

**PENERAPAN *CLUSTERING ANALYSIS* DAN *ASSOCIATE RULES*
MINING UNTUK STRATEGI PENINGKATAN PENJUALAN
BERDASARKAN POLA PEMBELIAN KONSUMEN GROBBAK
KOREA PANDEYAN**

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 Farhan Ramadhan
No. Mahasiswa : 21522301

**PROGRAM STUDI TEKNIK INDUSTRI PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2025**

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, 02 Oktober 2025



A handwritten signature in black ink, appearing to read 'Muhammad Farhan Ramadhan'.

(Muhammad Farhan Ramadhan)

21522301

SURAT BUKTI PENELITIAN



FAKULTAS
TEKNOLOGI INDUSTRI

Gedung RI, Mas Mansur
Kampus Terpadu Universitas Islam Indonesia
Jl. Kalurang km 14,5 Yogyakarta 55584
T. (0274) 898444 ext. 4100, 4101
F. (0274) 895007
E. ti@uii.ac.id
W. ti.uii.ac.id

SURAT KETERANGAN PENELITIAN

Nomor: 14/Ka.Lab.Datmin/70/Lab.Datmin/IX/2025

Assalamu'alaikum Warahmatullahi Wabarakatuh

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

Nama : Muhammad Farhan Ramadahn
No. Mhs : 21522301
Dosen Pembimbing : Annisa Uswatun Khasanah, S.T., M.Sc

Telah selesai melaksanakan penelitian yang berjudul "Penerapan Clustering Analysis Dan Associate Rules Mining Untuk Strategi Peningkatan Penjualan Berdasarkan Pola Pembelian Konsumen Grobbak Korea Pandeyan" di Laboratorium Data Mining, Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia tercatat mulai tanggal 16 Mei sampai dengan tanggal 17 Juni 2025.

Demikian surat keterangan kami keluarkan, agar dapat dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Yogyakarta 09 September 2025

Kepala Laboratorium
Data Mining

Dr. Dwi Adi Purnama, S.T.

LEMBAR PENGESAHAN PEMBIMBING
PENERAPAN CLUSTERING ANALYSIS DAN ASSOCIATE RULES
MINNING UNTUK STRATEGI PENINGKATAN PENJUALAN
BERDASARKAN POLA PEMBELIAN KONSUMEN GROBBAK
KOREA

TUGAS AKHIR

Disusun Oleh :

Nama : Muhammad Farhan Ramadhan

No. Mahasiswa : 21522301

Yogyakarta, 09 September 2025

Dosen Pembimbing



(Annisa Uswatun Khasanah, S.T., M.Sc.)

LEMBAR PENGESAHAN DOSEN PENGUJI
PENERAPAN *CLUSTERING ANALYSIS* DAN *ASSOCIATE RULES MINNING*
UNTUK STRATEGI PENINGKATAN PENJUALAN BERDASARKAN POLA
PEMBELIAN KONSUMEN GROBBAK KOREA

TUGAS AKHIR

Disusun Oleh :

Nama : Muhammad Farhan Ramadhan

No. Mahasiswa : 21522301

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, 02 Oktober 2025

Tim Penguji

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

Ketua

Putri Dwi Annisa, S.T., M.Sc.

Anggota I

Wahyudhi Sutrisno, S.T., M.M., M.T

Anggota II



Mengetahui,

Ketua Program Studi Teknik Industri Program Sarjana
Fakultas Teknologi Industri
Universitas Islam Indonesia



Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM.

NIK.015220101

HALAMAN PERSEMBAHAN

Bismillahirrahmanirrahim, Dengan menyebut nama Allah SWT, saya mengungkapkan rasa syukur atas rahmat, karunia, dan petunjuk-Nya. Dengan ini, saya persembahkan hasil penelitian Tugas Akhir ini kepada diri saya sendiri sebagai bentuk apresiasi atas perjuangan, ketekunan, dan kesabaran dalam melewati berbagai tantangan demi meraih gelar Strata Satu pada Jurusan Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia. Saya juga mempersembahkan karya ini kepada kedua orang tua tercinta yang selalu memberikan doa, kasih sayang, dan dukungan tanpa henti, serta kepada seluruh pihak yang telah membantu, membimbing, memotivasi, dan memberikan semangat terbaik selama proses studi hingga penyusunan Tugas Akhir ini.

MOTTO

"Sesungguhnya Allah tidak akan mengubah keadaan suatu kaum sampai mereka mengubah keadaan yang ada pada diri mereka sendiri."

(QS. Ar-Ra'd: 11)

"Stay hungry, stay foolish."

(Steve Jobs)

"Kemerdekaan bukan tanda untuk berhenti berjuang, tapi justru titik tolak untuk berjuang lebih keras lagi."

(Mohammad Hatta)

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh,
Segala puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan baik. Shalawat serta salam senantiasa tercurah kepada junjungan Nabi Muhammad SAW, beserta keluarga, sahabat, dan seluruh umatnya hingga akhir zaman.

Tugas Akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu (S1) pada Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia. Melalui penelitian ini, penulis berkesempatan untuk mengintegrasikan pengetahuan teoritis yang diperoleh selama perkuliahan dengan aplikasi nyata di lapangan, khususnya dalam bidang *data mining* dan pemasaran.

