distance

i = Banyaknya objek

(x,y) = Koordinat object

(s,t) = Koordinat *centroid*

4. Untuk setiap kelompok k, mencari *centroid* kelompok dan memperbarui lokasi setiap pusat kelompok dengan nilai baru dari *centroid*.

5. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan centroid. Data akan diklasifikasikan ke dalam kelompok yang memiliki jarak terkecil dengan centroid.

$$v_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (2.6)$$

Keterangan:

$v_{ij}$  = Centroid/ rata-rata *cluster* ke-i untuk variable ke-j

$N_i$  = Jumlah data yang menjadi anggota *cluster* ke-i

i,k = Indeks dari *cluster*

j = Indeks dari variabel

$x_{kj}$  = Nilai data ke-k yang ada di dalam *cluster* tersebut untuk variable ke-j

6. Mengulangi langkah 3-5 hingga mencapai konvergensi atau berhenti.

Algoritma *K-means* telah menjadi salah satu algoritma *clustering* yang populer dalam bidang analisis data. Ini memiliki beberapa keunggulan yang membuatnya diminati oleh para peneliti dan praktisi. Berikut adalah beberapa alasan di balik popularitas algoritma *K-means* (Kantardzic, 2020):

1. Algoritma ini memiliki kompleksitas waktu yang efisien, terutama ketika jumlah sampel, jumlah klaster, dan jumlah iterasi telah ditentukan sebelumnya.
2. Ruang yang dibutuhkan oleh algoritma ini juga relatif kecil, sehingga jika data dapat disimpan di memori utama, akses ke elemen-elemen data dapat dilakukan dengan cepat.
3. Salah satu kelebihan utama algoritma *K-means* adalah independen terhadap urutan data. Artinya, walaupun urutan sampel yang diberikan kepada algoritma berbeda, hasil partisi data yang dihasilkan akan tetap sama.

Dalam penelitian berjudul “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau” yang di susun oleh Kamila et al. (2019). Berdasarkan hasil penelitian yang membandingkan efisiensi antara algoritma K-Means dan K-Medoids, ditemukan bahwa algoritma K-Means memiliki waktu pemrosesan yang jauh lebih cepat daripada K-Medoids. Dengan menggunakan data berjumlah 3089 baris, algoritma K-Means hanya memerlukan waktu rata-rata 1 detik untuk menyelesaikan proses clustering, sedangkan K-Medoids memerlukan waktu rata-rata 1 menit 38 detik. Oleh karena itu, dapat disimpulkan bahwa dalam hal kecepatan pemrosesan data, algoritma K-Means lebih efisien dibandingkan dengan K-Medoids.

### 2.2.8 Metode Elbow

Metode siku (*Elbow method*) adalah sebuah metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah kluster terbaik dengan melihat persentase perbandingan antara jumlah kluster yang akan membentuk siku pada suatu titik (Nainggolan et al., 2019). Pada awalnya, penambahan kluster baru cenderung mengurangi nilai SSE (*Sum of Squared Errors*) secara signifikan, tetapi pada suatu titik, penurunan SSE akan menjadi lebih lambat dan menghasilkan sudut tajam pada grafik *elbow* (Bholowalia & Kumar, 2014). Titik di mana penurunan SSE menjadi lambat adalah titik di mana "*elbow*" terbentuk, dan itulah titik yang digunakan untuk memilih jumlah kluster yang optimal. *Sum of Square Error* (SSE) adalah formula yang digunakan untuk mengukur perbedaan antara data yang diperoleh oleh model prediksi yang telah dilakukan sebelumnya (Nainggolan et al., 2019). Rumus SSE adalah sebagai berikut (Irwanto et al., 2012):

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} \|x_i - c_k\|_2^2 \quad (2.7)$$

Keterangan:

$k$  = Cluster

$x_i$  = data ke-I dalam cluster

$c_k$  = nilai rata-rata k cluster

Langkah untuk melakukan metode siku, sebagai berikut (Kodinariya & Makwana, 2013):

1. Inisialisasi Nilai K: Langkah pertama adalah memulai dengan menginisialisasi nilai K sebagai 2.
2. Penambahan Nilai K: Lanjutkan dengan menambahkan nilai K sebanyak 1 pada setiap langkah. Ini berarti kita akan menghitung kluster dan biaya terkait untuk K=3, K=4, dan seterusnya.
3. Pengukuran Kluster dan Biaya: Untuk setiap nilai K yang diuji, lakukan pengelompokan data ke dalam kluster dan hitung biaya yang terkait dengan proses pelatihan. Biaya ini umumnya diukur menggunakan metrik seperti *Sum of Squared Errors* (SSE), yang mengukur seberapa jauh titik data dari pusat klusternya.
4. Pemantauan Perubahan Biaya: Saat K meningkat, amati perubahan dalam biaya. Pada suatu titik, Anda akan melihat bahwa biaya menurun secara signifikan.
5. Titik Siku: Prinsip utama metode siku adalah bahwa pada titik di mana biaya menurun secara dramatis, Anda akan melihat sebuah "siku" dalam grafik perubahan biaya. Ini

menunjukkan titik di mana penambahan lebih banyak kluster tidak lagi memberikan pengurangan biaya yang signifikan.

6. Fase Plateau: Setelah mencapai titik siku, grafik biaya biasanya mencapai fase plateau. Ini berarti penurunan biaya menjadi lebih lambat saat jumlah kluster terus ditingkatkan.
7. Identifikasi K Optimal: Nilai K yang sesuai dengan titik siku sering dianggap sebagai jumlah kluster yang optimal. Ini mencerminkan keseimbangan antara kemampuan menggambarkan variasi dalam data dan risiko overfitting di mana kluster terlalu mendekati titik data individual.
8. Visualisasi: Proses ini dapat divisualisasikan dalam bentuk grafik biaya (SSE atau metrik serupa) terhadap berbagai nilai K. Titik siku pada grafik ini akan menandai titik di mana penambahan kluster lebih lanjut menghasilkan pengurangan biaya yang melambat.

### 2.2.9 Silhouette Score

Analisis *Silhouette Score* digunakan untuk mempelajari dan memahami jarak pemisahan antara kluster yang dihasilkan, dengan mengukur seberapa dekat setiap objek dalam satu kluster terhadap objek lain dalam kluster lain (Ogbuabor & Ugwoke, 2018). Nilai skor Silhouette berkisar antara -1 hingga +1, nilai tersebut dapat diinterpretasikan sebagai berikut (Shahapure & Nicholas, 2020):

1. *Silhouette Score* dengan nilai mendekati +1 menunjukkan bahwa titik data berada dalam kluster yang benar.
2. *Silhouette Score* dengan nilai mendekati 0 menunjukkan bahwa titik data mungkin seharusnya berada dalam kluster lain.
3. *Silhouette Score* dengan nilai mendekati -1 menunjukkan bahwa titik data berada dalam kluster yang salah.

Berikut tahapan dari menghitung nilai *silhouette score* (Dhewayani et al., 2022):

1. Untuk setiap objek data  $i$  dalam kluster tertentu, hitung jarak rata-rata antara objek data  $i$  dengan semua objek data lain yang berada dalam kluster yang sama.
2. Untuk setiap objek data  $i$  dalam kluster tertentu, hitung rata-rata jarak antara objek data  $i$  dengan semua objek data dalam kluster lain. Pilih nilai rata-rata terkecil dari langkah ini dan tetapkan sebagai  $b_i$ .
3. Hitung nilai skor siluet untuk objek data  $i$  dengan menggunakan rumus.

$$S_i = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2.7)$$



4. Ulangi langkah 1-3 untuk semua objek data dalam kluster, dan kemudian hitung rata-rata dari semua nilai skor siluet yang dihasilkan untuk mendapatkan nilai skor siluet keseluruhan.

### 2.2.10 Python

*Python* adalah bahasa pemrograman yang dijalankan dengan cara menginterpretasikan baris per baris kode, dengan sintaks yang mudah dimengerti dan sering diibaratkan sebagai pseudokode yang bisa langsung dieksekusi (Oliphant, 2007). Bahasa ini dikembangkan oleh Guido van Rossum pada awal tahun 1990-an (Lindstrom, 2005). Meskipun awalnya banyak digunakan dalam lingkungan akademis, Python juga semakin populer digunakan dalam industri (Pedregosa et al., 2011). Alasan penggunaan *python* menurut Nagpal & Gabrani (2019), sebagai berikut:

1. Lebih Sedikit Kode dengan Tingkat Keterbacaan Tinggi

*Python* melibatkan penggunaan kode yang sangat sedikit dalam hal baris kode dibandingkan dengan banyak bahasa pemrograman lain, yang dapat digunakan untuk mengembangkan banyak aplikasi. Fitur ini memungkinkan pengujian yang mudah sehingga para pengembang dapat lebih fokus pada pemrograman sebenarnya. *Python* menggunakan hingga 1/5 kode dari bahasa OOP (*Object-Oriented Programming*) lainnya untuk mengimplementasikan logika yang sama.

2. Portabilitas dan Fleksibilitas

*Python* yang fleksibel menawarkan pilihan bagi pengembang untuk memilih menggunakan pendekatan OOP atau melakukan *scripting*. Oleh karena itu *python* cocok untuk pengembang untuk banyak tujuan. *Python* dapat digunakan untuk menghubungkan berbagai struktur data dan bisa digunakan sebagai bahasa *backend*. Sebagian besar kodenya diperiksa dalam IDE (*Integrated Development Environment*). Ini sangat membantu pengembang dalam menulis berbagai algoritma di berbagai domain seperti AI, jaringan saraf, *deep learning*, dan algoritma jaringan komputer.

3. Platform Independen

*Python* memberikan fleksibilitas kepada pengembang untuk menyediakan API dari bahasa pemrograman saat ini yang ternyata sangat fleksibel bagi para pengembang Python baru. Pengembang dapat mengubah kode sumber (*source code*) proyek mereka dengan cara yang mudah untuk menjalankan proyek mereka di OS yang berbeda, sehingga menghemat banyak waktu dan pekerjaan.

#### 4. Keseimbangan Pemrograman Level Rendah dan Tinggi

*Python* memiliki kemampuan untuk menyeimbangkan pemrograman level tinggi dengan level rendah. Ini adalah salah satu fitur paling penting dalam *Python*. Ini memberikan kinerja tinggi pada objek level lebih tinggi seperti arrays dan matriks. Salah satu contohnya adalah vektorisasi dalam algoritma. Ini membuatnya memungkinkan untuk bekerja dengan seluruh *array* daripada angka tunggal.

#### 5. Kontinuitas

*Python* hampir menjadi kandidat ideal sebagai bahasa pemrograman pertama. Manfaat lainnya adalah fakta bahwa itu tidak hanya dibuat untuk tujuan pendidikan tetapi juga untuk digunakan dalam kehidupan profesional mereka di masa depan. Ini digunakan misalnya dalam administrasi jaringan, pengembangan web, pemrograman permainan komputer, dan banyak program memiliki dukungan terintegrasi untuk skrip *Python*.

#### 6. Struktur Data

*Python* memiliki set, daftar, kamus, tuple, antrian yang aman dari thread, string, dan lain-lain. Penting bagi seorang pengembang untuk menggunakan struktur data yang tepat untuk suatu algoritma. Hal ini diperlukan terutama dalam bidang pengembangan kode yang berorientasi pada penelitian. Ketidakhadiran fitur-fitur tersebut membuat seorang pengembang tidak dapat mempelajari dan menggunakan pola desain yang baik.

#### 7. Tersedia *library Open Source*

*libraries* yang didedikasikan membantu menghemat waktu pengembang dari pengkodean algoritma dasar. Selain itu, ada komunitas pengembang *Python* yang besar yang bersedia membantu pengembang *Python* lain dalam berbagai tahap siklus pengembangan mereka.

Salah satu keunggulan dari bahasa pemrograman *Python* adalah menyediakan perpustakaan (*library*) yang kaya dengan modul dan paket-paket untuk digunakan dalam komputasi ilmiah dan pembelajaran mesin (Sodhi et al., 2019). *Libraries* yang akan digunakan peneliti dalam menerapkan algoritma *K-Means* dan *FP-Growth* adalah *Scikit-learn* dan *mlxtend*.

*Scikit-learn* adalah proyek *software open source* yang bertujuan untuk membuat pembelajaran mesin (*machine learning*) dapat diakses oleh semua orang, baik itu di lingkungan akademis maupun industri (Varoquaux et al., 2015). *Scikit-learn* terbukti menjadi pustaka yang unggul dengan dokumentasi yang sangat baik dan intuitif, serta menyediakan berbagai contoh yang mendukung (Stančin & Jović, 2019). Salah satu algoritma yang dapat dijalankan melalui

*library* tersebut adalah *K-Means Clustering*. *library* ini berbeda dari beberapa *library* yang ada di *python*, karena (Pedregosa et al., 2011):

1. Didistribusikan dengan lisensi BSD.
2. Menggabungkan kode terkompilasi untuk efisiensi, tidak seperti MDP dan *pybrain* yang memerlukan kompilasi saat runtime.
3. Hanya bergantung pada *numpy* dan *scipy* untuk memudahkan distribusi, tidak seperti *pymvpa* yang memiliki dependensi opsional seperti *R* dan *shogun*.
4. Lebih fokus pada pemrograman imperatif, tidak seperti *pybrain* yang menggunakan kerangka kerja aliran data.

*Scikit-learn* dapat digunakan dalam berbagai studi kasus, contohnya dalam jurnal “Analysis of electricity consumption at home using K-means clustering algorithm” yang ditulis oleh Choi et al. (2019). Penelitian tersebut menggunakan *scikit-learn* untuk menjalankan algoritma *K-Means*. Tujuan dari penelitian tersebut adalah menganalisis hasil klusterisasi dari penggunaan listrik di rumah. Hasil dari penelitian tersebut algoritma *K-Means* pada *library Scikit-learn* dapat menjalankan algoritma *K-means* dengan baik sehingga dapat untuk menghasilkan analisis dari cluster yang dibentuk.

*MLxtend* adalah sebuah perpustakaan yang mengimplementasikan berbagai algoritma inti dan utilitas untuk pembelajaran mesin dan penambangan data (Raschka, 2018). *mlxtend* merupakan satu-satunya *library* yang mengimplementasikan algoritma aturan asosiasi dan pembelajaran ensemble stacking (Stančin & Jović, 2019). Penerapan dari *MLxtend* untuk menjalankan algoritma aturan asosiasi dapat dilakukan pada berbagai jenis data industri, seperti data rumah sakit, toko ritel, finansial, dan lain-lain. Seperti pada jurnal berjudul “Application of Machine Learning in medical data analysis illustrated with an example of association rules” yang ditulis oleh Butryn et al. (2021), penelitian tersebut bertujuan untuk menemukan hubungan antara karakteristik individu dengan penyakit untuk mengidentifikasi risiko terhadap suatu penyakit. Hasil dari penelitian ini menunjukkan penggunaan *library MLxtend* untuk melakukan aturan asosiasi pada data medis terbukti dapat memberikan hubungan antar variabel.

Selanjutnya, pada jurnal berjudul “Analisis Intensitas Hujan Provinsi Jawa Barat Tahun 2020 Menggunakan Association Rule Apriori dan FP-Growth” yang ditulis oleh Almahenzar & Wijayanto (2022). Jurnal tersebut membandingkan penggunaan aturan asosiasi pada *software* WEKA dan *library MLxtend Python*. Dari jurnal tersebut menunjukkan hasil yang sama antara aturan asosiasi menggunakan *software* WEKA dan *library MLxtend Python*.

## **BAB III**

### **METODE PENELITIAN**

#### **3.1 Objek Penelitian**

Objek pada penelitian ini adalah data transaksi penjualan pada toko Intimart Gedongan yang terletak di Kabupaten Sleman, Daerah Istimewa Yogyakarta. Data transaksi yang digunakan ada data 3 bulan transaksi pada toko Intimart Gedongan. Metode yang akan digunakan adalah metode *AR-MBA (Association Rule-Market Basket Analysis)* menggunakan algoritma *FP-Growth* dan metode *Clustering* menggunakan algoritma *K-means*.

#### **3.2 Metode Pengumpulan Data**

Metode pengumpulan data yang digunakan adalah sebagai berikut:

1. Data Historis Pembelian

Data historis yang digunakan pada penelitian ini adalah data transaksi penjualan pada bulan November 2022, Desember 2022, dan Januari 2023. Data tersebut diambil pada *database* toko Intimart Gedongan.

2. Wawancara

Wawancara dilakukan peneliti sebagai data pendukung dari hasil data atau dapat menjadi pendukung latar belakang dari penelitian yang dilakukan. Wawancara disini dilakukan oleh peneliti langsung kepada pemilik toko.

3. Kajian Literatur

Sumber yang digunakan untuk referensi pada penelitian ini merupakan jurnal-jurnal terdahulu yang memiliki topik penelitian serupa dengan penelitian yang sedang dilakukan.

#### **3.3 Jenis Data**

Berdasarkan sumbernya, data yang digunakan pada penelitian ini dibagi menjadi 2, yaitu:

1. Data Primer

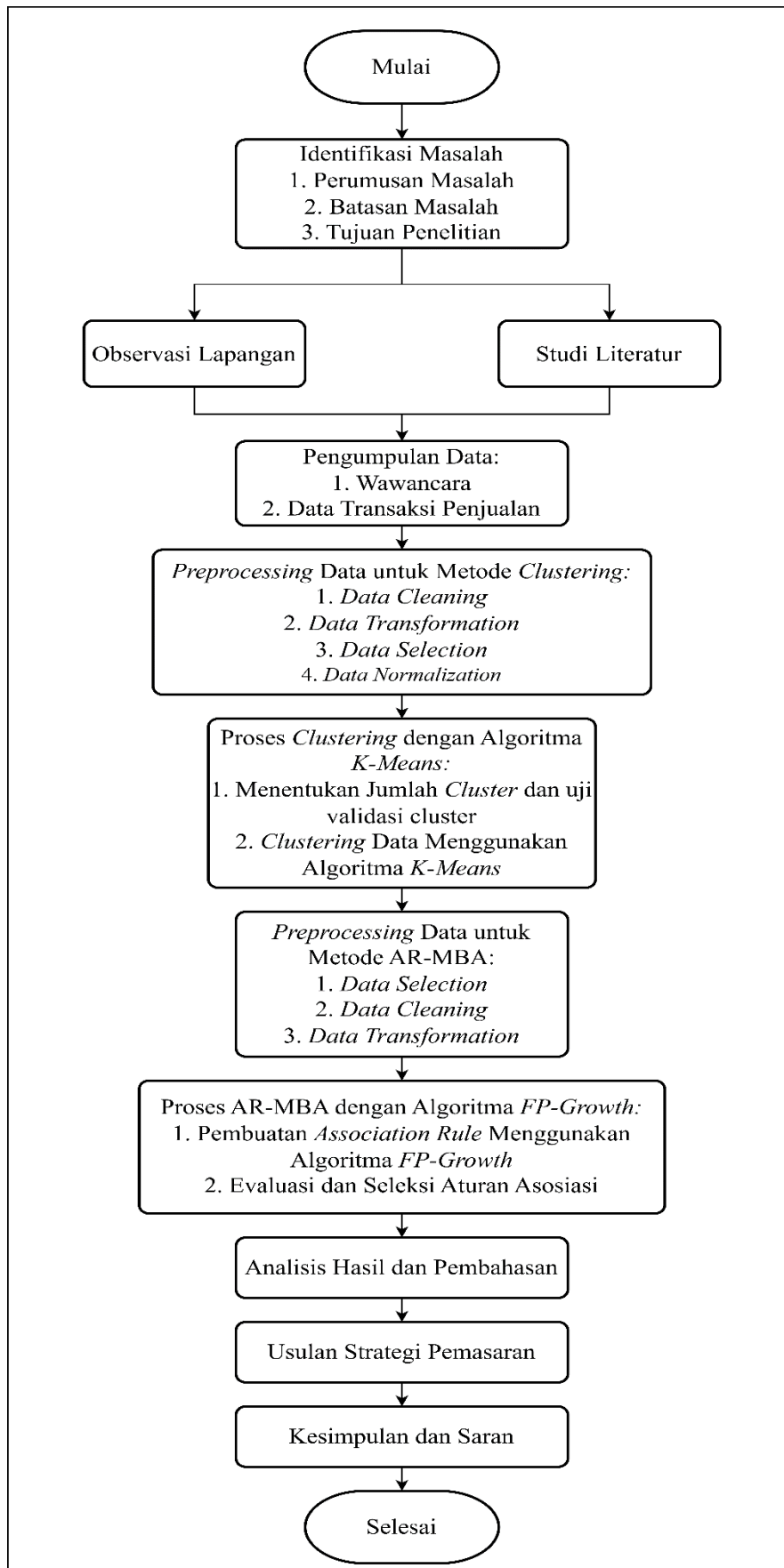
Data primer adalah jenis data yang dikumpulkan pertama kali dan secara langsung oleh peneliti untuk tujuan penelitian tertentu. Dalam konteks penelitian ini, data primer merujuk pada informasi yang diperoleh langsung oleh peneliti melalui proses wawancara. Proses wawancara dengan pemilik toko menjadi sumber informasi yang penting dalam penelitian ini, dan data primer yang dihasilkan dari wawancara tersebut digunakan sebagai landasan utama untuk menganalisis dan mendukung temuan dalam penelitian.

## 2. Data Sekunder

Data sekunder dalam penelitian ini diperoleh melalui pengumpulan informasi yang telah dilakukan oleh pihak lain dari sumber-sumber yang telah ada sebelumnya. Sumber data sekunder utama adalah data transaksi penjualan dari toko Intimart Gedongan. Data ini diambil dari database toko Intimart Gedongan dan mencakup periode November 2022 hingga Januari 2023. Selain itu, data sekunder juga diperoleh dari referensi jurnal dan buku yang relevan dengan topik penelitian. Data sekunder tersebut merupakan sumber informasi tambahan yang digunakan untuk mendukung analisis dan temuan dalam penelitian ini.

### **3.4 Alur Penelitian**

Langkah kerja pada penelitian ini diketahui dengan diagram alir. Berikut merupakan alur penelitian ini:



Gambar 3. 1 Alur Penelitian

Berikut merupakan penjelasan dari alur penelitian diatas:

#### 1. Mulai

Proses awal dalam penelitian ini dimulai dengan melakukan observasi langsung terhadap kondisi lapangan dan mencari permasalahan yang dapat menjadi dasar penelitian. Hal ini akan membentuk latar belakang penelitian yang relevan dengan tujuan yang ingin dicapai.

#### 2. Identifikasi Masalah

Tahapan identifikasi masalah dilakukan untuk mencari dasar permasalahan dari penelitian yang dilakukan. Berdasarkan rumusan masalah dapat ditetapkan tujuan dan batasan pada penelitian. Pada penelitian ini peneliti mengidentifikasi permasalahan yang ada pada *sales*/total pendapatan yang ada di toko Intimart Gedongan. Permasalahan yang ditemukan dapat dilanjutkan untuk dilakukan analisis dan ditemukan solusinya.

#### 3. Observasi Lapangan

Pada tahap observasi lapangan, peneliti secara langsung mengamati dan memperhatikan kondisi di mana penelitian dilakukan. Dengan melakukan observasi, peneliti dapat memperoleh informasi secara detail mengenai beberapa aspek yang relevan dengan penelitian, seperti tempat penelitian, lingkungan sekitar, dan praktik-praktik yang dilakukan oleh subjek penelitian.

#### 4. Studi Literatur

Studi literatur memainkan peran penting dalam proses penelitian. Melalui studi literatur, peneliti mengumpulkan referensi dan teori-teori yang relevan dengan topik penelitian, memperoleh pemahaman mendalam tentang isu-isu yang terkait, konsep-konsep teoritis yang terkait, dan penelitian terdahulu yang telah dilakukan. Referensi dan teori-teori ini menjadi pondasi yang kokoh dan mendukung penelitian, membantu dalam pembahasan dan analisis selanjutnya. Dalam penelitian ini, peneliti akan menggunakan metode *AR-MBA* (*Association Rule-Market Basket Analysis*) dengan algoritma *FP-Growth* untuk mengidentifikasi pola asosiasi antara produk dalam data transaksi penjualan. Selain itu, metode *clustering* dengan algoritma *K-means* akan digunakan untuk mengelompokkan produk atau pelanggan berdasarkan kesamaan karakteristik.

#### 5. Pengumpulan Data

Pada tahap pengumpulan data dalam alur penelitian ini, peneliti melakukan akuisisi data historis yang relevan dengan penelitian. Data historis yang digunakan dalam penelitian ini adalah data transaksi penjualan pada bulan November 2022, Desember 2022, dan Januari 2023. Data ini merupakan hasil dari transaksi penjualan yang tercatat di *database* toko Intimart Gedongan.

## 6. *Preprocessing Data* untuk Metode *Clustering*

Tahapan *preprocessing data* untuk metode *clustering* dalam penelitian skripsi, yang akan digunakan untuk melakukan analisis, memiliki peran penting dalam mempersiapkan data sebelum dilakukan analisis *clustering*. Pada tahap ini, peneliti menggunakan software *VSCode* dengan menggunakan bahasa pemrograman *Python*. Pada tahapan ini, peneliti akan melakukan *data selection*, *data cleaning*, *data transformation*, dan normalisasi data.

- Tahap *data cleaning* bertujuan untuk membersihkan data dari *noise*, *outliers*, atau nilai yang hilang. Langkah-langkah yang dapat dilakukan termasuk identifikasi dan penghapusan data duplikat, pengelolaan nilai yang hilang dengan mengisi atau menghapusnya, serta deteksi dan penanganan *outliers* yang tidak sesuai.
- Tahap *data transformation*, data akan diubah agar sesuai dengan kebutuhan analisis *clustering*. Contoh dari tahap *data transformation* dalam *clustering* adalah dengan menambahkan kolom "waktu" yang berisi informasi pada jam berapa transaksi dilakukan. Kolom ini dapat diekstraksi dari kolom "JAM" pada dataset utama. Misalnya, dataset utama memiliki kolom "JAM" yang berisi informasi waktu transaksi dalam format "HH:MM:SS". Untuk melaksanakan tahap data transformation, kita dapat mengekstrak bagian jam dari kolom "JAM" dan menambahkannya sebagai kolom baru dengan nama "waktu".
- Pada tahap *data selection*, peneliti akan memilih variabel atau atribut yang relevan untuk analisis *clustering*. Hal ini melibatkan pemilihan subset data yang akan digunakan dalam proses *clustering*, seperti memilih atribut yang memiliki informasi yang penting dan relevan.
- Tahap normalisasi data bertujuan untuk menyamakan skala atribut, sehingga atribut dengan skala yang berbeda tidak mendominasi dalam proses *clustering*. Pada penelitian ini, peneliti menggunakan metode normalisasi yang umum digunakan *Z-Score Scaling*.

Tujuan dari *preprocessing* ini adalah untuk memastikan kualitas data yang optimal dan mempersiapkannya sesuai dengan kebutuhan analisis *clustering*.

## 7. Proses *Clustering* dengan Algoritma *K-means*

Dalam tahapan ini, peneliti menggunakan algoritma *K-means* untuk mengelompokkan data ke dalam kelompok-kelompok homogen berdasarkan kemiripan fitur. Pada tahap ini, peneliti menggunakan software *VSCode* dengan menggunakan bahasa pemrograman *Python*.



- Pada tahap menentukan jumlah *cluster* dan uji validitas *cluster*, peneliti perlu menentukan jumlah *cluster* yang diinginkan. Hal ini dapat dilakukan berdasarkan pengetahuan domain, tujuan analisis, atau dengan menggunakan metode seperti *elbow method* atau *silhouette method* untuk menemukan jumlah *cluster* yang optimal.
- Pada tahap *clustering data* menggunakan algoritma *K-means*, Setelah jumlah *cluster* ditentukan, algoritma *K-means* digunakan untuk melakukan *clustering* data. Algoritma ini membagi data menjadi *k cluster* dengan mencari pusat *cluster (centroid)* yang mewakili kelompok data. Data akan dikelompokkan berdasarkan jarak *euclidean* antara setiap data dengan *centroid* terdekat.

Hasil dari proses *clustering* ini akan digunakan untuk analisis lebih lanjut, seperti pembuatan aturan asosiasi yang relevan dalam penelitian skripsi.

#### 8. *Preprocessing Data* untuk Metode *AR-MBA*

Pada tahap ini, peneliti menggunakan *software VSCode* dengan menggunakan bahasa pemrograman *Python* untuk melakukan *Preprocessing* data. Tahap ini melibatkan beberapa langkah penting untuk mempersiapkan data sebelum dilakukan analisis menggunakan metode *AR-MBA (Association Rule-Market Basket Analysis)*.

- Pada tahap *data selection*, peneliti akan memilih asosiasi yang sesuai dengan hasil analisis *cluster* sebelumnya. Asosiasi ini akan digunakan untuk mempelajari pola hubungan antara item-item yang sering muncul bersama dalam transaksi.
- Tahap *data cleaning* dilakukan untuk membersihkan data transaksi. Salah satu langkah penting dalam tahap ini adalah menghapus transaksi yang hanya memiliki 1 item belanja. Hal ini dilakukan untuk memfokuskan analisis pada transaksi yang memiliki *multiple-item*, yang cenderung lebih relevan dalam pencarian pola asosiasi.
- Tahap *data transformation* dilakukan dengan menggabungkan ID transaksi yang sama dalam dataset. Tujuan dari langkah ini adalah untuk memastikan bahwa setiap ID transaksi hanya mewakili satu transaksi yang unik dalam dataset. Tahap ini juga melibatkan perubahan data transaksi menjadi format Boolean. Dalam format ini, setiap item akan direpresentasikan sebagai nilai *True* atau *False* tergantung pada keberadaannya dalam transaksi. Pemetaan ini memungkinkan analisis yang lebih efisien dan mudah terhadap pola asosiasi.

#### 9. Proses *AR-MBA* dengan Algoritma *FP-Growth*

Pada tahap ini, algoritma *FP-Growth* digunakan untuk menganalisis pola asosiasi antara produk dalam data transaksi penjualan. Pada tahap ini, peneliti menggunakan software *VSCode* dengan menggunakan bahasa pemrograman *Python*.

- Pada tahap pembuatan *Association Rule* Menggunakan Algoritma *FP-Growth*, peneliti akan menggunakan algoritma *FP-Growth* untuk melakukan pembuatan aturan asosiasi. Algoritma ini memungkinkan pencarian aturan asosiasi secara efisien tanpa harus menghasilkan kandidat itemsets, sehingga mengurangi kebutuhan memori. Dalam proses ini, data akan diproses menggunakan algoritma *FP-Growth* untuk mengidentifikasi *itemsets* yang sering muncul bersama dalam transaksi.
- Evaluasi dan Seleksi Aturan Asosiasi: Setelah pembuatan aturan asosiasi, peneliti akan melakukan evaluasi dan seleksi aturan untuk memilih aturan yang paling relevan dan bermakna. Evaluasi dapat dilakukan dengan menggunakan metrik seperti *support*, *confidence*, dan *lift*. Selain itu, peneliti juga dapat mempertimbangkan faktor bisnis dan kebutuhan analisis untuk memilih aturan yang paling sesuai. Tujuan dari tahap ini adalah untuk mendapatkan aturan asosiasi yang memiliki nilai *support* dan *confidence* yang tinggi serta relevan untuk kebutuhan analisis yang dilakukan.

Melalui tahap ini, peneliti dapat memperoleh pemahaman yang mendalam mengenai pola penjualan yang relevan dan mengembangkan strategi peningkatan penjualan yang efektif.

#### 10. Analisis Hasil dan Pembahasan

Dalam tahapan ini, peneliti melakukan analisis hasil dan pembahasan yang fokus pada dua metode utama, yaitu metode *clustering* dan metode *AR-MBA*. Peneliti menganalisis data yang telah diproses menggunakan metode *clustering*, seperti algoritma *K-means*, untuk mengelompokkan data penjualan ke dalam kelompok-kelompok yang homogen. Hasil *clustering* ini digunakan untuk mengidentifikasi pola penjualan yang tersembunyi dan menggambarkan karakteristik setiap kelompok secara rinci. Selain itu, peneliti juga menerapkan metode *AR-MBA*, terutama dengan menggunakan algoritma *FP-Growth*, untuk mengungkapkan asosiasi produk yang kuat dan mengidentifikasi itemset yang sering muncul bersama dalam transaksi. Pembahasan hasil analisis ini memberikan pemahaman yang lebih mendalam tentang perilaku penjualan, hubungan antara produk, serta strategi peningkatan penjualan yang dapat diambil berdasarkan temuan yang signifikan dari metode *clustering* dan *AR-MBA*.

#### 11. Usulan Strategi Pemasaran

Dalam tahap usulan strategi pemasaran, peneliti mengusulkan strategi kepada pihak Toko Intimart Gedongan. Strategi yang diberikan berasal dari metode *clustering* dan *AR-MBA* yang sebelumnya telah dilakukan. Peneliti memberikan usulan strategi ini karena analisis yang dilakukan dari kedua metode tersebut memberikan wawasan yang sangat berharga mengenai perilaku pelanggan dan pola pembelian produk di toko. Dengan memanfaatkan informasi yang diperoleh dari hasil *clustering*, toko dapat menyusun strategi pemasaran yang lebih tepat sasaran dan sesuai dengan preferensi masing-masing kelompok pelanggan. Selain itu, dengan menerapkan aturan asosiasi dari metode *AR-MBA*, toko dapat menciptakan paket bundling produk yang lebih menarik bagi pelanggan, serta mengoptimalkan promosi dan penjualan produk yang memiliki hubungan asosiasi kuat.

#### 12. Kesimpulan dan Saran

Dalam tahapan ini, peneliti menyajikan kesimpulan utama dari penelitian dan memberikan saran-saran yang relevan berdasarkan temuan dan analisis yang telah dilakukan. Kesimpulan disusun berdasarkan temuan-temuan signifikan dari analisis hasil dan pembahasan, menghubungkannya dengan tujuan penelitian, dan menjawab rumusan masalah yang telah disusun di awal penelitian.

#### 13. Selesai

## BAB IV

### PENGUMPULAN DAN PENGOLAHAN DATA

#### 4.1 Pengumpulan Data

Pengumpulan data adalah proses mengumpulkan informasi atau fakta-fakta yang relevan dengan penelitian atau studi yang dilakukan. Hal ini bertujuan untuk mendapatkan pemahaman yang lebih baik tentang fenomena yang diteliti, menganalisis hubungan antara variabel, dan mendukung pembuatan kesimpulan yang valid. Pada penelitian ini, pengumpulan data dilakukan di Minimarket Intimart Gedongan sebagai fokus penelitiannya. Toko minimarket Intimart Gedongan berupaya melakukan perubahan secara terus-menerus guna menghadapi berbagai tantangan yang ada. Intimart Gedongan merupakan salah satu minimarket yang beroperasi di bawah pengelolaan CV Intimart. CV Intimart telah berdiri sejak tahun 2018 dan memiliki lebih dari 10 cabang minimarket yang tersebar di Yogyakarta dan Solo. Lokasi toko Intimart Gedongan berada di Jalan Selokan Mataram, Gedongan, Sinduadi, Kecamatan Mlati, Kabupaten Sleman, Daerah Istimewa Yogyakarta.

Pada penelitian ini menggunakan data historis perusahaan, yaitu data historis penjualan selama 3 bulan dimulai dari November 2022, Desember 2022, dan Januari 2023. Pengumpulan data diambil pada *database* dari toko Intimart Gedongan. Data yang dikumpulkan berisi 8 kolom dan 65536 baris. Kolom pada data tersebut adalah.

Tabel 4. 1 Nama Kolom dan Penjelasannya

Nama Kolom	Keterangan
KODE_BARANG	Kode unik untuk mengidentifikasi barang dalam toko atau ritel.
NAMA_BARANG	Nama barang yang dijual dalam toko atau ritel.
JUMLAH	Jumlah barang yang terjual
HARGA_JUAL	Harga jual per unit dari barang
TOTAL	Total harga atau nilai keseluruhan dari barang yang terjual
KODE_FAKTUR	Kode unik untuk mengidentifikasi faktur atau transaksi penjualan
TANGGAL_FAKTUR	Tanggal transaksi penjualan atau tanggal pada faktur
JAM	Waktu transaksi penjualan

## 4.2 Pengolahan Data

### 4.2.1 Data Transaksi

Penelitiann yang dilakukan menggunakan data transaksi toko Intimart Gedongan selama 3 bulan yaitu pada bulan November 2022, Desember 2022, dan Januari 2023. Berikut merupakan contoh 5 data teratas transaksi pada toko Intimart Gedongan.

Tabel 4. 2 Data Transaksi Toko Intimart Gedongan

Kode Barang	Nama Barang	Jumlah	Harga Jual	Total	Kode Faktur	Tanggal Faktur	Jam
0104050039	AMO AIR MINERAL 600ML	1	2000	2000	JF00001112 2	11/1/2022	6:01:30
0112020043	GG SURYA 16S	1	27900	27900	JF00002112 2	11/1/2022	6:22:56
0104030033	MOGU2 MELON 320ML	1	9900	9900	JF00002112 2	11/1/2022	6:22:56
0112020115	MARLBO RO KRTK BIRU 12	1	9500	9500	JF00003112 2	11/1/2022	6:28:57
0112020095	ESSE BANA POP 12S	1	20500	20500	JF00003112 2	11/1/2022	6:28:57

Dari tabel diatas diketahui informasi mengenai kode barang, nama barang, jumlah, harga jual, total, kode faktur, tanggal faktur, dan jam. Namun, terdapat problematika dalam dataset tersebut di mana kode faktur belum digabungkan dalam satu transaksi dan masih terpisah-pisah antara setiap barang yang dibeli dalam satu transaksi. Oleh karena itu, diperlukan proses *preprocessing data* untuk menggabungkan kode faktur yang terkait dengan transaksi yang sama.

#### 4.2.2 Preprocessing Dataset untuk Melakukan Clustering

Tahapan preprocessing data dalam metode *clustering* pada penelitian skripsi memiliki peran penting dalam persiapan data sebelum dilakukan analisis *clustering*. Peneliti menggunakan perangkat lunak VSCode dan bahasa pemrograman Python untuk melaksanakan tahap ini. Tahapan preprocessing data mencakup data selection, data cleaning, data transformation, dan normalisasi data. Proses ini bertujuan untuk mengoptimalkan data sebelum dianalisis dengan algoritma *clustering*.

##### 4.2.2.1 Data Cleaning pada Preprocessing Metode Clustering

Pada tahap data cleaning dalam Preprocessing Metode *Clustering*, peneliti akan memeriksa adanya data kosong pada dataset. Berikut ini adalah daftar data kosong dari setiap kolom atau atribut:

```
Missing Values of df_0:
KODE_BARANG      0
NAMA_BARANG     12
JUMLAH           0
HARGA_JUAL       0
TOTAL            0
KODE_FAKTUR      0
TANGGAL_FAKTUR  0
JAM              0
dtype: int64
```

Gambar 4. 1 Jumlah *Missing Value* pada Data

Pada tahap ini, peneliti akan melakukan pembersihan data transaksi. Dalam data transaksi ini, peneliti akan menghapus kolom "NAMA\_BARANG" yang memiliki nilai NaN. Hal ini dilakukan karena pada kolom "NAMA\_BARANG", terdapat baris yang memiliki nilai NaN, dan saat dianalisis, baris tersebut memiliki nilai nol (0) atau satu (1) pada atribut "JUMLAH", "HARGA\_JUAL", dan "TOTAL". Terdapat pula data yang kosong ini hanya berjumlah 12 baris, yang relatif sedikit jika dibandingkan dengan jumlah total baris dalam data awal. Penghapusan kolom "NAMA\_BARANG" yang memiliki nilai NaN dilakukan untuk menjaga kebersihan dan integritas data transaksi. Berikut ini merupakan jumlah *missing value* pada data.

Tabel 4. 3 Tampilan Data yang Memiliki *Missing Value*

Kode Barang	Nama Barang	Jumlah	Harga Jual	Total	Kode Faktur	Tanggal Faktur	Jam
----------------	----------------	--------	---------------	-------	----------------	-------------------	-----

112020119	NaN	1	0	0	JF03744112	11/11/202	18:38:45
					2	2	
112020119	NaN	0	0	0	JF05556112	11/17/202	11:03:17
					2	2	
112020119	NaN	0	0	0	JF06444112	11/19/202	18:53:12
					2	2	
112020119	NaN	0	0	0	JF07217112	11/21/202	20:08:30
					2	2	
112020119	NaN	1	0	0	JF03744112	11/11/202	18:38:45
					2	2	

Dengan menghapus kolom ini, peneliti dapat memastikan bahwa hanya atribut yang relevan dan memiliki nilai yang valid yang akan digunakan dalam analisis *clustering*. Selain itu, dengan membuang data yang kosong tersebut, peneliti dapat menghindari bias atau kesalahan dalam analisis *clustering* yang dapat muncul karena ketidaklengkapan data.

#### 4.2.2.2 Data Transformation pada Preprocessing Metode Clustering

Pada tahap ini, peneliti melakukan transformasi dataset untuk memastikan data yang diambil sesuai dengan kebutuhan penelitian. Dalam proses data transformation, peneliti melakukan penggabungan pada nilai yang memiliki kode yang sama pada kolom "KODE\_FAKTUR". Tujuan dari penggabungan ini adalah untuk menyatukan kode faktur yang sebelumnya terpisah-pisah dalam beberapa baris, meskipun termasuk dalam satu transaksi atau struk, menjadi satu baris yang berisi daftar nama barang dan total dari satu kode faktur (transaksi). Dengan melakukan penggabungan ini, kolom "NAMA\_BARANG" yang sebelumnya hanya berisi satu barang sekarang menjadi daftar barang yang dibeli dalam transaksi tersebut. Selain itu, kolom "TOTAL" juga berisi total nilai dari satu transaksi. Berikut ini merupakan tabel dari transformasi diatas.

Tabel 4. 4 Data Transformasi Kode Faktur

Kode Faktur	Jam	Nama Barang	Total
JF000010123	6:40:01	['TELUR AY.RAS CURAH GR']	26500
JF000011122	6:01:30	['AMO AIR MINERAL 600ML']	2000

JF000011222	6:38:26	['BONCABE MAKARONI LVL 15']	7975
JF000020123	6:40:41	['NABATI NEXTAR BLUEBERRY 106G', POCKY STRAW STICK 47GR']	17475
JF000021122	6:22:56	['GG SURYA 16S', 'MOGU2 MELON 320ML']	37800

Pada tahap ini, peneliti juga melakukan pengambilan nilai jam dari kolom "Jam" dalam dataset. Tujuan dari langkah ini adalah untuk menggunakan kolom jam sebagai salah satu atribut dalam metode *clustering*. Untuk mencapai hal tersebut, nilai jam pada kolom "Jam" yang awalnya memiliki format "%H:%M:%S" akan diambil hanya bagian "H" (jam) saja, dan nilai tersebut akan disimpan dalam kolom baru yang disebut "waktu". Selain itu, peneliti juga melakukan perhitungan terkait panjang list atau jumlah variasi produk yang terdapat dalam satu transaksi. Hal ini dilakukan dengan menghitung panjang dari setiap list yang terdapat dalam kolom "Nama Barang". Hasil perhitungan tersebut kemudian dimasukkan ke dalam kolom baru yang dinamakan "variasi produk".

Dengan melakukan transformasi data seperti yang dijelaskan di atas, peneliti berhasil mengubah format dan mengambil nilai yang relevan dari kolom "Jam" serta menghitung "Variasi Produk" dalam satu transaksi. Tabel hasil transformasi data ini memberikan informasi yang lebih terstruktur dan memungkinkan penggunaan kolom "waktu" dan " variasi produk" sebagai atribut dalam analisis *clustering*. Berikut merupakan tabel yang menunjukkan perubahan dan transformasi data yang telah dilakukan oleh peneliti.

Tabel 4. 5 Data Transformasi Nama Barang dan Jam Transaksi

Kode Faktur	Nama Barang	Total Transaksi	Variasi Produk	Waktu
JF000010123	['TELUR AY.RAS CURAH GR']	26500	1	6
JF000011122	['AMO AIR MINERAL 600ML']	2000	1	6
JF000011222	['BONCABE MAKARONI LVL 15']	7975	1	6
JF000020123	['NABATI NEXTAR BLUEBERRY 106G',	17475	2	6



	'POCKY STRAW STICK 47GR']			
JF000021122	['GG SURYA 16S', 'MOGU2 MELON 320ML']	37800	2	6

#### 4.2.2.3 Data Selection pada Preprocessing Metode Clustering

Pada tahap ini, peneliti melakukan *Data Selection* dimana peneliti melakukan pemilihan kolom atau atribut yang akan digunakan ketika melakukan *clustering*. Tujuan utama dari tahap ini adalah untuk memilih atribut-atribut yang dianggap relevan dan memiliki potensi dalam mengungkapkan pola dan insight yang berharga dari data transaksi yang akan dianalisis. Disini peneliti akan menggunakan atribut waktu, macam barang, dan total transaksi. Untuk keterangan dari setiap atribut diatas dapat dilihat pada tabel dibawah ini.

Tabel 4. 6 Deskripsi Variabel *Clustering*

<b>Nama Kolom</b>	<b>Keterangan</b>
Variasi Produk	Transformasi kolom "NAMA_BARANG" yang berisi banyaknya daftar barang dalam satu transaksi
Total Transaksi	Total harga yang harus dibayarkan pada satu transaksi diambil dari kolom "TOTAL"
Waktu	Mencakup nilai jam yang diambil dari kolom "JAM"

Pemilihan atribut di atas dilakukan oleh peneliti dengan tujuan untuk mendapatkan pengelompokan perilaku pelanggan berdasarkan tiga atribut yang dipilih. Dalam analisis *clustering*, atribut "waktu" digunakan untuk mengetahui jam-jam ideal untuk melakukan promosi atau kegiatan pemasaran. Informasi ini dapat membantu peneliti dalam menentukan waktu yang tepat untuk menarik minat pelanggan dan meningkatkan efektivitas promosi produk. Selanjutnya, atribut "variasi produk" dipilih untuk mengetahui apakah pada jam-jam tertentu terdapat perbedaan dalam jumlah dan variasi barang yang dibeli oleh pelanggan. Peneliti ingin melihat apakah ada tren tertentu dalam preferensi pelanggan terkait dengan jumlah variasi produk yang dibeli pada waktu-waktu tertentu. Selain itu, melalui atribut "Total Transaksi", peneliti ingin mengetahui seberapa besar jumlah uang yang dibelanjakan oleh pelanggan pada jam-jam dan jenis barang tertentu. Hal ini dapat memberikan wawasan tentang

tingkat pengeluaran pelanggan pada waktu-waktu dan jumlah dari keranjang belanja tertentu, yang penting dalam memahami potensi pendapatan dan kontribusi transaksi terhadap bisnis.

Tabel 4. 7 Tampilan Data *Clustering*

Waktu	Variasi Produk	Total Transaksi
6	1	26500
6	1	2000
6	1	7975
6	2	17475
6	2	37800

Selain melakukan pemilihan atribut berdasarkan preferensi peneliti, peneliti juga mengacu pada beberapa jurnal terkait. Pada penelitian berjudul “Penerapan Algoritma Association Rules Dalam Penentuan Pola Pembelian Berdasarkan Hasil Clustering” yang disusun oleh Octavia et al. (2023). Penelitian tersebut menggunakan variabel “Item dibeli” dan “Item tidak dibeli”. Variabel “Item dibeli” pada penelitian tersebut memiliki definisi yaitu jumlah variasi/kategori buku yang dibeli pada 1 kali transaksi. Variabel tersebut memiliki definisi yang sama dengan variabel yang digunakan oleh penulis yaitu “Variasi Produk”.

Selanjutnya pada penelitian berjudul “Sistem Rekomendasi Penawaran Produk Pada Online Shop Menggunakan K-Means *Clustering*” yang dirancang oleh Naufal et al. (2022). Penelitian tersebut menggunakan 8 variabel untuk melakukan *clustering* pada data transaksi. Variabel yang digunakan meliputi nomor transaksi, bulan transaksi, pelanggan, umur pelanggan, merk, tipe, harga jual, dan harga beli. Dalam penelitian ini, terdapat dua variabel yang sama dengan yang digunakan oleh penulis, yaitu "harga jual" dan "bulan transaksi". Variabel "harga jual" dalam penelitian ini setara dengan variabel "Total Transaksi" yang digunakan oleh penulis. Begitu pula, variabel "bulan transaksi" dalam penelitian ini setara dengan variabel "Waktu" yang digunakan oleh penulis. Namun, terdapat perbedaan dalam pengukuran waktu yang digunakan. Penelitian tersebut menggunakan bulan sebagai satuan waktu, sementara penulis menggunakan jam sebagai satuan waktu dalam variabel "Waktu".

Dengan mempertimbangkan aspek-aspek tersebut, peneliti berharap bahwa analisis *clustering* pada atribut "Waktu", "Variasi Produk", dan "Total Transaksi" dapat memberikan pemahaman yang lebih dalam tentang perilaku pelanggan, preferensi pembelian, dan tingkat pengeluaran yang dapat digunakan untuk mengoptimalkan strategi pemasaran dan peningkatan pendapatan dalam konteks bisnis yang dijalankan oleh minimarket Intimart Gedongan.

#### 4.2.2.4 Data Normalization pada Preprocessing Metode Clustering

Pada tahap normalisasi, peneliti melakukan langkah untuk menyamakan skala atribut agar atribut dengan skala yang berbeda tidak mendominasi dalam proses *clustering*. Hal ini penting karena beberapa atribut dalam dataset, seperti kolom "Variasi Produk" dan "Total Transaksi", memiliki skala yang berbeda, yaitu satuan jumlah dan rupiah. Untuk mengatasi perbedaan skala tersebut, peneliti menggunakan normalisasi dengan metode z-score yang tersedia dalam library *scikit-learn*. Dengan demikian, setiap variabel akan memiliki nilai yang diubah menjadi nilai relatif terhadap mean dan standar deviasi dari atribut tersebut. Hasil normalisasi ini memungkinkan peneliti untuk membandingkan dan menganalisis atribut dengan skala yang seragam. Berikut ini merupakan nilai dari tiap variabel setelah dilakukan normalisasi.

Tabel 4. 8 Tampilan Data Setelah dilakukan Normalisasi

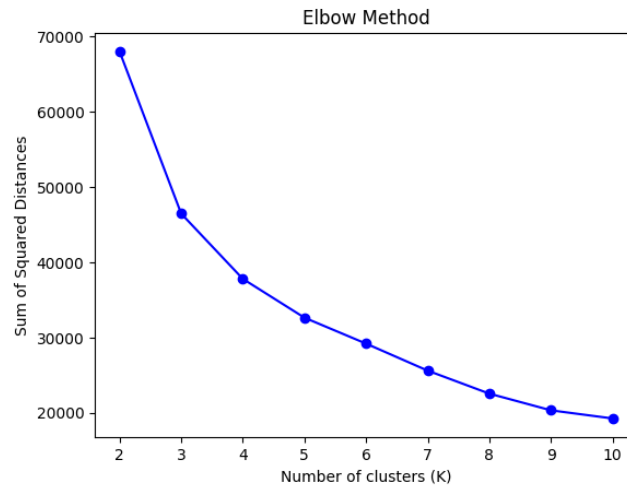
Waktu	Variasi Produk	Total Transaksi
-1.936547	-0.66748	0.005408
-1.936547	-0.66748	-0.84161
-1.936547	-0.66748	-0.635041
-1.936547	-0.035997	-0.306606
-1.936547	-0.035997	0.396074

#### 4.2.3 Proses Clustering dengan K-means

Dalam tahapan ini, peneliti menggunakan algoritma *K-means* untuk mengelompokkan data ke dalam kelompok-kelompok homogen berdasarkan kemiripan fitur. Pada tahap ini, peneliti menggunakan software *VSCode* dan memanfaatkan *library scikit-learn* dalam bahasa pemrograman *Python*.

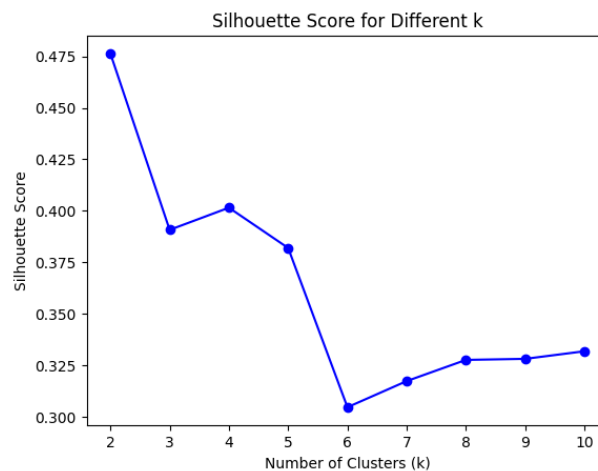
##### 4.2.3.1 Penentuan jumlah Cluster menggunakan Elbow Method dan Silhouette Score

Peneliti menggunakan metode *elbow* dan *silhouette score* untuk menentukan jumlah kluster yang optimal dalam analisis *clustering*. Pada metode *elbow*, peneliti melakukan plot grafik dengan jumlah kluster dari 2 sampai 10. Dalam grafik tersebut, terlihat adanya siku-siku yang terbentuk pada kluster 3 dan kluster 4. Namun, untuk memastikan pilihan kluster yang tepat, peneliti melanjutkan penentuan jumlah kluster menggunakan metode *silhouette score*.



Gambar 4. 2 Grafik *Elbow Method*

Pada tahap *silhouette score*, peneliti melakukan plotting dan memperhatikan bahwa kluster 4 memiliki nilai *silhouette score* tertinggi di antara kluster 3 dan kluster 4. Hal ini menunjukkan bahwa kluster 4 memiliki tingkat kesesuaian data yang lebih baik dibandingkan dengan kluster lainnya. Oleh karena itu, peneliti memilih jumlah kluster sebanyak 4 untuk analisis selanjutnya terkait perilaku pelanggan, preferensi pembelian, dan pola pengeluaran dalam konteks bisnis minimarket Intimart Gedongan.



Gambar 4. 3 Grafik *Silhouette Score*

Dengan menggunakan kombinasi metode *elbow* dan *silhouette score*, peneliti dapat memperoleh hasil yang lebih terpercaya dalam menentukan jumlah kluster yang optimal. Dengan demikian, penggabungan kedua metode tersebut membantu peneliti dalam memilih jumlah kluster yang terbaik untuk analisis *clustering* dalam penelitian ini.

#### 4.2.3.2 Proses Clustering menggunakan Algoritma K-Means

Proses *clustering* algoritma *K-Means* menggunakan *software VSCode* dengan bahasa pemrograman *Python 3.11.4*. Setelah peneliti melakukan proses *clustering* dengan menggunakan jumlah *cluster* sebanyak 4, peneliti menambahkan label hasil *clustering* tersebut ke dalam data transaksinya. Dengan demikian, dapat terlihat hasil dari proses *clustering* pada data transaksi seperti terlihat pada contoh berikut.

Tabel 4. 9 Tampilan Hasil *Cluster* pada Data

ID Transaksi	Waktu	Variasi Produk	Total Transaksi	Kluster
JF031481122	9	1	14500	1
JF096651122	21	2	6000	3
JF083060123	12	5	39000	2
JF034151222	18	2	43700	3
JF022231222	11	6	217795	4

Dalam contoh di atas, kolom "Label" menunjukkan label hasil *clustering* yang diberikan oleh algoritma *K-Means*. Dengan adanya label ini, peneliti dapat memperoleh informasi tambahan tentang kelompok mana tiap transaksi masuk ke dalamnya. Hal ini memungkinkan peneliti untuk menganalisis karakteristik dan pola yang mungkin ada dalam setiap kelompok transaksi, serta mengambil tindakan yang sesuai berdasarkan hasil *clustering* yang didapatkan.

#### 4.2.4 Hasil Clustering K-Means

Analisis perhitungan *clustering K-Means* menggunakan *software VSCode* dengan bahasa pemrograman *Python 3.11.4*. Pada perhitungan pusat *cluster* setelah menjalankan algoritma *K-Means* dapat dilihat seperti dibawah ini.

Tabel 4. 10 Pusat *Cluster* dalam *Z-Score*

Kluster	Waktu (Z-Score)	Variasi Produk (Z-Score)	Total Transaksi (Z-Score)
1	-0.969171	-0.299578	-0.224008
2	0.274084	1.749916	0.793434
3	0.802976	-0.311235	-0.233447
4	-0.17616	1.687596	5.224968

Dilihat dari tabel output pusat *cluster (Z-Score)* diatas, dari hasil perhitungan pusat *cluster* diatas dapat diambil keputusan yaitu sebagai berikut:

- Nilai negatif (-) pada Z-Score: Data berada dibawah rata-rata total variabel.
- Nilai positif (+) pada Z-Score: Data berada diatas rata-rata total variabel.

Dari tabel pusat *cluster* (*Z-Score*) dan ketentuan dari nilai *z-score* diatas maka dapat data dapat dijelaskan sebagai berikut:

1. *Cluster 1*

Pada *cluster 1* terdapat variabel Waktu, Variasi Produk, dan Total Transaksi. Dari 3 variabel tersebut variabel Waktu memiliki nilai dominan dibawah rata-rata total, variabel Variasi Produk memiliki nilai yang dominan dibawah rata-rata total, dan variabel Total Transaksi memiliki nilai yang dominan dibawah rata-rata total.

2. *Cluster 2*

Pada *cluster 2* terdapat variabel Waktu, Variasi Produk, dan Total Transaksi. Dari 3 variabel tersebut variabel Waktu memiliki nilai dominan diatas rata-rata total, variabel Variasi Produk memiliki nilai yang dominan diatas rata-rata total, dan variabel Total Transaksi memiliki nilai yang dominan diatas rata-rata total.

3. *Cluster 3*

Pada *cluster 3* terdapat variabel Waktu, Variasi Produk, dan Total Transaksi. Dari 3 variabel tersebut variabel Waktu memiliki nilai dominan diatas rata-rata total, variabel Variasi Produk memiliki nilai yang dominan dibawah rata-rata total, dan variabel Total Transaksi memiliki nilai yang dominan dibawah rata-rata total.

4. *Cluster 4*

Pada *cluster 4* terdapat variabel Waktu, Variasi Produk, dan Total Transaksi. Dari 3 variabel tersebut variabel Waktu memiliki nilai dominan dibawah rata-rata total, variabel Variasi Produk memiliki nilai yang dominan diatas rata-rata total, dan variabel Total Transaksi memiliki nilai yang dominan diatas rata-rata total.

Setelah mengetahui dan mendeskripsikan value dari nilai pusat *cluster z-score*, peneliti melakukan perhitungan rata-rata, *std. deviation*, Q1, dan Q3 pada tiap-tiap *cluster*. Berikut merupakan tabel rata-rata (pusat *cluster*) dan *std. deviation*:

Tabel 4. 11 Deskripsi Statistik dari Tiap *Cluster*

Kluster	Waktu		Variasi Produk		Total Transaksi	
	Mean	std. deviation	Mean	std. deviation	Mean	std. deviation

1	10.33	2.46	1.58	0.77	19866.89	14767.45
2	15.88	3.88	4.83	1.64	49271.20	23131.71
3	18.25	1.95	1.56	0.72	19591.12	13124.91
4	13.87	4.33	4.73	4.48	177475.84	76439.49

Lalu berikut merupakan tabel dari *lower bound*/batas bawah (Q1) dan *upper bound*/ batas atas (Q3).

Tabel 4. 12 Rentang Persebaran Variabel dari Tiap *Cluster*

Kluster	Waktu		Variasi Produk		Total Transaksi	
	Lower bound (Q1)	Upper bound (Q3)	Lower bound (Q1)	Upper bound (Q3)	Lower bound (Q1)	Upper bound (Q3)
1	8	12	1	2	8925	26450
2	13	19	4	5	31000	63435
3	17	20	1	2	9500	26400
4	10	18	1	7	128050	209884

Dari tabel rata-rata, *std. deviation*, Q1 (*lower bound*), dan Q3 (*upper bound*). Dapat dijelaskan seperti dibawah ini.

1. Pada *Cluster 1*, terdapat 3 variabel yaitu Waktu, Variasi Produk, dan Total Transaksi. Pada “Variabel Waktu”, memiliki nilai rata-rata (pusat *cluster*) sebesar 10.33, nilai *std. deviation* sebesar 2.46, nilai Q1 (*lower bound*) sebesar 8, dan Q3 (*upper bound*) sebesar 12. Kemudian pada variabel “Variasi Produk” memiliki nilai rata-rata (pusat *cluster*) sebesar 1.58, nilai *std. deviation* sebesar 0.77, nilai Q1 (*lower bound*) sebesar 1, dan Q3 (*upper bound*) sebesar 2. Lalu pada variabel “Total Transaksi” memiliki nilai rata-rata (pusat *cluster*) sebesar 19866.89, nilai *std. deviation* sebesar 14767.45, nilai Q1 (*lower bound*) sebesar 8925, dan Q3 (*upper bound*) sebesar 26450.
2. Pada *Cluster 2*, terdapat 3 variabel yaitu Waktu, Variasi Produk, dan Total Transaksi. Pada “Variabel Waktu”, memiliki nilai rata-rata (pusat *cluster*) sebesar 15.88, nilai *std. deviation* sebesar 3.88, nilai Q1 (*lower bound*) sebesar 13, dan Q3 (*upper bound*) sebesar 19. Kemudian pada variabel “Variasi Produk” memiliki nilai rata-rata (pusat *cluster*) sebesar 4.83, nilai *std. deviation* sebesar 1.64, nilai Q1 (*lower bound*) sebesar 4, dan Q3 (*upper bound*) sebesar 5. Lalu pada variabel “Total Transaksi” memiliki nilai rata-rata (pusat *cluster*) sebesar 49271.20, nilai *std. deviation* sebesar 23131.71, nilai Q1 (*lower bound*) sebesar 31000, dan Q3 (*upper bound*) sebesar 63435.

3. Pada *Cluster 3*, terdapat 3 variabel yaitu Waktu, Variasi Produk, dan Total Transaksi. Pada “Variabel Waktu”, memiliki nilai rata-rata (pusat *cluster*) sebesar 18.25, nilai *std. deviation* sebesar 1.95, nilai Q1 (*lower bound*) sebesar 17, dan Q3 (*upper bound*) sebesar 20. Kemudian pada variabel “Variansi Produk” memiliki nilai rata-rata (pusat *cluster*) sebesar 1.56, nilai *std. deviation* sebesar 0.72, nilai Q1 (*lower bound*) sebesar 1, dan Q3 (*upper bound*) sebesar 2. Lalu pada variabel “Total Transaksi” memiliki nilai rata-rata (pusat *cluster*) sebesar 19591.12, nilai *std. deviation* sebesar 13124.91, nilai Q1 (*lower bound*) sebesar 9500, dan Q3 (*upper bound*) sebesar 26400.
4. Pada *Cluster 4*, terdapat 3 variabel yaitu Waktu, Variasi Produk, dan Total Transaksi. Pada “Variabel Waktu”, memiliki nilai rata-rata (pusat *cluster*) sebesar 13.87 nilai *std. deviation* sebesar 4.33, nilai Q1 (*lower bound*) sebesar 10, dan Q3 (*upper bound*) sebesar 18. Kemudian pada variabel “Variansi Produk” memiliki nilai rata-rata (pusat *cluster*) sebesar 4.73, nilai *std. deviation* sebesar 4.48, nilai Q1 (*lower bound*) sebesar 1, dan Q3 (*upper bound*) sebesar 7. Lalu pada variabel “Total Transaksi” memiliki nilai rata-rata (pusat *cluster*) sebesar 177475.84, nilai *std. deviation* sebesar 76439.49, nilai Q1 (*lower bound*) sebesar 128050, dan Q3 (*upper bound*) sebesar 209884.

Dari hasil perhitungan kepada empat *cluster* diatas, *cluster 4* memiliki nilai rata-rata yang paling tinggi untuk variabel total transaksi, yang menunjukkan bahwa pelanggan dalam kelompok ini cenderung melakukan pembelian dengan nilai yang lebih tinggi. *Cluster 2* juga menunjukkan kinerja yang baik dengan rata-rata total transaksi yang cukup tinggi. Sementara itu, *Cluster 1* dan *Cluster 3* memiliki rata-rata total transaksi yang lebih rendah dibandingkan dengan *cluster* lainnya.

Selain itu, pada variabel variasi produk (variasi item yang dibeli) *Cluster 4* memiliki rata-rata variasi produk yang tinggi, menunjukkan bahwa pelanggan dalam kelompok ini cenderung membeli beragam produk. Hal ini menunjukkan minat yang luas dan kemungkinan pelanggan yang lebih eksploratif dalam mencoba produk baru. Sementara itu, *Cluster 1* dan *Cluster 3* memiliki variasi produk yang lebih rendah, mengindikasikan preferensi pembelian yang lebih terfokus pada produk tertentu.

Dalam kesimpulannya, berdasarkan analisis pengelompokan transaksi pelanggan, dapat disimpulkan bahwa *Cluster 4* adalah kelompok dengan kinerja yang paling baik, ditandai dengan total transaksi yang tinggi dan variasi produk yang luas. *Cluster 2* juga menunjukkan kinerja yang baik dengan total transaksi yang cukup tinggi. Sementara *Cluster 1* dan *Cluster 3* memiliki kinerja yang lebih rendah dalam hal total transaksi dan variasi produk.



#### 4.2.5 Preprocessing Dataset untuk Melakukan AR-MBA

Pada tahap ini, peneliti menggunakan *software VSCode* dengan menggunakan *library mlxtend* pada bahasa pemrograman *Python* untuk melakukan *Preprocessing* data. Tahap ini melibatkan beberapa langkah penting untuk mempersiapkan data sebelum dilakukan analisis menggunakan metode *AR-MBA (Association Rule-Market Basket Analysis)*.

##### 4.2.5.1 Data Selection pada Metode AR-MBA

Pada tahap ini, peneliti akan melakukan seleksi data untuk metode AR-MBA. Variabel yang tidak diperlukan dalam proses tersebut, seperti Id Transaksi, Waktu, Variasi Produk, dan lain-lain, akan dihapus. Peneliti hanya akan menggunakan variabel Nama Barang, yang berisi daftar barang yang dibeli dalam satu transaksi. Sebelum mengambil nilai dari variabel tersebut, peneliti akan mengurutkan transaksi sesuai dengan pengelompokan cluster yang telah diprediksi sebelumnya. Berikut ini adalah contoh tabel yang menunjukkan pembagian tersebut.

Tabel 4. 13 Tabel *Data Selection* pada *AR-MBA*

Kluster	ID Transaksi	Nama Barang
1	JF000020123	['Nabati Nextar Blueberry 106g', 'Pocky Straw Stick 47gr']
	JF000021122	['Gg Surya 16s', 'Mogu2 Melon 320ml']
	JF000021222	['Telur Ay.Ras Curah Gr', 'Sari/R Rt Spesial']
2	JF000040123	['Sasa Tp Bbu Original 225', 'Sedaap Baso Special 77gr', 'Sedaap Goreng 72gr', 'Lifebuoy Mildcare 85gr', 'Sunlight Jr Nipis 95ml']
	JF000061222	['Gg Sign 12s', 'Indomie Gor Spc 84gr', 'Super/Bbr Ayam 45gr', 'Roma Malkist Kopyor 252gr', 'Migelas Baso Sapi 28g', 'Mie Gelas Grg Sosis 28g', 'Stella Car.Fres Energy 70gr']
	JF000071122	['Nutrijell Org 10gr', 'Nutrijell Anggur 10gr', 'Sasa Sant.Klp 65', 'Royco Sapi 100 Gr', 'Indo/F Racik N.Grg Sosis20']
3	JF001460123	['Dsa Sarung Tangan Motor', 'Glade Air/F Orange Peach 70 Gr']
	JF001500123	['Abc Exo Milk 230ml', 'Le Minerale Btl 1,5l', 'Soba Mie SmbL.Bldo 24gr']
	JF001501222	['Nissin Choco Soes 100gr', 'Sari/R Rm Sobek Cokl']
4	JF000371122	['Telur Ay.Ras Curah Gr', 'Rinso Det.B A.Noda 800g', 'Close Up Icy White 160g', 'Oral B 123 M 1"s", 'Fraiswell Myk Grg Btl 500ml', 'Bon Cabe Lv15 Btl 45gr']
	JF000460123	['Telur Ay.Ras Curah Gr', 'Dlg Beras C4 Raja 5kg', 'Ultra T/Kotak Lemon 200ml']
	JF000761122	['Fiesta Zoo 500gr', 'Sido/M T.Angin+Md 15ml', 'Stt French.F 2000 28gr', 'Larutan Penyegar Straw', 'Fortune M.Gor Reff 1l', 'Promina Puff Blbry 15',

'Nabati Richoco Cokl 50gr', 'Gery Malk.Swt Ches 110gr', 'Chiki Ball Cheese 10gr', 'Glico Pocky Choc.Bnn 47gr', 'Coca Cola Pet 250ml', 'Aice Ic Mochi Durian 45gr', 'Telur Ay.Ras Curah Gr', 'Mamy/P Pant Std XI26', 'Sgm Eks 3+ Mdu 900', 'Absolute Green Tea 60 MI', 'Rinso Det.B A.Noda 44g', 'Downy Pwangi Sun.R.Frs 25', 'Molto Edp Enchanting 10ml']

Pada tabel di atas, terlihat bahwa transaksi telah diurutkan agar sesuai dengan cluster yang telah ditentukan sebelumnya. Tujuan dari penyortiran ini adalah agar peneliti dapat memberikan rekomendasi yang sesuai dengan karakteristik masing-masing cluster dan aturan asosiasi pada keranjang belanja. Dengan demikian, peneliti dapat mengidentifikasi pola pembelian yang konsisten dalam setiap cluster dan memberikan saran yang relevan untuk meningkatkan penjualan atau memberikan rekomendasi produk kepada pelanggan berdasarkan kebutuhan dan preferensi mereka dalam kelompok yang sama.

#### 4.2.5.2 Data Cleaning pada Metode AR-MBA

Pada tahap ini, dilakukan proses pembersihan data transaksi. Salah satu langkah penting dalam tahap ini adalah menghapus transaksi yang hanya memiliki satu item belanja. Hal ini dilakukan untuk memfokuskan analisis pada transaksi yang memiliki lebih dari satu item, karena transaksi tersebut cenderung lebih relevan dalam mencari pola asosiasi yang signifikan. Berikut merupakan contoh data transaksi yang hanya memiliki 1 item.

Tabel 4. 14 Tampilan Data Memiliki 1 Barang

Kluster	ID Transaksi	Nama Barang
1	JF022260123	[Aice Ic Sweet Corn 52gr]
1	JF022261122	[Dunhil Filter Htm 16s]
2	JF085651222	[Telur Ay.Ras Curah Gr]
3	JF047771122	[Clasmild Redmax 16]
3	JF087281122	[Fruit/T Markisa 350ml]
4	JF056151122	[Gas Elpiji Isi 12 Kg]
4	JF066651222	[Aqua Air.Min Btl 600ml]

Berikut ini merupakan kondisi data sebelum setelah dilakukan filter terhadap keranjang belanja pada tiap transaksi.

Tabel 4. 15 Sebelum dan Sesuda Pembersihan Data

Kluster	Panjang Data Sebelum difilter	Panjang Data Setelah difilter
1	12825	5471
2	4191	4190
3	14172	6067
4	547	383
Total	31735	16111

Dilihat dari tabel diatas bahwa pengurangan data yang signifikan dari 31.735 menjadi 16.111 baris adalah langkah yang penting dalam melakukan aturan asosiasi pada data transaksi. Dengan menghilangkan transaksi yang hanya memiliki satu item belanja, fokus analisis dapat dipersempit pada transaksi yang memiliki lebih dari satu item belanja, yang cenderung memberikan informasi yang lebih relevan dalam mengungkap pola pembelian. Selain itu, pengurangan data juga mengurangi kompleksitas perhitungan dan memfokuskan sumber daya pada analisis yang lebih terstruktur dan bermakna. Dengan demikian, pengurangan data yang dilakukan memungkinkan peneliti untuk mendapatkan wawasan yang lebih mendalam tentang hubungan antar item dan pola asosiasi yang lebih signifikan dalam data transaksi.

#### 4.2.5.3 Data Transformation pada Metode AR-MBA

Pada tahap ini peneliti melakukan transformasi data pada keranjang belanja pada tiap transaksi. Transformasi data ini akan mengubah data yang mempunyai jenis data *list/object* menjadi *boolean (True/False)*. Berikut ini merupakan tampilan data yang diubah menjadi kedalam bentuk *boolean*.

Tabel 4. 16 Hasil Transformasi Data Menjadi *Boolean*

1000 Bintang Gula Pth 900gr	3 Ayam M.Telur Kng 200gr	A/Easy Det.B Bloss700g	A/Jazz1 Det.B S.Cinta900	A/Jazz1 Det.B.S.Cinta900	Abc Alkalin Aa- Lr6/4
False	False	False	False	False	False
False	False	False	False	False	False
False	False	False	False	False	False
False	False	False	False	False	False

Dalam tampilan tersebut, nilai *False* menunjukkan bahwa item tersebut tidak ada dalam transaksi. Sedangkan nilai *True* menunjukkan bahwa pelanggan memasukkan item tersebut ke dalam transaksi belanja mereka. Dengan transformasi ini, data keranjang belanja menjadi lebih terstruktur dan dapat digunakan dalam analisis pola asosiasi.

#### 4.2.6 Proses AR-MBA dengan FP-Growth

Pada tahap ini, peneliti akan menggunakan algoritma *FP-Growth* untuk membuat aturan asosiasi berdasarkan data yang telah ditransformasikan sebelumnya. Proses pembuatan aturan asosiasi akan dilakukan menggunakan software VSCode dan memanfaatkan library *mlxtend* dalam bahasa pemrograman Python. Peneliti akan mengatur nilai *minimum support* sebesar 0.001 untuk kluster 1, kluster 2, dan kluster 3, serta menggunakan nilai *minimum support* sebesar 0.01 untuk kluster 4. Selain itu, nilai *minimum confidence* juga ditetapkan sebesar 0.4.

Sebelum menentukan *minimum support*, peneliti akan melakukan perbandingan nilai *support* dari tiap-tiap cluster. Berikut merupakan perbandingan *minimum support* dari tiap-tiap cluster:

Tabel 4. 17 Perbandingan Nilai *Minimum Support*

Kluster	Min Support	Min Confidence	Jumlah Rule	Lift ratio Terbesar
1	0.001	0.4	23	248.681818
	0.002	0.4	8	107.538084
	0.005	0.4	2	44.818927
	0.01	0.4	0	NaN
2	0.001	0.4	104	206.9
	0.002	0.4	40	112.2
	0.005	0.4	8	5.2
	0.01	0.4	5	4.8
3	0.001	0.4	5	148.304
	0.002	0.4	0	0
	0.005	0.4	0	0
	0.01	0.4	0	0
4	0.005	0.4	1193	191.5
	0.007	0.4	102	127.7
	0.01	0.4	35	25.5
	0.05	0.4	2	1.3

Dari tabel diatas peneliti menentukan nilai *minimum support* untuk kluster 1,2, dan 3 adalah 0.001 karena memiliki jumlah rule yang terbanyak, tidak terlalu lama memakan waktu dalam

menjalankan algoritmanya, dan memiliki nilai *lift ratio* terbesar diatas 1. Sedangkan pada kluster 4 peneliti memilih nilai *minimum support* sebesar 0.01 dikarenakan jumlah rule pada nilai *minimum support* tersebut sudah dirasa cukup, memiliki waktu yang cepat dalam ketika menjalankan algoritmanya, dan memiliki nilai *lift ratio* diatas 1. Berikut merupakan tabel jumlah aturan asosiasi yang dihasilkan pada tiap *cluster*.

Tabel 4. 18 Deskripsi Aturan Asosiasi Tiap *Cluster*

<b>Kluster</b>	<b>Min support (%)</b>	<b>Min Confidence (%)</b>	<b>Jumlah Aturan Asosiasi</b>	<b>Lift ratio Terbesar</b>
<b>1</b>	0.001	0.4	23	248.681818
<b>2</b>	0.001	0.4	104	206.91358
<b>3</b>	0.001	0.4	5	148.304444
<b>4</b>	0.01	0.4	35	25.533333

Nilai dari *lift ratio* dapat di interpretasi sebagai berikut:

- Jika *lift ratio* = 1, maka tidak ada hubungan antara *antecedents* dan *consequents*. Keduanya muncul secara acak.
- Jika *lift ratio* > 1, maka ada kecenderungan positif antara *antecedents* dan *consequents*. Kemunculan *antecedents* lebih mungkin diikuti oleh kemunculan *consequents*.
- Jika *lift ratio* < 1, maka ada kecenderungan negatif antara *antecedents* dan *consequents*. Kemunculan *antecedents* lebih mungkin mengurangi kemunculan *consequents*.

Dari tabel diatas, hasil dari tabel tersebut adalah sebagai berikut:

1. Pada Kluster 1, dengan menggunakan *min support* 0.001 dan *min confidence* 0.4, terdapat 23 aturan asosiasi dengan *lift ratio* terbesar sebesar 248.68. *Lift ratio* yang lebih besar dari 1 menunjukkan adanya hubungan yang positif antara *antecedents* dan *consequents* pada aturan-aturan tersebut.
2. Pada Kluster 2, dengan menggunakan *min support* 0.001 dan *min confidence* 0.4, terdapat 104 aturan asosiasi dengan *lift ratio* terbesar sebesar 206.91. *Lift ratio* yang lebih besar dari 1 menunjukkan adanya hubungan yang positif antara *antecedents* dan *consequents* pada aturan-aturan tersebut.
3. Pada Kluster 3, dengan *min support* 0.001 dan *min confidence* 0.4, ditemukan 5 aturan asosiasi dengan *lift ratio* terbesar sebesar 148.30. *Lift ratio* yang lebih besar dari 1 menunjukkan adanya kecenderungan positif antara *antecedents* dan *consequents* dalam aturan-aturan tersebut.

4. Pada Kluster 4, dengan *min support* 0.01 dan *min confidence* 0.4, terdapat 35 aturan asosiasi dengan *lift ratio* terbesar sebesar 25.53. *Lift ratio* yang lebih besar dari 1 menunjukkan adanya kecenderungan positif antara *antecedents* dan *consequents* pada aturan-aturan tersebut.

Setelah melakukan interpretasi terhadap nilai *lift ratio* masing-masing cluster. Peneliti akan menampilkan rule yang dihasilkan dari tiap-tiap cluster. Dibawah ini merupakan tabel yang menampilkan aturan asosiasi pada *cluster* 1. Peneliti akan memperlihatkan aturan asosiasi dengan nilai *confidence* tertinggi, dan hanya akan menampilkan 5 aturan asosiasi dari total 23 aturan yang ada pada cluster 1.

Tabel 4. 19 Aturan Asosiasi pada *Cluster* 1

<b>Antecedents</b>	<b>Consequents</b>	<b>Support</b>	<b>Confidence</b>	<b>Lift</b>
(Ultra Mimi Straw 125ml, Aqua Air.Min Btl 600ml)	(Ultra Mimi Van Tpk 125ml)	0.001279	1	248.681818
(Sampoerna Mild 12s, Bear/B Steril Klg 189ml)	(Sido/M T.Angin Flu)	0.001645	1	91.183333
(Regal Marie 12g, Ultra Mimi Van Tpk 125ml)	(Ultra Mimi Straw 125ml)	0.001462	1	147.864865
(Ultra Mimi Straw 125ml, Regal Marie 12g)	(Ultra Mimi Van Tpk 125ml)	0.001462	0.888889	221.050505
(Aqua Air.Min Btl 600ml, Ultra Mimi Van Tpk 125ml)	(Ultra Mimi Straw 125ml)	0.001279	0.875	129.381757

Dari kelima aturan asosiasi yang terbentuk pada kluster 1 dengan nilai *confidence* tertinggi, peneliti dapat melihat produk apa saja yang cenderung dibeli oleh *customer*. Kluster ini menunjukkan kecenderungan pelanggan untuk membeli snack seperti Regal Marie 12g, minuman berbentuk susu seperti Ultra Mimi Van Tpk 125ml, dan minuman air putih seperti Aqua Air.Min Btl 600ml. Dari analisis tersebut, dapat disimpulkan bahwa pelanggan dalam

kluster 1 cenderung memiliki pola pembelian yang sering mencakup kombinasi dari snack, minuman susu, dan minuman air putih.

Selanjutnya, dibawah ini merupakan tabel yang menampilkan aturan asosiasi pada *cluster* 2. Peneliti akan memperlihatkan aturan asosiasi dengan nilai *confidence* tertinggi, dan hanya akan menampilkan 5 aturan asosiasi dari total 104 aturan yang ada pada cluster 2.

Tabel 4. 20 Aturan Asosiasi pada *Cluster* 2

<b>Antecedents</b>	<b>Consequents</b>	<b>Support</b>	<b>Confidence</b>	<b>Lift</b>
(Mie Gelas Grg Sosis 28g, Indomie Gor Rendang 91gr)	(Telur Ay.Ras Curah Gr)	0.002147971	1	4.935217903
(Mie Gelas Grg Sosis 28g, Indomie Gor Aceh 90gr)	(Telur Ay.Ras Curah Gr)	0.002147971	1	4.935217903
(Energen Choc Sct 30g, Sedaap Kari Spc 75gr)	(Energen Crl.Van 10x29gr)	0.001193317	1	110.2631579
(Mie Gelas Grg Sosis 28g, Indomie Gor Aceh 90gr, Indomie Gor Rendang 91gr)	(Telur Ay.Ras Curah Gr)	0.001670644	1	4.935217903
(Lucky Strike Filters 20s)	(Dunhil Filter Htm 16s)	0.001909308	0.888888889	206.9135802

Dari kelima aturan asosiasi yang terbentuk pada kluster 2 dengan nilai *confidence* tertinggi, peneliti dapat melihat produk-produk apa saja yang cenderung dibeli oleh pelanggan. Kluster ini menunjukkan kecenderungan pelanggan untuk membeli makanan instan seperti Mie Gelas Grg Sosis 28g, Indomie Gor Rendang 91gr, dan Indomie Gor Aceh 90gr. cluster ini juga cenderung membeli produk Telur Ay.Ras Curah Gr ketika melakukan pembelian makanan instan. Dari aturan-aturan tersebut dapat memberikan penjelasan terkait penyebab tingginya perbedaan rata-rata total transaksi yang dilakukan pada kluster 2 dengan kluster 4 meskipun memiliki variasi produk yang sama-sama *diverse*.

Selanjutnya, dibawah ini merupakan tabel yang menampilkan aturan asosiasi pada *cluster* 3. Peneliti akan memperlihatkan aturan asosiasi dengan nilai *confidence* tertinggi, dan hanya akan menampilkan 5 aturan asosiasi dari total 5 aturan yang ada pada cluster 3.

Tabel 4. 21 Aturan Asosiasi pada *Cluster* 3

<b>Antecedents</b>	<b>Consequents</b>	<b>Support</b>	<b>Confidence</b>	<b>Lift</b>
(Tora/B Susu Moka10x33gr)	(La Light Ice 16s)	0.001813	0.733333	148.304444
(Sun Kara Santan 65ml)	(Telur Ay.Ras Curah Gr)	0.001154	0.583333	5.998446
(Dlg Beras C4 Raja 5kg)	(Telur Ay.Ras Curah Gr)	0.001978	0.5	5.141525
(Ff Skm Gold Rcg 6x40gr)	(Sampoerna Mild 16s)	0.001319	0.470588	38.581876
(Minyak Kita Btl 1lt)	(Telur Ay.Ras Curah Gr)	0.001978	0.413793	4.255056

Dari kelima aturan asosiasi yang terbentuk pada kluster 3 dengan nilai *confidence* tertinggi, peneliti dapat melihat produk-produk apa saja yang cenderung dibeli oleh pelanggan. Kluster ini menunjukkan kecenderungan pelanggan untuk membeli bahan-bahan dapur seperti Dlg Beras C4 Raja 5kg, Minyak Kita Btl 1lt, Sun Kara Santan 65ml, dan Telur Ay.Ras Curah Gr.

Selanjutnya, dibawah ini merupakan tabel yang menampilkan aturan asosiasi pada *cluster* 4. Peneliti akan memperlihatkan aturan asosiasi dengan nilai *confidence* tertinggi, dan hanya akan menampilkan 5 aturan asosiasi dari total 35 aturan yang ada pada cluster 4.

Tabel 4. 22 Aturan Asosiasi pada Cluster 4

<b>Antecedents</b>	<b>Consequents</b>	<b>Support</b>	<b>Confidence</b>	<b>Lift</b>
(Champ Burger)	(Champ Sosis Ayam 375gr, "Meg Keju Slice 8s")	0.005221932	1	191.5
(Nissin Choco Soes 100gr, Energen Crl.Van 10x29gr)	(Sari/R Rm Sobek Cokl)	0.005221932	1	95.75
(Sari/R Rm Sobek Cokl, Nissin Choco Soes)	(Energen Crl.Van 10x29gr)	0.005221932	1	76.6



100gr, Energen Choc Sct 30g) (Sari/R Rm Sobek Cokl, Energen Choc Sct 30g, Energen Crl.Van 10x29gr) (Nissin Choco Soes 100gr, Energen Choc Sct 30g, Energen Crl.Van 10x29gr)	(Nissin Choco Soes 100gr)	0.005221932	1	191.5
	(Sari/R Rm Sobek Cokl)	0.005221932	1	95.75

---

Dari kelima aturan asosiasi yang terbentuk pada kluster 4 dengan nilai confidence tertinggi, peneliti dapat melihat produk-produk apa saja yang cenderung dibeli oleh pelanggan. Kluster ini memiliki kecenderungan untuk membeli minuman cereal seperti Energen Choc Sct 30g dan Energen Crl.Van 10x29gr. pelanggan pada kluster ini juga membeli berbagai produk makanan ringan seperti roti dan juga biskuit. Dari pola pembelian tersebut terdapat perbedaan jenis barang yang dibeli oleh kluster 1, kluster 2, kluster 3, dan kluster 4. Perbedaan-perbedaan ini dapat menjadi faktor penyebab perbedaan total transaksi pada tiap-tiap kluster.

## BAB V

### ANALISIS DAN PEMBAHASAN

#### 5.1 Analisis Clustering

Peneliti melakukan analisis *clustering* untuk mengelompokkan data transaksi pelanggan berdasarkan karakteristik yang serupa. Dalam analisis ini, peneliti menggunakan metode *K-means* untuk mengelompokkan data menjadi empat kluster yang berbeda. Setiap kluster memiliki ciri khasnya sendiri berdasarkan variabel waktu, variasi produk, dan total transaksi.

##### 5.1.1 Analisis Deskripsi Data

Penelitian ini menggunakan data transaksi selama 3 bulan dari bulan November 2022, Desember 2022, dan Januari 2023. Dari 3 bulan data transaksi tersebut peneliti mendapatkan data yang berisi 8 kolom dan 65536 baris. Peneliti menggunakan *software VSCode* dan bahasa pemrograman *Python* untuk melakukan *clustering K-Means*. Berikut ini merupakan deskripsi dari data transaksi Toko Intimart Gedongan yang akan digunakan analisis *clustering*.

Tabel 5. 1 Deskripsi Data Metode *Clustering*

Nama Variabel	Jumlah Data	Minimum	Maximum	Mean (Rata-rata)
Waktu	31735	5	22	14.7
Variasi Produk	31735	1	27	2.1
Total Transaksi	31735	460	754500	26343.6

Berdasarkan tabel diatas, dapat diamati beberapa informasi penting terkait transaksi pada Toko Intimart Gedongan. Pada variabel waktu, didapati waktu minimum pembelian pada jam 5 pagi dan waktu maksimal pada jam 10 malam dengan rata-rata transaksi dilakukan pada jam 2 siang. Dari nilai rata-rata waktu pembeli melakukan transaksi menunjukkan banyak transaksi yang terjadi pada jam 2 siang.

Selanjutnya pada variabel Variasi Produk, didapati minimum variasi produk yang dibeli oleh pelanggan ada 1 dan memiliki nilai maksimum sebesar 27 produk dengan rata-rata variasi produk yang dibeli pelanggan sebanyak 2 produk. Hal ini menunjukkan bahwa Sebagian besar pelanggan membeli beberapa produk dalam satu transaksi.

Kemudian pada variabel Total Transaksi, didapati minimum total transaksi yang dibeli pelanggan sebesar Rp. 460,- dan maksimum sebesar Rp. 754.500,- dengan nilai rata-rata total transaksi sebesar Rp. 26.343,-. Hal ini menunjukkan adanya perbedaan besar dalam jumlah pembelian yang dilakukan oleh pelanggan, dengan beberapa transaksi memiliki nilai yang jauh lebih tinggi dibandingkan yang lain.

### 5.1.2 Proporsi Anggota Kluster

Pada tahap ini, proporsi anggota cluster digunakan untuk mengukur distribusi komponen suatu variabel dalam populasi cluster. Proporsi anggota cluster ini dihitung dalam persentase untuk memberikan pemahaman yang lebih jelas tentang karakteristik dan perbandingan antara cluster. Proporsi anggota cluster memberikan gambaran tentang sejauh mana suatu cluster mendominasi dalam populasi data. Hal ini penting untuk memahami perbedaan dan keunikan antara cluster, serta memperoleh wawasan tentang komposisi variabel dalam setiap cluster.

Tabel 5. 2 Proporsi Jumlah Cluster

Cluster	Jumlah	Presentase (%)
1	12825	40.4%
2	4191	13.2%
3	14172	44.7%
4	547	1.7%
Total	31735	100.0%

Berdasarkan tabel proporsi anggota cluster di atas, dapat dilihat bahwa cluster 1 memiliki jumlah data sebanyak 12.825, yang merupakan proporsi sebesar 40,4% dari total data. Cluster 2 memiliki jumlah data sebanyak 4.191, dengan proporsi sebesar 13,2%. Cluster 3 memiliki jumlah data terbanyak yaitu 14.172, dengan proporsi sebesar 44,7%. Sementara itu, cluster 4 memiliki jumlah data yang paling sedikit, hanya 547 data, dengan proporsi sebesar 1,7%. Total keseluruhan data dalam analisis adalah 31.735.

Informasi ini memberikan gambaran tentang sebaran data di masing-masing cluster. Cluster 1, 2, dan 3 memiliki proporsi yang signifikan dalam data, sementara cluster 4 memiliki proporsi yang relatif kecil. Analisis lebih lanjut dapat dilakukan dengan memperhatikan karakteristik dan pola yang muncul di setiap cluster untuk mendapatkan wawasan yang lebih dalam mengenai kelompok-kelompok pelanggan yang berbeda.

### 5.1.3 Analisis Karakteristik Kluster

Pada tahap analisis karakteristik kluster, peneliti akan melakukan *profiling* terhadap kluster yang telah dibentuk oleh algoritma *K-Means*. Agar peneliti dapat memprofilkan kluster yang dibentuk, diperlukan deskripsi data hasil cluster yaitu rata-rata, batas bawah (Q1), dan batas atas (Q3). Berikut ini merupakan tabel rata-rata tiap kluster.

Tabel 5. 3 Tabel Rata-Rata Pusat *Cluster*

	<b>Waktu</b>	<b>Variasi Produk</b>	<b>Total Transaksi (Rp)</b>
Kluster 1	10.33	1.58	19866.89
Kluster 2	15.88	4.83	49271.20
Kluster 3	18.25	1.56	19591.12
Kluster 4	13.87	4.73	177475.84

Berikut ini merupakan persebaran data dari masing-masing kluster. Disini peneliti menggunakan Q1 dan Q3 sebagai batas bawah dan batas atas dari persebaran data. Penggunaan Q1 (Kuartil Pertama) dan Q3 (Kuartil Ketiga) sebagai batas bawah dan batas atas dari persebaran data juga memberikan wawasan tambahan tentang distribusi data di dalam kluster. Q1 dan Q3 masing-masing merepresentasikan nilai persentil ke-25 dan ke-75 dari data. Dengan demikian, rentang antara keduanya (IQR - *Interquartile Range*) mencakup 50% dari data, yang mencerminkan distribusi data yang terpusat pada kisaran nilai tersebut.

Tabel 5. 4 Tabel Persebaran Data Tiap *Cluster*

<b>Cluster</b>		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
<b>Waktu</b>	Batas Bawah (Q1)	8	13	17	10
	Batas Atas (Q3)	12	19	20	18
<b>Variasi Produk</b>	Batas Bawah (Q1)	1	4	1	1
	Batas Atas (Q3)	2	5	2	7
<b>Total Transaksi (Rp)</b>	Batas Bawah (Q1)	8925	31000	9500	128050
	Batas Atas (Q3)	26450	63434.75	26400	209883.75

Berdasarkan nilai rata-rata, batas bawah (Q1), dan batas atas (Q3) dari masing-masing kluster, dapat diambil kesimpulan sebagai berikut:

1. Cluster 1:

- Pada Variabel Waktu Transaksi, rata-rata yang didapatkan adalah 10.3, dengan persebaran data antara 8 dan 12. Hal ini menunjukkan bahwa sebagian besar transaksi dalam kluster ini dilakukan pada jam 8 pagi hingga jam 12 siang.
- Pada Variabel Variasi Produk, rata-rata yang didapatkan adalah 1.6, dengan persebaran data antara 1 hingga 2 variasi produk. Kluster ini menunjukkan bahwa kluster ini mencerminkan pelanggan yang cenderung membeli produk yang sama atau memiliki sedikit variasi produk.
- Pada Variabel Total Transaksi (Rp), rata-rata nilai transaksi adalah Rp. 19.866,-, dengan rentang data dari Rp. 8.925,- sampai Rp. 26.540,-. Hal ini menunjukkan bahwa pelanggan pada kluster ini memiliki nilai transaksi yang berada dalam kisaran sedang atau menengah.
- Kluster ini dapat disebut "*Mid-Morning Moderates*", karena kluster ini memiliki nilai transaksi dalam kisaran sedang atau menengah. Mereka cenderung membeli produk yang sama atau memiliki sedikit variasi, dan melakukan transaksi pada waktu pagi hingga siang. "*Mid-Morning*" menggambarkan waktu dominan saat pelanggan dalam kluster ini melakukan transaksi. "*Moderates*" menekankan pada tingkat pembelian yang stabil dan berimbang.

2. Cluster 2:

- Pada Variabel Waktu Transaksi, rata-rata yang didapatkan adalah 15.9, dengan persebaran data antara 13 dan 19. Hal ini menunjukkan bahwa Sebagian besar transaksi dalam kluster ini dilakukan pada jam 1 siang hingga jam 7 malam.
- Pada Variabel Variasi Produk, rata-rata yang didapatkan adalah 4.8, dengan persebaran data antara 4 hingga 5 variasi produk. Kluster ini menunjukkan pelanggan yang cenderung membeli berbagai variasi produk.
- Pada Variabel Total Transaksi, rata-rata nilai transaksi adalah Rp. 49.271,- dengan rentang data dari Rp. 31.000,- hingga Rp. 63.434,-. Hal ini menunjukkan bahwa pelanggan pada kluster ini memiliki nilai transaksi yang berada dalam kisaran cukup tinggi.
- Kluster ini dapat disebut "*Diverse Afternoon Buyers*", karena kluster ini mencerminkan pelanggan yang memiliki nilai transaksi yang berada dalam kisaran cukup tinggi.

Mereka cenderung membeli berbagai variasi produk, dan melakukan transaksi pada waktu siang hingga sore. "*Diverse*" menggambarkan kecenderungan pelanggan dalam kluster ini untuk mencoba berbagai produk yang ditawarkan. "*Afternoon Buyers*" menekankan pada waktu dominan saat pelanggan melakukan transaksi.

- Kluster ini merupakan kelompok pelanggan yang penting untuk fokuskan dalam strategi pemasaran yang berorientasi pada nilai transaksi tinggi dan variasi produk, khususnya pada waktu siang hingga sore hari.

### 3. Cluster 3:

- Pada Variabel Waktu Transaksi, rata-rata yang didapatkan adalah 18.3, dengan persebaran data antara 17 dan 20. Hal ini menunjukkan bahwa Sebagian besar transaksi dalam kluster ini dilakukan pada jam 5 sore hingga jam 8 malam.
- Pada Variabel Variasi Produk, rata-rata yang didapatkan adalah 1.6, dengan persebaran data antara 1 hingga 2 variasi produk. Kluster ini mirip dengan kluster 1, di mana pelanggan cenderung membeli produk dengan sedikit variasi.
- Pada Variabel Total Transaksi, rata-rata nilai transaksi adalah Rp. 19.591,- dengan rentang data dari Rp. 9.500,- hingga Rp. 26.400,-. Hal ini menunjukkan bahwa pelanggan pada kluster ini cenderung melakukan transaksi dengan nilai yang berada dalam kisaran sedang atau menengah, tidak terlalu rendah namun juga tidak terlalu tinggi.
- Kluster ini dapat disebut "*Evening Moderates*", karena kluster ini mencerminkan pelanggan yang memiliki nilai transaksi dalam kisaran sedang atau menengah. Mereka cenderung membeli produk dengan sedikit variasi, dan melakukan transaksi pada waktu sore hingga malam. "*Evening*" menggambarkan waktu dominan saat pelanggan dalam kluster ini melakukan transaksi. "*Moderates*" menekankan pada kecenderungan pelanggan untuk bertransaksi dengan tingkat yang stabil dan berimbang.
- Kluster ini merupakan kelompok pelanggan yang melakukan transaksi pada waktu sore hingga malam, memilih produk dengan variasi yang terbatas, dan berorientasi pada transaksi yang stabil dan konsisten.

### 4. Cluster 4:

- Pada Variabel Waktu Transaksi, rata-rata yang didapatkan adalah 13.9, dengan persebaran data antara 10 dan 18. Hal ini menunjukkan bahwa Sebagian besar transaksi dalam kluster ini dilakukan pada jam 10 pagi hingga jam 6 sore.

- Pada Variabel Variasi Produk, rata-rata yang didapatkan adalah 4.7, dengan persebaran data antara 1 hingga 7 variasi produk. Kluster ini menunjukkan pelanggan yang cenderung membeli berbagai variasi produk.
- Pada Variabel Total Transaksi, rata-rata nilai transaksi adalah Rp. 177.475,- dengan rentang data dari Rp. 128.050,- hingga Rp. 209.883,-. Kluster ini mencerminkan pelanggan yang melakukan beberapa transaksi dengan nilai yang sangat tinggi, jauh di atas rata-rata transaksi dari kluster lainnya.
- Kluster ini dapat disebut "*High-Value Customers*", Kluster ini dapat disebut "*High-Value Customers*" karena mencerminkan pelanggan yang memiliki nilai transaksi yang sangat tinggi. Sebutan ini menggambarkan bahwa kluster ini merupakan kelompok pelanggan yang sangat berharga dan dapat menjadi fokus utama dalam strategi pemasaran yang bertujuan untuk mempertahankan dan meningkatkan loyalitas pelanggan dengan nilai transaksi tinggi.
- Pelanggan dalam kluster ini cenderung melakukan beberapa transaksi dengan nilai yang jauh di atas rata-rata kluster lainnya, sehingga mereka memiliki kontribusi yang signifikan pada pendapatan toko Intimart Gedongan.

Dari keempat kluster di atas, kluster yang paling baik untuk dilakukan perbaikan dan diberikan rekomendasi adalah Kluster 2, yaitu "*Diverse Afternoon Buyers*". Kluster ini memiliki nilai transaksi yang berada dalam kisaran cukup tinggi dan pelanggan cenderung membeli berbagai variasi produk. Pelanggan dalam kluster ini melakukan transaksi pada waktu siang hingga sore, yang merupakan waktu dengan aktivitas belanja yang tinggi.

## 5.2 Analisis Association Rule Market Basket Analysis

Market Basket Analysis (MBA) adalah teknik data mining yang digunakan untuk menemukan pola asosiasi dalam transaksi jual beli, mengungkap kebiasaan pembelian konsumen dalam satu keranjang belanja, dan memahami pola pembelian barang dengan menghasilkan aturan atau rules yang menggambarkan hubungan antara produk yang sering dibeli bersamaan (Ramadana et al., 2020).

### 5.2.1 Analisis Perbandingan Association Rule

Pada analisis perbandingan aturan asosiasi ini, peneliti akan menggunakan data transaksi pada bulan November 2022, Desember 2022, dan Januari 2023. Data transaksi yang digunakan terdiri dari 4190 transaksi yang masuk ke dalam kluster 2, yaitu "*Diverse Afternoon Buyers*".

Sebelum dilakukan perhitungan aturan asosiasi dengan algoritma *FP-Growth*, data transaksi telah diolah dan direduksi untuk mempersiapkan data yang akan digunakan dalam analisis. Proses perhitungan aturan asosiasi dilakukan menggunakan bahasa pemrograman *Python* dengan bantuan *library mlxtend*.

Pengolahan data untuk menemukan aturan asosiasi dilakukan dengan membandingkan *minimum support*. Berikut merupakan tabel hasil perbandingan yang terbentuk.

Tabel 5. 5 Perbandingan *Minimum Support*

<i>Min Support</i>	<i>Min Confidence</i>	Jumlah Rule	<i>Lift ratio</i> Terbesar
0.001	0.4	104	206.9
0.002	0.4	40	112.2
0.005	0.4	8	5.2
0.01	0.4	5	4.8

Pada tabel di atas, terlihat bahwa penggunaan *minimum support* yang berbeda menghasilkan jumlah *rule* yang beragam. Dalam analisis ini, peneliti menggunakan empat nilai *minimum support*, yaitu 0.001, 0.002, 0.005, dan 0.01, dengan menggunakan *minimum confidence* tetap sebesar 0.4. Hasil perbandingan menunjukkan bahwa peneliti mendapatkan jumlah *rule* dan *lift ratio* terbesar pada *minimum support* sebesar 0.001, dengan 104 *rule* dan *lift ratio* sebesar 206.9.

Dengan demikian, peneliti akan menggunakan *minimum support* sebesar 0.001 dalam analisis aturan asosiasi menggunakan algoritma *FP-Growth*. Penggunaan nilai *minimum support* yang lebih rendah memungkinkan untuk mengidentifikasi lebih banyak pola asosiasi yang mungkin tersembunyi dalam data transaksi. Hal ini dapat memberikan wawasan yang lebih mendalam tentang kebiasaan pembelian pelanggan dalam kluster "*Diverse Afternoon Buyers*" dan memungkinkan penentuan strategi pemasaran yang lebih tepat dan efektif untuk meningkatkan nilai transaksi dan profitabilitas toko Intimart Gedongan.

### 5.2.2 Analisis Frekuensi Pembelian

Data transaksi yang digunakan dalam analisis aturan asosiasi merupakan daftar produk-produk yang dimasukkan ke dalam keranjang belanja oleh pelanggan di Toko Intimart Gedongan. Data yang digunakan mencakup transaksi pada bulan November 2022, Desember 2022, dan Januari 2023, dengan total 4190 transaksi. Data tersebut merupakan data transaksi



yang masuk kedalam kluster 2, yaitu "*Diverse Afternoon Buyers*". Dalam analisis ini, peneliti akan menampilkan frekuensi kemunculan setiap produk dalam keranjang belanja pelanggan.

Tabel 5. 6 Frekuensi Pembelian Kluster 2

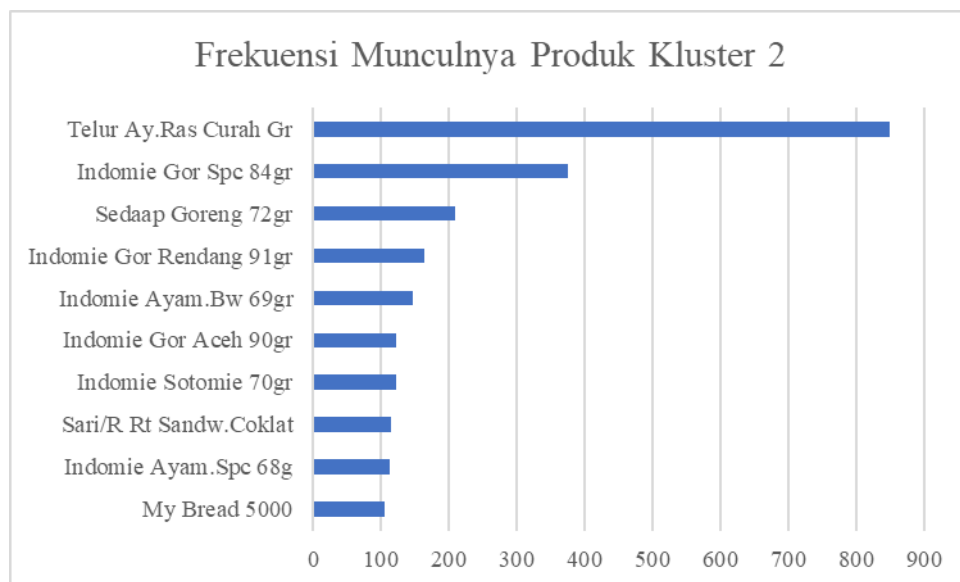
No	Produk	Frekuensi	Presentase
1	Telur Ay.Ras Curah Gr	849	4.20%
2	Indomie Gor Spc 84gr	376	1.86%
3	Sedaap Goreng 72gr	209	1.03%
4	Indomie Gor Rendang 91gr	163	0.81%
5	Indomie Ayam.Bw 69gr	146	0.72%
6	Indomie Gor Aceh 90gr	123	0.61%
7	Indomie Sotomie 70gr	122	0.60%
8	Sari/R Rt Sandw.Coklat	114	0.56%
9	Indomie Ayam.Spc 68g	113	0.56%
10	My Bread 5000	106	0.52%
:	:	:	:
2358	Hot In Cream 60gr	1	0.0049%
2359	Rapika Lavendr/S 400ml	1	0.0049%
2360	Absolute Green Tea 60 Ml	1	0.0049%
<b>Total</b>		<b>20232</b>	<b>100%</b>

Berdasarkan tabel frekuensi pembelian kluster 2 di atas, peneliti dapat mengetahui frekuensi dan presentase dari setiap produk yang ada pada data transaksi tersebut. Berikut merupakan analisis dari 10 produk yang memiliki frekuensi terbanyak pada data transaksi kluster 2.

1. "Telur Ay.Ras Curah Gr" memiliki frekuensi tertinggi sebanyak 849 kali atau sekitar 4.20% dari total transaksi.
2. "Indomie Gor Spc 84gr" merupakan produk dengan frekuensi kedua tertinggi, tercatat sebanyak 376 kali atau sekitar 1.86% dari total transaksi.
3. "Sedaap Goreng 72gr" menduduki posisi ketiga dengan frekuensi 209 kali atau sekitar 1.03% dari total transaksi.
4. "Indomie Gor Rendang 91gr" tercatat dengan frekuensi 163 kali atau sekitar 0.81% dari total transaksi.
5. "Indomie Ayam.Bw 69gr" memiliki frekuensi 146 kali atau sekitar 0.72% dari total transaksi.
6. "Indomie Gor Aceh 90gr" tercatat dengan frekuensi 123 kali atau sekitar 0.61% dari total transaksi.

7. "Indomie Sotomie 70gr" memiliki frekuensi 122 kali atau sekitar 0.60% dari total transaksi.
8. "Sari/R Rt Sandw.Coklat" tercatat dengan frekuensi 114 kali atau sekitar 0.56% dari total transaksi.
9. "Indomie Ayam.Spc 68g" memiliki frekuensi 113 kali atau sekitar 0.56% dari total transaksi.
10. "My Bread 5000" menduduki posisi kesepuluh dengan frekuensi 106 kali atau sekitar 0.52% dari total transaksi.

Data frekuensi ini membantu kita mengidentifikasi produk-produk yang paling populer dan sering dibeli oleh pelanggan dalam kluster 2 (*Diverse Afternoon Buyers*). Berikut ini merupakan diagram batang yang memuat 10 produk yang paling sering muncul kedalam keranjang kustomer.



Gambar 5. 1 Frekuensi Kemunculan Produk pada Kluster2

### 5.2.3 Output Association Rule

Pada perbandingan nilai minimum *support* yang dilakukan sebelumnya, peneliti memilih minimum *support* sebesar 0.001 dan minimum *confidence* sebesar 0.04. Dengan menggunakan kedua parameter tersebut, peneliti berhasil menemukan 104 rule dengan *lift ratio* terbesar mencapai 206.9. Berikut adalah tabel yang menampilkan 10 rule dengan *support* tertinggi pada analisis aturan asosiasi yang dilakukan oleh peneliti.

Tabel 5. 7 Aturan Asosiasi dengan *Support* Tertinggi

<b>Antecedents</b>	<b>Consequents</b>	<b>Support</b>	<b>Confidence</b>	<b>Lift</b>
(Indomie Gor Spc 84gr)	(Telur Ay.Ras Curah Gr)	0.0360	0.402	1.982
(Indomie Gor Spc 84gr, Indomie Gor Aceh 90gr)	(Telur Ay.Ras Curah Gr)	0.0026	0.407	2.011
(Mie Gelas Grg Sosis 28g)	(Indomie Gor Rendang 91gr, Indomie Gor Aceh 90gr, Telur Ay.Ras Curah Gr)	0.0017	0.412	115.020
(Mie Gelas Grg Sosis 28g)	(Indomie Gor Rendang 91gr, Indomie Gor Aceh 90gr)	0.0017	0.412	44.238
(Indomie Gor Spc 84gr, Sedaap Soto 72gr)	(Telur Ay.Ras Curah Gr)	0.0017	0.412	2.032
(Bango Kecap Mns 220ml)	(Telur Ay.Ras Curah Gr)	0.0012	0.417	2.056
(Delfi Funt.Mlk Vnl26)	(Gery Choco.Grande 17gr)	0.0012	0.417	96.991
(Gg Cig Patra 12)	(Telur Ay.Ras Curah Gr)	0.0012	0.417	2.056
(Indomie Gor Geprek 85gr, Indomie Gor Aceh 90gr)	(Telur Ay.Ras Curah Gr)	0.0012	0.417	2.056
(Sedaap Kari Spc 75gr, Sedaap Baso Special 77gr)	(Telur Ay.Ras Curah Gr)	0.0012	0.417	2.056

Pemilihan nilai *support* tertinggi dilakukan karena *support* merupakan ukuran seberapa sering aturan tersebut muncul dalam data transaksi. Dengan memilih nilai *support* yang tinggi, peneliti dapat fokus pada aturan yang memiliki dampak signifikan dan sering terjadi dalam pembelian produk oleh pelanggan. Pada tabel diatas, merupakan hasil pemilihan 10 aturan asosiasi yang memiliki nilai *support* tertinggi dan memiliki nilai *lift ratio* diatas 1. Berikut merupakan penjelasan dari tiap-tiap rule:

1. Aturan pertama menghasilkan kombinasi antara produk "Indomie Gor Spc 84gr" dan "Telur Ay.Ras Curah Gr" dengan *support* sebesar 0.036 dan *confidence* sebesar 0.402. Hal ini menunjukkan bahwa 40.2% dari pelanggan yang membeli "Indomie Gor Spc 84gr" juga membeli "Telur Ay.Ras Curah Gr". Selain itu, *lift ratio* yang bernilai 1.98 menunjukkan adanya hubungan asosiasi yang cukup kuat antara kedua produk tersebut.
2. Aturan kedua merupakan kombinasi dari produk "Indomie Gor Spc 84gr" dan "Indomie Gor Aceh 90gr" yang menghasilkan pembelian "Telur Ay.Ras Curah Gr" dengan *support* sebesar 0.0026 dan *confidence* sebesar 0.407. Ini berarti 40.7% dari pelanggan yang membeli "Indomie Gor Spc 84gr" dan "Indomie Gor Aceh 90gr" juga membeli "Telur Ay.Ras Curah Gr". *Lift ratio* sebesar 2.01 menunjukkan adanya hubungan asosiasi yang kuat antara kombinasi produk tersebut dengan pembelian "Telur Ay.Ras Curah Gr".
3. Aturan ketiga memiliki kombinasi produk yang lebih kompleks, produk "Mie Gelas Grg Sosis 28g" menghasilkan pembelian "Indomie Gor Rendang 91gr", "Indomie Gor Aceh 90gr", dan "Telur Ay.Ras Curah Gr". Aturan ini memiliki *support* yang sangat rendah, hanya sekitar 0.0016, tetapi *confidence* yang cukup tinggi, yaitu sekitar 0.412. *Lift ratio* yang sangat tinggi mencapai 115.02 menunjukkan adanya hubungan asosiasi yang sangat kuat ketika melakukan pembelian "Mie Gelas Grg Sosis 28g".
4. Aturan keempat juga memiliki kombinasi produk yang kompleks, yaitu pembelian produk "Mie Gelas Grg Sosis 28g" akan menghasilkan pembelian "Indomie Gor Rendang 91gr" dan "Indomie Gor Aceh 90gr". Aturan ini memiliki *support* dan *confidence* yang mirip dengan aturan ketiga, tetapi *lift ratio* yang lebih rendah, yaitu sekitar 44.24. Ini menunjukkan adanya hubungan asosiasi yang kuat antara kombinasi produk tersebut, namun tidak sekuat pada aturan ketiga.
5. Aturan kelima merupakan kombinasi antara produk "Indomie Gor Spc 84gr" dan "Sedaap Soto 72gr" yang menghasilkan pembelian "Telur Ay.Ras Curah Gr". Aturan ini memiliki *support*, *confidence*, dan *lift ratio* yang cukup tinggi, yaitu sekitar 0.0017, 0.412, dan 2.03, menunjukkan adanya hubungan asosiasi yang kuat antara kedua produk tersebut dengan pembelian "Telur Ay.Ras Curah Gr".
6. Aturan keenam merupakan pembelian produk "Bango Kecap Mns 220ml" yang menghasilkan pembelian "Telur Ay.Ras Curah Gr". Aturan ini memiliki *support* sebesar 0.0012, *confidence* sebesar 0.417, dan *lift ratio* sebesar 2.06. Hal ini menunjukkan bahwa 41.7% dari pelanggan yang membeli "Bango Kecap Mns 220ml" juga membeli "Telur Ay.Ras Curah Gr". *Lift ratio* yang lebih besar dari 1 menunjukkan adanya hubungan asosiasi yang positif antara kedua produk tersebut.

7. Aturan ketujuh merupakan pembelian produk "Delfi Funt.Mlk Vnl26" yang menghasilkan pembelian "Gery Choco.Grande 17gr". Aturan ini memiliki *support* sebesar 0.0012, *confidence* sebesar 0.417, dan *lift ratio* sebesar 96.99. *Lift ratio* yang sangat tinggi menunjukkan adanya hubungan asosiasi yang kuat antara kedua produk tersebut. Hal ini menunjukkan bahwa 41.7% dari pelanggan yang membeli "Delfi Funt.Mlk Vnl26" juga cenderung membeli "Gery Choco.Grande 17gr".
8. Aturan kedelapan merupakan pembelian produk "Gg Cig Patra 12" yang menghasilkan pembelian "Telur Ay.Ras Curah Gr". Aturan ini memiliki *support* sebesar 0.0012, *confidence* sebesar 0.417, dan *lift ratio* sebesar 2.06. Hal ini menunjukkan bahwa 41.7% dari pelanggan yang membeli "Gg Cig Patra 12" juga membeli "Telur Ay.Ras Curah Gr". *Lift ratio* yang lebih besar dari 1 menunjukkan adanya hubungan asosiasi yang positif antara kedua produk tersebut.
9. Aturan kesembilan merupakan kombinasi antara produk "Indomie Gor Geprek 85gr" dan "Indomie Gor Aceh 90gr" yang menghasilkan pembelian "Telur Ay.Ras Curah Gr". Aturan ini memiliki *support* sebesar 0.0012, *confidence* sebesar 0.417, dan *lift ratio* sebesar 2.06. Hal ini menunjukkan bahwa 41.7% dari pelanggan yang membeli "Indomie Gor Geprek 85gr" dan "Indomie Gor Aceh 90gr" juga membeli "Telur Ay.Ras Curah Gr". *Lift ratio* yang lebih besar dari 1 menunjukkan adanya hubungan asosiasi yang positif antara kedua produk tersebut dengan pembelian "Telur Ay.Ras Curah Gr".
10. Aturan kesepuluh merupakan kombinasi antara produk "Sedaap Kari Spc 75gr" dan "Sedaap Baso Special 77gr" yang menghasilkan pembelian "Telur Ay.Ras Curah Gr". Aturan ini memiliki *support* sebesar 0.0012, *confidence* sebesar 0.417, dan *lift ratio* sebesar 2.06. Hal ini menunjukkan bahwa 41.7% dari pelanggan yang membeli "Sedaap Kari Spc 75gr" dan "Sedaap Baso Special 77gr" juga membeli "Telur Ay.Ras Curah Gr". *Lift ratio* yang lebih besar dari 1 menunjukkan adanya hubungan asosiasi yang positif antara kedua produk tersebut dengan pembelian "Telur Ay.Ras Curah Gr".

### 5.3 Rekomendasi Hasil Penelitian

Rekomendasi yang diajukan dalam penelitian ini berfokus pada strategi promosi yang dapat diterapkan oleh Toko Intimart Gedongan, seperti *product bundling*, pemberian poin loyalitas, dan pemberian diskon. Keputusan untuk tidak merekomendasikan perubahan dalam tata letak toko didasarkan pada hasil identifikasi selama proses wawancara dengan pemilik toko. Pemilik toko dengan jelas menyatakan bahwa ia merasa puas dengan tata letak toko yang ada dan tidak merasa ada masalah yang signifikan. Oleh karena itu, dengan mempertimbangkan masukan

langsung dari pemilik toko, penelitian ini memutuskan untuk tidak mengusulkan perubahan dalam tata letak toko. Sebaliknya, fokus direkomendasikan pada pengembangan strategi promosi yang dapat meningkatkan kinerja dan loyalitas pelanggan secara keseluruhan. Hal ini diharapkan dapat membantu Toko Intimart Gedongan mencapai hasil yang lebih baik dalam bisnisnya. Dari analisis yang telah peneliti lakukan, maka peneliti dapat memberikan rekomendasi strategi peningkatan penjualan kepada Toko Intimart Gedongan. Berikut merupakan rekomendasi yang dapat peneliti berikan:

1. Berdasarkan 4 kluster yang telah dibentuk menggunakan algoritma *K-Means* peneliti menargetkan kluster 2 sebagai prioritas strategi peningkatan penjualan di Toko Intimart Gedongan. Pemilihan kluster kedua didasarkan pada nilai transaksi yang tinggi, dengan rata-rata nilai sebesar Rp. 49.271,- per transaksi. Selain itu, pelanggan dalam kluster ini cenderung membeli berbagai variasi produk, yaitu dengan persebaran 4 hingga 5 variasi produk, yang mendukung analisis aturan asosiasi. Selain itu, ditemukan bahwa kluster ini aktif bertransaksi pada rentang waktu antara jam 13.00 hingga jam 19.00, dengan puncak transaksi terjadi sekitar jam 14.00 hingga 15.00. Oleh karena itu, peneliti merekomendasikan Toko Intimart Gedongan untuk memfokuskan strategi peningkatan penjualan kepada pelanggan yang termasuk dalam kluster 2.
2. Peneliti merekomendasikan kepada pihak toko Intimart Gedongan untuk melakukan promosi khusus, diskon, atau penawaran menarik seperti produk *bundling* pada jam sibuk ketika customer dari kluster 2 berbelanja, yaitu pada rentang waktu jam 13.00 hingga jam 19.00. Dengan melakukan promosi pada periode tersebut, tidak hanya kluster 2 tetapi juga kluster 3 dan kluster 4 dapat ikut merasakan manfaat dari promosi yang dilakukan. Hal ini akan membantu menarik lebih banyak pelanggan, meningkatkan jumlah transaksi, dan memperluas pangsa pasar untuk semua kluster pelanggan yang berbelanja di toko.
3. Berdasarkan frekuensi pembelian pelanggan, terlihat bahwa produk "Telur Ay.Ras Curah Gr" menjadi produk yang paling diminati dengan frekuensi pembelian yang tinggi. Selanjutnya, berdasarkan analisis aturan asosiasi pada 10 aturan dengan *support* tertinggi, produk "Telur Ay.Ras Curah Gr" selalu termasuk dalam kombinasi pembelian pelanggan. Oleh karena itu, peneliti merekomendasikan kepada pihak Toko Intimart Gedongan untuk selalu menyediakan stok yang cukup untuk produk "Telur Ay.Ras Curah Gr".
4. Peneliti merekomendasikan untuk memberikan diskon pada produk *bundling* yang terbentuk dari aturan ke-3, yaitu dengan mengombinasikan produk "Mie Gelas Grg Sosis 28g", "Indomie Gor Rendang 91gr", "Indomie Gor Aceh 90gr", dan "Telur Ay.Ras Curah Gr". Rekomendasi pemberian diskon ini didasarkan pada fakta bahwa kombinasi produk

ini memiliki hubungan asosiasi yang kuat berdasarkan analisis aturan asosiasi. Untuk meningkatkan efektivitas strategi ini dalam meningkatkan penjualan, pihak Toko Intimart Gedongan dapat mempertimbangkan untuk menyesuaikan volume bundling. Sebagai contoh, mereka dapat menawarkan berbagai pilihan bundling, seperti paket bundling untuk keluarga dengan volume lebih besar, seperti 2 paket produk bundling ini, atau paket untuk pelanggan individu dengan volume yang lebih kecil, seperti 1 paket produk bundling. Diskon yang diberikan dapat didasarkan pada margin keuntungan produk bundling tersebut, misalnya, dengan memberikan diskon sebesar 5% dari margin keuntungan. Dalam penelitian *“The effects of price discount and product complementarity on consumer evaluations of bundle components”*, Menggunakan bundling produk dan memberikan diskon kepada pelanggan dapat memiliki efek positif. Diskon harga bundel dapat meningkatkan evaluasi positif dari konsumen terhadap bundel secara keseluruhan, meningkatkan penjualan, dan menarik pelanggan. Hal ini terjadi karena konsumen melihat bundel sebagai penawaran yang memberikan nilai lebih dibandingkan dengan membeli produk-produk secara individual (Sheng et al., 2007).

5. Peneliti merekomendasikan pihak Toko Intimart Gedongan untuk memberikan program poin atau *voucher* belanja kepada pelanggan. Program poin dapat diperoleh ketika pelanggan berbelanja dengan minimal pembelian sebesar Rp. 170.000,- dan kelipatannya. Angka tersebut didapatkan dari rata-rata pembelian yang dilakukan oleh pelanggan dalam kluster 4. Tujuan dari program poin ini adalah untuk mendorong pelanggan yang termasuk dalam kluster 2 agar dapat menjadi pelanggan yang masuk ke dalam kluster 4 (*“High-Value Customers”*). Strategi ini juga dilakukan pada penelitian berjudul *“Peningkatan Strategi Penjualan dengan metode Association Rule Pada Toko Ritel ABC”* yang dilakukan oleh Rahmattullah & Yanti (2022).
6. Peneliti merekomendasikan untuk membuat diskon atau promo menarik pada produk *bundling* yang memiliki frekuensi pembelian rendah dengan kombinasi produk yang memiliki hubungan asosiasi yang kuat. Seperti menggabungkan produk *“Absolute Green Tea 60 ml”* (produk dengan frekuensi pembelian rendah) dengan kombinasi aturan asosiasi 2 (*“Indomie Gor Spc 84gr”*, *“Indomie Gor Aceh 90gr”*, dan *“Telur Ay.Ras Curah Gr”*). Promo ini bertujuan untuk memberikan motivasi atau pematik kepada pelanggan agar tertarik untuk membeli produk yang kurang laku atau jarang dibeli oleh pelanggan lain. Penerapan strategi di atas juga diterapkan pada penelitian berjudul *“Penerapan Market Basket Analysis Menggunakan Proses KDD (Knowledge Discovery in Database) Sebagai*

*Strategi Penjualan Produk Swalayan (Studi Kasus: Swalayan X)*” yang ditulis oleh Alfiqra & Alfizi (2018).

7. Peneliti merekomendasikan untuk membuat bonus pack bagi produk-produk yang memiliki frekuensi pembelian rendah ketika konsumen membeli produk-produk asosiasi tinggi. Contohnya, dengan setiap pembelian 3 bundle produk ("Mie Gelas Grg Sosis 28g", "Indomie Gor Rendang 91gr", "Indomie Gor Aceh 90gr", dan "Telur Ay.Ras Curah Gr"), konsumen akan mendapatkan gratis produk dengan frekuensi rendah seperti "Absolute Green Tea 60 MI". Melalui langkah ini, toko dapat mendorong pembelian produk dengan frekuensi rendah dengan memberikan insentif kepada pelanggan untuk mencoba produk tersebut, yang pada akhirnya dapat meningkatkan penjualan produk yang dibeli oleh pelanggan. strategi tersebut juga telah dilakukan pada penelitian "*Pengaruh Price Discount Dan Bonus Pack Terhadap Impulse Buying Konsumen Indomaret di Kecamatan Sukun Kota Malang*" yang ditulis oleh Wilujeng (2017). Penelitian tersebut menyatakan Bonus pack secara positif dan signifikan berpengaruh terhadap pembelian konsumen secara *impulse buying* di Indomaret Kecamatan Sukun Kota Malang.
8. Peneliti merekomendasikan untuk membuat rak khusus yang menampilkan produk-produk yang telah ditentukan untuk dilakukan bundling, bersama dengan barang-barang yang memiliki frekuensi pembelian tinggi dan rendah, seperti menempatkan produk "Telur Ay.Ras Curah Gr" dan "Absolute Green Tea 60 MI" di depan toko atau dekat area kasir. Tujuannya adalah untuk meningkatkan kesadaran akan tampilan produk, yang dapat membantu meningkatkan penjualan dan memperkuat strategi *bundling*. Strategi tersebut telah diteliti pada penelitian berjudul "*Pengaruh Price Discount, Bonus Pack, Dan In-Store Display Terhadap Keputusan Impulse Buying Pada Giant Ekstra Banjar*" yang ditulis oleh Faisal (2018). Penelitian tersebut menyatakan bahwa Variabel *price discount, bonus pack, in-store display* berpengaruh secara simultan terhadap keputusan *impulse buying* pada Giant Ekstra Kabupaten Banjar Kalimantan Selatan.



## BAB VI

### KESIMPULAN DAN SARAN

#### 6.1 Kesimpulan

Berdasarkan hasil pengolahan data dan pembahasan yang peneliti lakukan, berikut merupakan kesimpulan dari penelitian ini.

1. Berdasarkan grafik *elbow method* dan *silhouette score*, didapatkan jumlah kluster sebanyak 4 kluster untuk mengelompokkan data transaksi pelanggan. Berdasarkan analisis hasil *clustering* menggunakan algoritma *K-Means* didapatkan proporsi kluster 1 sebanyak 40.4% dengan 12825 transaksi, kluster 2 sebesar 13.2% dengan 4191 transaksi, kluster 3 sebesar 44.7% dengan 14172 transaksi, dan kluster 4 sebesar 1.7% dengan 547 transaksi. Berikut merupakan penjelasan (*profiling*) dari tiap cluster:

- a. Kluster 1 (*Mid-Morning Moderates*)

Kluster "*Mid-Morning Moderates*" merupakan kelompok pelanggan dengan kecenderungan bertransaksi pada jam 8 pagi hingga jam 12 siang. Pelanggan dalam kluster ini cenderung membeli produk yang sama atau memiliki sedikit variasi produk, dengan rentang 1 sampai 2 varian produk. Nilai transaksi pelanggan berada dalam kisaran sedang atau menengah, dengan rata-rata Rp. 19.866,- dan rentang total transaksi antara Rp. 8.925,- sampai Rp. 26.540,-.

- b. Kluster 2 (*Diverse Afternoon Buyers*)

Kluster "*Diverse Afternoon Buyers*" merupakan kelompok pelanggan dengan kecenderungan bertransaksi pada jam 1 siang hingga jam 7 malam. Pelanggan dalam kluster ini cenderung membeli berbagai variasi produk, dengan rentang 4 hingga 5 variasi produk per transaksi. Nilai transaksi pelanggan berada dalam kisaran cukup tinggi, dengan rata-rata Rp. 49.271,- dan rentang total transaksi antara Rp. 31.000,- hingga Rp. 63.434,-. Kluster ini mencerminkan pelanggan yang aktif dan beragam dalam pembelian produk, khususnya pada waktu siang hingga sore hari. Sebagai "*Diverse Afternoon Buyers*," mereka menjadi target strategis untuk peningkatan penjualan dengan mengedepankan promosi produk bervariasi dan meningkatkan nilai transaksi.

- c. Kluster 3 (*Evening Moderates*)

Kluster "*Evening Moderates*" merupakan kelompok pelanggan dengan kecenderungan bertransaksi pada jam 5 sore hingga jam 8 malam. Pelanggan dalam kluster ini

cenderung membeli produk dengan sedikit variasi. Dengan rentang 1 hingga 2 variasi produk per transaksi. Nilai transaksi pelanggan berada dalam kisaran sedang atau menengah, dengan rata-rata Rp. 19.591,- dan rentang total transaksi antara Rp. 9.500,- hingga Rp. 26.400,-. Kluster ini memiliki kesamaan dengan kluster 1, di mana keduanya memiliki ciri-ciri pelanggan yang membeli produk dengan variasi yang terbatas.

d. Kluster 4 (*High-Value Customers*)

Kluster "*High-Value Customers*" merupakan kelompok pelanggan dengan kecenderungan bertransaksi pada jam 10 pagi hingga jam 6 sore. Pelanggan dalam kluster ini cenderung membeli produk dengan berbagai variasi yaitu dengan rata-rata 4.7 varian produk per transaksi. Nilai transaksi pelanggan berada dalam kisaran sangat tinggi, dengan rata-rata Rp. 177.475,- dan rentang antara Rp. 128.050,- hingga Rp. 209.883,-. Kluster ini menjadi kelompok pelanggan yang sangat berharga karena memiliki kontribusi yang signifikan pada pendapatan toko.

2. Berdasarkan hasil pengolahan data menggunakan metode *Association Rule Market Basket Analysis* algoritma *FP-Growth* dengan bantuan *library mlxtend* dalam bahasa pemrograman *Python*, didapatkan 104 aturan asosiasi pada kluster 2. Berikut merupakan 10 aturan asosiasi yang memiliki nilai *support* terbesar dan memiliki nilai *lift ratio* lebih dari 1:

- a. Jika pelanggan membeli "Indomie Gor Spc 84gr", kemungkinan mereka juga akan membeli "Telur Ay.Ras Curah Gr".
- b. Jika pelanggan membeli "Indomie Gor Spc 84gr" dan "Indomie Gor Aceh 90gr", kemungkinan mereka juga akan membeli "Telur Ay.Ras Curah Gr".
- c. Jika Pelanggan membeli "Mie Gelas Grg Sosis 28g", kemungkinan mereka juga akan membeli "Indomie Gor Rendang 91gr", "Indomie Gor Aceh 90gr", dan "Telur Ay.Ras Curah Gr".
- d. Jika pelanggan membeli "Mie Gelas Grg Sosis 28g", kemungkinan mereka juga akan membeli "Indomie Gor Rendang 91gr" dan "Indomie Gor Aceh 90gr".
- e. Jika pelanggan membeli "Indomie Gor Spc 84gr" dan "Sedaap Soto 72gr", kemungkinan mereka juga akan membeli "Telur Ay.Ras Curah Gr".
- f. Jika pelanggan membeli "Bango Kecap Mns 220ml", kemungkinan mereka juga akan membeli "Telur Ay.Ras Curah Gr".

- g. Jika pelanggan membeli "Delfi Funt.Mlk Vnl26", kemungkinan mereka juga akan membeli "Gery Choco.Grande 17gr".
  - h. Jika pelanggan membeli "Gg Cig Patra 12", kemungkinan mereka juga akan membeli "Telur Ay.Ras Curah Gr".
  - i. Jika pelanggan membeli "Indomie Gor Geprek 85gr" dan "Indomie Gor Aceh 90gr", kemungkinan mereka juga akan membeli "Telur Ay.Ras Curah Gr".
  - j. Jika pelanggan membeli "Sedaap Kari Spc 75gr" dan "Sedaap Baso Special 77gr", kemungkinan mereka juga akan membeli "Telur Ay.Ras Curah Gr".
3. Berdasarkan analisis *Clustering K-means* dan *AR-MBA FP-Growth*, berikut merupakan rekomendasi yang bisa diberikan peneliti kepada pihak Toko Intimart Gedongan sebagai strategi peningkatan penjualan:
- a. Berdasarkan 4 kluster yang telah dibentuk menggunakan algoritma *K-Means* peneliti menargetkan kluster 2 sebagai prioritas strategi peningkatan penjualan di Toko Intimart Gedongan. Pemilihan kluster kedua didasarkan pada nilai transaksi yang tinggi, dengan rata-rata nilai sebesar Rp. 49.271,- per transaksi. Selain itu, pelanggan dalam kluster ini cenderung membeli berbagai variasi produk, yaitu dengan persebaran 4 hingga 5 variasi produk, yang mendukung analisis aturan asosiasi. Selain itu, ditemukan bahwa kluster ini aktif bertransaksi pada rentang waktu antara jam 13.00 hingga jam 19.00, dengan puncak transaksi terjadi sekitar jam 14.00 hingga 15.00. Oleh karena itu, peneliti merekomendasikan Toko Intimart Gedongan untuk memfokuskan strategi peningkatan penjualan kepada pelanggan yang termasuk dalam kluster 2.
  - b. Peneliti merekomendasikan kepada pihak toko Intimart Gedongan untuk melakukan promosi khusus, diskon, atau penawaran menarik seperti produk *bundling* pada jam sibuk ketika customer dari kluster 2 berbelanja, yaitu pada rentang waktu jam 13.00 hingga jam 19.00. Dengan melakukan promosi pada periode tersebut, tidak hanya kluster 2 tetapi juga kluster 3 dan kluster 4 dapat ikut merasakan manfaat dari promosi yang dilakukan. Hal ini akan membantu menarik lebih banyak pelanggan, meningkatkan jumlah transaksi, dan memperluas pangsa pasar untuk semua kluster pelanggan yang berbelanja di toko.
  - c. Berdasarkan frekuensi pembelian pelanggan, terlihat bahwa produk "Telur Ay.Ras Curah Gr" menjadi produk yang paling diminati dengan frekuensi pembelian yang tinggi. Selanjutnya, berdasarkan analisis aturan asosiasi pada 10 aturan dengan *support* tertinggi, produk "Telur Ay.Ras Curah Gr" selalu termasuk dalam kombinasi pembelian pelanggan. Oleh karena itu, peneliti merekomendasikan kepada pihak Toko Intimart

Gedongan untuk selalu menyediakan stok yang cukup untuk produk "Telur Ay.Ras Curah Gr".

- d. Peneliti merekomendasikan untuk memberikan diskon pada produk bundling yang terbentuk dari aturan ke-3, yaitu dengan mengombinasikan produk "Mie Gelas Grg Sosis 28g", "Indomie Gor Rendang 91gr", "Indomie Gor Aceh 90gr", dan "Telur Ay.Ras Curah Gr". Rekomendasi pemberian diskon ini didasarkan pada fakta bahwa kombinasi produk ini memiliki hubungan asosiasi yang kuat berdasarkan analisis aturan asosiasi. Untuk meningkatkan efektivitas strategi ini dalam meningkatkan penjualan, pihak Toko Intimart Gedongan dapat mempertimbangkan untuk menyesuaikan volume bundling. Sebagai contoh, mereka dapat menawarkan berbagai pilihan bundling, seperti paket bundling untuk keluarga dengan volume lebih besar, seperti 2 paket produk bundling ini, atau paket untuk pelanggan individu dengan volume yang lebih kecil, seperti 1 paket produk bundling. Diskon yang diberikan dapat didasarkan pada margin keuntungan produk bundling tersebut, misalnya, dengan memberikan diskon sebesar 5% dari margin keuntungan. Dalam penelitian "*The effects of price discount and product complementarity on consumer evaluations of bundle components*", Menggunakan bundling produk dan memberikan diskon kepada pelanggan dapat memiliki efek positif. Diskon harga bundel dapat meningkatkan evaluasi positif dari konsumen terhadap bundel secara keseluruhan, meningkatkan penjualan, dan menarik pelanggan. Hal ini terjadi karena konsumen melihat bundel sebagai penawaran yang memberikan nilai lebih dibandingkan dengan membeli produk-produk secara individual (Sheng et al., 2007).
- e. Peneliti merekomendasikan pihak Toko Intimart Gedongan untuk memberikan program poin atau *voucher* belanja kepada pelanggan. Program poin dapat diperoleh ketika pelanggan berbelanja dengan minimal pembelian sebesar Rp. 170.000,- dan kelipatannya. Angka tersebut didapatkan dari rata-rata pembelian yang dilakukan oleh pelanggan dalam kluster 4. Tujuan dari program poin ini adalah untuk mendorong pelanggan yang termasuk dalam kluster 2 agar dapat menjadi pelanggan yang masuk ke dalam kluster 4 ("*High-Value Customers*"). Strategi ini juga dilakukan pada penelitian berjudul "*Peningkatan Strategi Penjualan dengan metode Association Rule Pada Toko Ritel ABC*" yang dilakukan oleh Rahmattullah & Yanti (2022).
- f. Peneliti merekomendasikan untuk membuat diskon atau promo menarik pada produk *bundling* yang memiliki frekuensi pembelian rendah dengan kombinasi produk yang memiliki hubungan asosiasi yang kuat. Seperti menggabungkan produk "Absolute

Green Tea 60 MI” (produk dengan frekuensi pembelian rendah) dengan kombinasi aturan asosiasi 2 (“Indomie Gor Spc 84gr”, “Indomie Gor Aceh 90gr”, dan “Telur Ay.Ras Curah Gr”). Promo ini bertujuan untuk memberikan motivasi atau pematik kepada pelanggan agar tertarik untuk membeli produk yang kurang laku atau jarang dibeli oleh pelanggan lain. Penerapan strategi diatas juga diterapkan pada penelitian berjudul “*Penerapan Market Basket Analysis Menggunakan Proses KDD (Knowledge Discovery in Database) Sebagai Strategi Penjualan Produk Swalayan (Studi Kasus: Swalayan X)*” yang ditulis oleh Alfiqra & Alfizi (2018).

- g. Peneliti merekomendasikan untuk membuat bonus pack bagi produk-produk yang memiliki frekuensi pembelian rendah ketika konsumen membeli produk-produk asosiasi tinggi. Contohnya, dengan setiap pembelian 3 bundle produk ("Mie Gelas Grg Sosis 28g", "Indomie Gor Rendang 91gr", "Indomie Gor Aceh 90gr", dan "Telur Ay.Ras Curah Gr"), konsumen akan mendapatkan gratis produk dengan frekuensi rendah seperti "Absolute Green Tea 60 MI". Melalui langkah ini, toko dapat mendorong pembelian produk dengan frekuensi rendah dengan memberikan insentif kepada pelanggan untuk mencoba produk tersebut, yang pada akhirnya dapat meningkatkan penjualan produk yang dibeli oleh pelanggan. strategi tersebut juga telah dilakukan pada penelitian “*Pengaruh Price Discount Dan Bonus Pack Terhadap Impulse Buying Konsumen Indomaret di Kecamatan Sukun Kota Malang*” yang ditulis oleh Wilujeng (2017). Penelitian tersebut menyatakan Bonus pack secara positif dan signifikan berpengaruh terhadap pembelian konsumen secara *impulse buying* di Indomaret Kecamatan Sukun Kota Malang.
- h. Peneliti merekomendasikan untuk membuat rak khusus yang menampilkan produk-produk yang telah ditentukan untuk dilakukan bundling, bersama dengan barang-barang yang memiliki frekuensi pembelian tinggi dan rendah, seperti menempatkan produk "Telur Ay.Ras Curah Gr" dan "Absolute Green Tea 60 MI" di depan toko atau dekat area kasir. Tujuannya adalah untuk meningkatkan kesadaran akan tampilan produk, yang dapat membantu meningkatkan penjualan dan memperkuat strategi *bundling*. Strategi tersebut telah diteliti pada penelitian berjudul “*Pengaruh Price Discount, Bonus Pack, Dan In-Store Display Terhadap Keputusan Impulse Buying Pada Giant Ekstra Banjar*” yang ditulis oleh Faisal (2018). Penelitian tersebut menyatakan bahwa Variabel *price discount, bonus pack, in-store display* berpengaruh secara simultan terhadap keputusan *impulse buying* pada Giant Ekstra Kabupaten Banjar Kalimantan Selatan.

## 6.2 Saran

Berikut merupakan saran yang dapat diberikan berdasarkan penelitian yang telah dilakukan, sebagai berikut:

### 1. Saran untuk pihak Toko Intimart Gedongan

Saran yang dapat peneliti berikan kepada pihak Toko Intimart Gedongan adalah untuk meninjau dan mengembangkan strategi pemasaran yang telah disusun agar dapat meningkatkan penjualan dari toko Intimart Gedongan. Selain itu, peneliti juga merekomendasikan agar pihak Intimart Gedongan mulai mengimplementasikan program *membership*, yang dapat memanfaatkan data pelanggan untuk melakukan *clustering* yang lebih efektif. Dengan program *membership*, pihak toko dapat memberikan promo dan insentif secara lebih personal kepada pelanggan, sehingga dapat meningkatkan minat pelanggan untuk berbelanja lebih banyak dan lebih sering.

### 2. Saran bagi peneliti selanjutnya

Saran yang dapat diberikan untuk peneliti selanjutnya adalah melakukan analisis *clustering* menggunakan data pelanggan dan menghubungkannya dengan analisis pola pembelian melalui keranjang belanja. Penelitian lebih lanjut dapat fokus untuk menggali faktor-faktor yang mempengaruhi pembelian produk *bundling* dan frekuensi pembelian produk yang rendah. Selain itu, peneliti diharapkan mempertimbangkan penggunaan algoritma lainnya dalam melakukan *clustering* dan aturan asosiasi untuk mengidentifikasi pola pembelian yang lebih kompleks dan akurat. Dengan demikian, penelitian selanjutnya dapat memberikan wawasan lebih mendalam tentang preferensi pelanggan, hubungan asosiasi produk, serta strategi pemasaran yang lebih efektif bagi toko atau perusahaan terkait.

## DAFTAR PUSTAKA

- Rizaty, Monavia Ayu. 2022. Jumlah Toko retail Indonesia mencapai 3,61 Juta Pada 202. (online): <https://dataindonesia.id/sektor-riil/detail/jumlah-toko-retail-indonesia-mencapai-361-juta-pada-2021> (17 Maret 2023)
- Fitriyani, E. 2023. Hippiindo Ungkap pemicu gerai ritel berguguran: Tempat kegedean, Kumparan. (online): <https://kumparan.com/kumparanbisnis/hippindo-ungkap-pemicu-gerai-ritel-berguguran-tempat-kegedean-1zyVPyvpBIH/3> (12 Maret 2023)
- Nuraini. 2022. 7 peluang industri FMCG di Indonesia untuk Terus Tumbuh di 2022. (online): <https://blog.usetada.com/id/peluang-industri-fmcg-di-indonesia-2022> (15 Mei 2023)
- Achmad, F., Nurdiawan, O., & Wijaya, Y. A. (2023). ANALISA POLA TRANSAKSI PEMBELIAN KONSUMEN PADA TOKO RITEL KESEHATAN MENGGUNAKAN ALGORITMA FP-GROWTH. *Jurnal Mahasiswa Teknik Informatika*, 7(1), 168–175. [www.researchgate.net](http://www.researchgate.net)
- Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-14142-8>
- Alfiqra, A., & Alfizi, F. Y. (2018). *Penerapan Market Basket Analysis Menggunakan Proses KDD (Knowledge Discovery in Database) Sebagai Strategi Penjualan Produk Swalayan (Studi Kasus: Swalayan X)*.
- Almahenzar, B., & Wijayanto, A. W. (2022). Analisis Intensitas Hujan Provinsi Jawa Barat Tahun 2020 Menggunakan Association Rule Apriori dan FP-Growth. *Journal of System and Computer Engineering (JSCE)*, 3(2), 258–271.
- Amelia, R., & Utomo, D. P. (2019). ANALISA POLA PEMESANAN PRODUK MODERN TRADE INDEPENDENT DENGAN MENEREPAKAN ALGORITMA FP. GROWTH (STUDI KASUS: PT. ADAM DANI LESTARI). *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1). <https://doi.org/10.30865/komik.v3i1.1622>
- Anggrawan, A., Mayadi, M., & Satria, C. (2021). Menentukan Akurasi Tata Letak Barang dengan Menggunakan Algoritma Apriori dan Algoritma FP-Growth. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(1), 125–138. <https://doi.org/10.30812/matrik.v21i1.1260>
- Ariestya, W. W., Supriyatin, W., & Astuti, I. (2019). Marketing strategy for the determination of staple consumer products using FP-growth and apriori algorithm. *Jurnal Ilmiah Ekonomi Bisnis*, 24(3), 225–235.
- Auliasari, K., & Kertaningtyas, M. (2019). Penerapan Algoritma K-Means untuk Segmentasi Konsumen Menggunakan R. *Jurnal Teknologi & Manajemen Informatika*, 5(1).
- Baskoro, D. A., Ambarwati, L., Wicaksana, I. W. S., & C, D. A. (2013). *Belajar Data Mining dengan Rapid Miner* (R. Sanjaya, Ed.).
- Bholowalia, P., & Kumar, A. (2014). EBK-means: A clustering technique based on elbow method and k-means in WSN. *International Journal of Computer Applications*, 105(9).

- Bunda, Y. P. (2020a). ALGORITMA FP-GROWTH UNTUK MENGANALISA POLA PEMBELIAN OLEH-OLEH (STUDI KASUS DI PUSAT OLEH-OLEH UMMI AUFA HAKIM). *Riau Journal of Computer Science*, 06(01), 34–44.
- Bunda, Y. P. (2020b). ALGORITMA FP-GROWTH UNTUK MENGANALISA POLA PEMBELIAN OLEH-OLEH (STUDI KASUS DI PUSAT OLEH-OLEH UMMI AUFA HAKIM). *Riau Journal of Computer Science*, 06(01), 34–44.
- Butryn, B., Chomiak-Orsa, I., Hauke, K., Pondel, M., & Siennicka, A. (2021). Application of Machine Learning in medical data analysis illustrated with an example of association rules. *Procedia Computer Science*, 192, 3134–3143.
- Chaniago, H. (2021). *MANAJEMEN RITEL & IMPLEMENTASINYA* (Y. Efawati, Ed.). EDUKASI RISET DIGITAL.
- Choi, H. W., Qureshi, N. M. F., & Shin, D. R. (2019). Analysis of electricity consumption at home using K-means clustering algorithm. *2019 21st International Conference on Advanced Communication Technology (ICACT)*, 639–643.
- Dawane, V., Waghodekar, P., & Pagare, J. (2021). RFM Analysis Using K-Means Clustering to Improve Revenue and Customer Retention. *International Conference on Smart Data Intelligence (ICSMDI 2021)* . <https://ssrn.com/abstract=3852887>
- Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., & Jajuli, M. (2022). Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi*, 12(1), 64–77.
- Elisa, E. (2018). *Market Basket Analysis Pada Mini Market Ayu Dengan Algoritma Apriori*. 2(2), 472–478. <http://jurnal.iaii.or.id>
- Erwin. (2009). Analisis Market Basket Dengan Algoritma Apriori dan FP-Growth. *Jurnal Generic*, 4(2), 26–30.
- Faisal, I. (2018). Pengaruh Price Discount, Bonus Pack, Dan In-Store Display Terhadap Keputusan Impulse Buying Pada Giant Ekstra Banjar. *Jurnal Sains Manajemen Dan Kewirausahaan*, 2(1), 51–60.
- Fathurrahman, M., Rizky Pratama, A., & Al-Mudzakir, T. (2023). Perbandingan Algoritma Apriori Dan Fp Growth Terhadap Market Basket Analysis Pada Data Penjualan Bakery. *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer Dan Manajemen)*, 4(2), 266–274.
- Fitria, R., Nengsih, W., & Qudsi, D. H. (2017). Implementasi Algoritma FP-Growth Dalam Penentuan Pola Hubungan Kecelakaan Lalu Lintas. *Jurnal Sistem Informasi*, 13(2), 118. <https://doi.org/10.21609/jsi.v13i2.551>
- Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). Clustering algorithms and validity measures. *Proceedings of the International Conference on Scientific and Statistical Database Management, SSDBM*, 3–22. <https://doi.org/10.1109/ssdm.2001.938534>



- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)* (3rd ed.). Morgan Kaufmann .
- Idris, A. I., Sampetoding, E. A. M., Ardhana, V. Y. P., Maritsa, I., Sakri, A., Ruslan, H., & Manapa, E. S. (2022). Comparison of Apriori, Apriori-TID and FP-Growth Algorithms in Market Basket Analysis at Grocery Stores. *The IJICS (International Journal of Informatics and Computer Science)*, 6(2), 107–112.
- Irwanto, I., Purwananto, Y., & Soelaiman, R. (2012). Optimasi Kinerja Algoritma Klusterisasi K-Means untuk Kuantisasi Warna Citra. *Jurnal Teknik ITS*, 1(1), A197–A202.
- Kamila, I., Khairunnisa, U., & Mustakim. (2019a). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 5(1), 119–125.
- Kamila, I., Khairunnisa, U., & Mustakim. (2019b). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 5(1), 119–125.
- Kantardzic, M. (2020). *Data Mining: Concepts, Models, Methods, and Algorithms* (3rd ed.). Wiley-IEEE Press.
- Kaur, M., & Kang, S. (2016). Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining. *Procedia Computer Science*, 85, 78–85. <https://doi.org/10.1016/j.procs.2016.05.180>
- Kavitha, M., & Selvi, T. (2016). Comparative Study on Apriori Algorithm and Fp Growth Algorithm with Pros and Cons. *International Journal of Computer Science Trends and Technology (IJCS T)*, 4(4). [www.ijcstjournal.org](http://www.ijcstjournal.org)
- Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal*, 1(6), 90–95.
- Kurnia, Y., Isharianto, Y., Giap, Y. C., & Hermawan, A. (2019). Study of application of data mining market basket analysis for knowing sales pattern (association of items) at the O! Fish restaurant using apriori algorithm. *Journal of Physics: Conference Series*, 1175(1), 012047.
- Larasati, D. P., Nasrun, M., & Ahmad, U. A. (2015). ANALISIS DAN IMPLEMENTASI ALGORITMA FP-GROWTH PADA APLIKASI SMART UNTUK MENENTUKAN MARKET BASKET ANALYSIS PADA USAHA RETAIL (STUDI KASUS : PT.X). *E-Proceeding of Engineering*, 2(1).
- Larose, D. T., & Larose, C. D. (2015). *DATA MINING AND PREDICTIVE ANALYTICS* (D. T. Larose, Ed.; 2nd ed.). Wiley.
- Li, Y., & Wu, H. (2012). A Clustering Method Based on K-Means Algorithm. *Physics Procedia*, 25, 1104–1109. <https://doi.org/10.1016/j.phpro.2012.03.206>

- Liew, H. P. (2018). Dietary habits and physical activity: Results from cluster analysis and market basket analysis. *Nutrition and Health*, 24(2), 83–92. <https://doi.org/10.1177/0260106018770942>
- Lindstrom, G. (2005). Programming with python. *IT Professional*, 7(5), 10–16.
- Mamahit, N., & Qoiriah, A. (2019). Penerapan Algoritma Fp-Growth dan K-Means pada Data Transaksi Minimarket. *JINACS*, 01(02), 78–83.
- Maraghi, M., Adibi, M. A., & Mehdizadeh, E. (2020). Using RFM model and market basket analysis for segmenting customers and assigning marketing strategies to resulted segments. *Journal of Applied Intelligent Systems and Information Sciences*, 1(1), 35–43.
- Muslim, M. A., Prasetyo, B., Mawarni, E. L. H., Herowati, A. J., Mirqotussa'adah, Rukmana, S. H., & Nurzahputra, A. (2019). *Data Mining Algoritma C4.5 Disertai contoh kasus dan penerapannya dengan program computer* (E. Listiana & N. Cahyani, Eds.; Vol. 1).
- Nagpal, A., & Gabrani, G. (2019). Python for data analytics, scientific and technical applications. *2019 Amity International Conference on Artificial Intelligence (AICAI)*, 140–145.
- Nainggolan, R., Perangin-angin, R., Simarmata, E., & Tarigan, A. F. (2019). Improved the performance of the K-means cluster using the sum of squared error (SSE) optimized by using the Elbow method. *Journal of Physics: Conference Series*, 1361(1), 012015.
- Naufal, F., Chrisnanto, Y. H., & Ningsih, A. K. (2022). Sistem Rekomendasi Penawaran Produk Pada Online Shop Menggunakan K-Means Clustering. *Informatics and Digital Expert (INDEX)*, 4(1), 10–17.
- Nisa, K. (2020). ANALISIS PERILAKU KONSUMEN DALAM MEMILIH TEMPAT BELANJA (Studi Kasus Minimarket Dan Toko Kelontong Di Kelurahan Kali Rungkut Kecamatan Rungkut Kota Surabaya). *Jurnal Inovasi Penelitian*, 1(7), 1281–1288.
- Nurchahyo, R. (2019). *PENGELOLAAN BISNIS RITEL*. PT. Kuantum Buku Sejahtera.
- Nurwahdania, S., Nur, I. M., & Haris, M. A. (2021). *MARKET BASKET ANALYSIS MENGGUNAKAN ALGORITMA ASSOCIATION RULES NETWORKS DI SYENNY MART*. <http://repository.unimus.ac.id>
- Octavia, S. F., Mustakim, M., Permana, I., & Monalisa, S. (2023). Penerapan Algoritma Association Rules Dalam Penentuan Pola Pembelian Berdasarkan Hasil Clustering. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 7(3), 956–965.
- Ogbuabor, G., & Ugwoke, F. N. (2018). Clustering algorithm for a healthcare dataset using silhouette score value. *Int. J. Comput. Sci. Inf. Technol*, 10(2), 27–37.
- Oliphant, T. E. (2007). Python for scientific computing. *Computing in Science & Engineering*, 9(3), 10–20.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, 12, 2825–2830.

- Rahmattullah, R., & Yanti, R. (2022). Peningkatan Strategi Penjualan dengan metode Association Rule Pada Toko Ritel ABC. *Seminar Dan Konferensi Nasional IDEC 2022*, 1–6.
- Ramadana, W. D., Satyahadewi, N., & Perdana, H. (2020). Penerapan Market Basket Analysis Pada Pola Pembelian Barang Oleh Konsumen Menggunakan Metode Algoritma Apriori. *Bimaster: Buletin Ilmiah Matematika, Statistika Dan Terapannya*, 11(3).
- Rani, L. N., Defit, S., & Muhammad, L. J. (2021). Determination of Student Subjects in Higher Education Using Hybrid Data Mining Method with the K-Means Algorithm and FP Growth. *International Journal of Artificial Intelligence Research*, 5(1). <https://doi.org/10.29099/ijair.v5i1.223>
- Raschka, S. (2018). MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *Journal of Open Source Software*, 3(24), 638.
- Sani, A. (2018). PENERAPAN METODE K-MEANS CLUSTERING PADA PERUSAHAAN. *Jurnal Ilmiah Teknologi Informasi* 353, 1–7.
- Setiawan, A., & Gita Anugrah, I. (2019). Penentuan Pola Pembelian Konsumen pada Indomaret GKB Gresik dengan Metode FP-Growth. *Jurnal Nasional Komputasi Dan Teknologi Informasi*, 02(02).
- Setiawan, Amani, H., & Tripiawan, W. (2021). PERANCANGAN SEGMENTASI PELANGGAN DENGAN METODE CLUSTERING K-MEANS DAN MODEL RFM PADA KLINIK KECANTIKAN SEOUL SECRET. *E-Proceeding of Engineering*, 8(2), 2286–2293.
- Shahapure, K. R., & Nicholas, C. (2020). Cluster quality analysis using silhouette score. *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, 747–748.
- Shaliha, K. M., Angelyna, A., Nugraha, A. A., Wahisyam, M. H., & Sandi, T. K. (2021). Implementasi K-Means Clustering pada Online Retail berdasarkan Recency, Frequency, dan Monetary (Implementation of K-Means Clustering in Online Retail based on Recency, Frequency, and Monetary). *Gunung Djati Conference Series*, 3. <https://conferences.uinsgd.ac.id/gdcs>
- Sheng, S., Parker, A. M., & Nakamoto, K. (2007). The effects of price discount and product complementarity on consumer evaluations of bundle components. *Journal of Marketing Theory and Practice*, 15(1), 53–64.
- Sodhi, P., Awasthi, N., & Sharma, V. (2019). Introduction to machine learning and its basic application in python. *Proceedings of 10th International Conference on Digital Strategies for Organizational Success*.
- Soliha, E. (2008). ANALISIS INDUSTRI RITEL DI INDONESIA. *Jurnal Bisnis Dan Ekonomi (JBE)*, 15(2), 128–142.

- Stančín, I., & Jović, A. (2019). An overview and comparison of free Python libraries for data mining and big data analysis. *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 977–982.
- Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan. *Jurnal TEKNO KOMPAK*, 15(2), 25–36.
- Syakur, M., Khotimah, B. K., Rohman, E., & Dwi Satoto, B. (2018). Using K-means algorithm and FP-growth base on FP-tree structure for recommendation customer SME. *Journal of Theoretical and Applied Information Technology*, 96, 1102–1113.
- Syukra, I., Hidayat, A., & Fauzi, M. Z. (2019). Implementation of K-Medoids and FP-Growth Algorithms for Grouping and Product Offering Recommendations. *Indonesian Journal of Artificial Intelligence and Data Mining*, 2(2), 107–115.
- Takdirillah, R. (2020). Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan. *Edumatic : Jurnal Pendidikan Informatika*, 4(1), 37–46. <https://doi.org/10.29408/edumatic.v4i1.2081>
- Ünvan, Y. A. (2021). Market basket analysis with association rules. *Communications in Statistics - Theory and Methods*, 50(7), 1615–1628. <https://doi.org/10.1080/03610926.2020.1716255>
- Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Pedregosa, F., & Mueller, A. (2015). Scikit-learn: Machine learning without learning the machinery. *GetMobile: Mobile Computing and Communications*, 19(1), 29–33.
- Wanto, A., Siregar, M. N. H., Windarto, A. P., Hartama, D., Ginantra, N. L. W. S. R., Napitupulu, D., Negara, E. S., Lubis, M. R., Dewi, S. V., & Prianto, C. (2020). *Data Mining : Algoritma dan Implementasi* (T. Limbong, Ed.). Yayasan Kita Menulis.
- Wei, J. T., Lin, S.-Y., Yang, Y.-Z., & Wu, H.-H. (2019). The application of data mining and RFM model in market segmentation of a veterinary hospital. *Journal of Statistics and Management Systems*, 22(6), 1049–1065. <https://doi.org/10.1080/09720510.2019.1565445>
- Wilujeng, S. (2017). Pengaruh Price Discount Dan Bonus Pack Terhadap Impulse Buying Konsumen Indomaret di Kecamatan Sukun Kota Malang. *Journal FEB Universitas Kanjuruhan Malang*, 457.

## LAMPIRAN

## Data Transaksi Toko Intimart Gedongan

KODE BARANG	NAMA BARANG	JUMLAH	HARGA JUAL	TOTAL	KODE FAKTUR	TANGGAL FAKTUR	JAM
010405003 9	AMO AIR MINERAL 600ML	1	2000	2000	JF000011122	11/1/2022	6:01:30
011202004 3	GG SURYA 16S	1	27900	27900	JF000021122	11/1/2022	6:22:56
010403003 3	MOGU2 MELON 320ML	1	9900	9900	JF000021122	11/1/2022	6:22:56
011202011 5	MARLBORO KRTK BIRU 12	1	9500	9500	JF000031122	11/1/2022	6:28:57
011202009 5	ESSE BANA POP 12S	1	20500	20500	JF000031122	11/1/2022	6:28:57
030701006 5	MASKER DB WARNA 5'S	1	6000	6000	JF000041122	11/1/2022	6:41:40
040201000 1	TELUR AY.RAS CURAH GR	960	24.9	23904	JF000051122	11/1/2022	6:41:59
010801002 2	INDOMIE ST.LAMONG 90GR	1	3100	3100	JF000051122	11/1/2022	6:41:59
.							
010705001 2	BANGO KECAP.MNS 60ML	1	3000	3000	JF105730123	1/31/2023	13:14:26
010702002 4	KOBE BONCABE LV30	1	1000	1000	JF105730123	1/31/2023	13:14:26
011202005 6	MAQNUM MILD 16S	1	26500	26500	JF105740123	1/31/2023	13:16:04

## Code Python

```

1. # Import Libraries untuk manipulasi struktur data
2. import pandas as pd
3. import numpy as np
4.
5. # Import Libraries untuk visualisasi data
6. import matplotlib.pyplot as plt
7.
8. # Import Libraries untuk preprocessing
9. from sklearn.preprocessing import StandardScaler
10. from mlxtend.preprocessing import TransactionEncoder
11.
12. # Import Libraries untuk clustering
13. from sklearn.cluster import KMeans
14. from sklearn.metrics import silhouette_score, silhouette_samples
15.
16. # Import Libraries untuk menggunakan fp-growth
17. from mlxtend.frequent_patterns import fpgrowth, association_rules
18.

```

```

1. # Mendefinisikan nilai k yang akan dicoba
2. k_values = range(2,10)
3.
4. # Memasukan nilai sse untuk mencari nilai k yang sesuai
5. sse = []
6.
7. # Looping untuk mencari nilai k yang sesuai
8. for k in k_values:
9.     kmeans = KMeans(n_clusters=k, random_state=42, n_init = 'auto')
10.    kmeans.fit(X_scaled)
11.    sse.append(kmeans.inertia_)
12.
13. # Melakukan visualisasi nilai elbow curve
14. plt.plot(k_values, sse, 'bo-')
15. plt.xlabel('Number of clusters (K)')
16. plt.ylabel('Sum of Squared Distances')
17. plt.title('Elbow Method')
18. plt.show();

```

```

1. # Mendefinisikan nilai k yang akan dicoba
2. k_values = range(2,10)
3.
4. # Menghitung nilai Shihouette Score
5. sil = []
6.
7. # Looping untuk mencari nilai k yang sesuai
8. for k in k_values:
9.     # Melakukan K-Means clustering
10.    kmeans = KMeans(n_clusters=k, random_state= 42, n_init = 'auto')
11.    labels = kmeans.fit(X_scaled)
12.
13.    # Menghitung Shihouette Score
14.    sil_avg = silhouette_score(X_scaled, labels.labels_, metric= 'euclidean')
15.    sil.append(sil_avg)
16.
17. # Melakukan Visualisasi nilai silhouette scores
18. plt.plot(k_values, sil, 'bo-')
19. plt.xlabel('Number of Clusters (k)')
20. plt.ylabel('Silhouette Score')
21. plt.title('Silhouette Score for Different k')
22. plt.show()

```

```

1. # Melakukan K-means pada Dataset
2. kmeans = KMeans(n_clusters= 4, random_state= 42, n_init = 'auto')
3.
4. # Melatih model
5. kmeans.fit(X_scaled)
6.
7. # Melakukan Pelabelan dengan K-means
8. labels = kmeans.predict(X_scaled)
9.
10. # Menambahkan 1 kedalam cluster labels
11. labels += 1
12.
13. # Memasukan hasil prediksi kedalam dataframe df_clean_fix
14. df_clean_fix['labels'] = labels
15.

```

```

1. def encode_fpgrowth(input_list: list) -> pd.DataFrame:
2.     """
3.     Menerapkan encode menggunakan algoritma FP-Growth pada daftar transaksi.
4.
5.     Parameter:
6.         input_list (list): Daftar transaksi, di mana setiap transaksi direpresentasikan
sebagai daftar item.
7.
8.     Kembalian:
9.         pd.DataFrame: DataFrame yang mewakili transaksi yang telah diencode, di mana setiap
kolom mewakili satu item dan setiap baris mewakili satu transaksi.
10.    """
11.    # Membuat model TransactionEncoder
12.    my_transactionencoder = TransactionEncoder()
13.
14.    # Melakukan transformasi menggunakan model
15.    encoded_transactions = my_transactionencoder.fit_transform(input_list)
16.
17.    # Mengubah array transaksi yang diencode menjadi DataFrame
18.    encoded_transactions_df = pd.DataFrame(encoded_transactions,
columns=my_transactionencoder.columns_)
19.
20.    return encoded_transactions_df
21.

```

```

1. def fpgrowth_fn(input_df: pd.DataFrame,
2.                 min_support: float = 0.01,
3.                 threshold: float = 1) -> tuple[pd.DataFrame, pd.DataFrame]:
4.     """
5.     Perform Association Rule analysis using the FP-Growth algorithm on the input DataFrame.
6.
7.     Parameters:
8.         input_df (pd.DataFrame): DataFrame containing encoded transaction data, where each
column represents an item and each row represents a transaction.
9.         min_support (float): Minimum support value used in FP-Growth. Default: MIN_SUPPORT.
10.        threshold (float): Minimum threshold used in association rule generation. Default:
THRESHOLD.
11.
12.     Returns:
13.         Tuple[pd.DataFrame, pd.DataFrame]: A tuple containing two DataFrames:
frequent_itemsets that contains frequently occurring itemsets, and association_rules that
contains the generated association rules.
14.    """
15.
16.    frequent_itemsets = fpgrowth(input_df,
17.                                min_support=min_support,
18.                                use_colnames=True)
19.    ar_itemset = association_rules(frequent_itemsets,

```

```
20.             metric='confidence',
21.             min_threshold=threshold).sort_values(by='confidence',
22.                                                 ascending=False)
23.
24.     return frequent_itemsets, ar_itemset
25.
```