Dalam proses penyusunan Tugas Akhir ini, penulis telah menerima dukungan, arahan, serta bantuan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis mengucapkan terima kasih kepada:

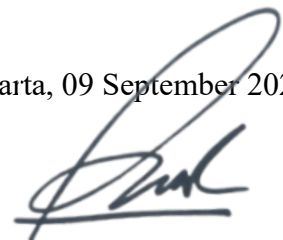
1. Bapak Prof. Dr. Ir. Hari Purnomo, M.T., IPU., ASEAN.Eng. selaku Dekan Fakultas Teknologi Industri Universitas Islam Indonesia.
2. Bapak Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM. selaku Ketua Program Studi Teknik Industri Universitas Islam Indonesia.
3. Ibu Annisa Uswatun Khasanah, S.T., M.Sc. selaku dosen pembimbing yang dengan penuh kesabaran telah memberikan bimbingan, arahan, dan ilmu yang sangat berharga.
4. Tim Grobbak Korea Pandeyan yang telah bersedia bekerja sama dan memberikan data penelitian sehingga Tugas Akhir ini dapat terlaksana.
5. Kedua orang tua tercinta yang senantiasa memberikan doa, kasih sayang, dukungan, dan motivasi tanpa henti.
6. Ketiga adik tercinta yang selalu memberikan semangat, doa, serta menjadi inspirasi dalam menyelesaikan setiap tahap perjuangan.
7. Rekan-rekan seperjuangan di Program Studi Teknik Industri UII yang selalu memberikan semangat, doa, dan kebersamaan dalam setiap langkah akademik penulis.

Penulis menyadari sepenuhnya bahwa Tugas Akhir ini masih jauh dari sempurna dan memiliki berbagai keterbatasan. Oleh karena itu, kritik dan saran yang membangun sangat di harapkan demi perbaikan di masa yang akan datang. Semoga Tugas Akhir ini dapat memberikan manfaat bagi penulis, institusi, maupun pihak lain yang membutuhkan.

Aamiin Yaa Rabbal 'Aalamiin.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Yogyakarta, 09 September 2025



Muhammad Farhan Ramadhan
NIM 21522301

ABSTRAK

Persaingan bisnis kuliner yang semakin ketat menuntut pelaku usaha untuk memahami perilaku konsumen secara lebih mendalam. Analisis berbasis data menjadi pendekatan strategis untuk mendukung pengambilan keputusan pemasaran yang tepat sasaran. Penelitian ini dilakukan pada Grobbak Korea Pandeyan dengan memanfaatkan teknik *data mining* guna mengeksplorasi pola transaksi dan menyusun strategi pemasaran yang relevan. Data penelitian berupa catatan transaksi harian yang mencakup jumlah produk, total nilai pembelian, waktu transaksi, serta platform pemesanan. Tahapan analisis meliputi *data cleaning*, transformasi variabel kategorikal menggunakan *one hot encoding*, normalisasi variabel numerik, serta penerapan algoritma *K-Means clustering* untuk mengidentifikasi segmen konsumen berdasarkan kesamaan perilaku. Selanjutnya, analisis *association rule mining* dengan metode *FP-Growth* digunakan untuk menemukan keterkaitan produk yang sering dibeli bersama. Hasil penelitian menunjukkan terbentuknya beberapa klaster konsumen dengan karakteristik yang berbeda, baik dari segi intensitas transaksi, daya beli, waktu dominan berbelanja, maupun preferensi *platform*. Berdasarkan temuan tersebut, dirumuskan strategi pemasaran berupa program loyalitas sesuai *platform*, promosi berbasis waktu transaksi, penawaran menu premium bagi segmen berdaya beli tinggi, serta strategi harga adaptif untuk segmen sensitif harga. Strategi ini diharapkan mampu meningkatkan kepuasan pelanggan sekaligus memperkuat daya saing Grobbak Korea.

Kata Kunci: *Data mining, clustering, association rules, strategi pemasaran*

DAFTAR ISI

PERNYATAAN KEASLIAN	ii
SURAT BUKTI PENELITIAN	iii
LEMBAR PENGESAHAN PEMBIMBING	iv
LEMBAR PENGESAHAN DOSEN PENGUJI	v
HALAMAN PERSEMBAHAN	vi
MOTTO	vii
KATA PENGANTAR	viii
ABSTRAK	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	5
1.5 Batasan Penelitian	5
1.6 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	7
2.1 Kajian Literatur	7
2.1.1 Clustering	7
2.1.2 Asosiasi Rules Mining	9
2.1.3 Kombinasi Clustering dan Asosiasi Rules	11
2.1.4 State of art	18
2.1.5 Posisi Penelitian	19
2.2 Landasan Teori	20
2.2.1 Food and Beverage	20
2.2.2 Data Mining	20
2.2.3 Clustering	21
2.2.4 Association Rules Mining	23
2.2.5 Market Basket Analysis	25
2.2.6 Algoritma FP-Growth	26
2.2.7 Algoritma K-Means	26
2.2.8 Silhouette Score	27
2.2.9 Segmentasi Pelanggan	30
BAB III METODE PENELITIAN	31
3.1 Objek Penelitian	31
3.2 Metode Pengumpulan Data	31
3.3 Jenis Data	31
3.4 Variabel Data	32
BAB IV PENGOLAHAN DATA	37
4.1 Pengumpulan Data	37
4.2 Pengolahan Data Clustering	38
4.2.1 Preprocessing Data sebelum Clustering	38
4.2.2 Proses Clustering	43
4.2.3 Hasil Clustering dengan algoritma K-means	48

4.3 Pengolahan Data Association Rules Mining.....	53
4.3.1 Preprocessing Dataset	53
4.3.2 Proses Association Rules Mining dengan Algoritma FP-Growth.....	56
BAB V PEMBAHASAN	64
5.1 Analisis Penerapan Metode <i>Clustering</i>	64
5.1.1 Analisis Deskripsi Data	64
5.1.2 Analisis Proporsi Anggota tiap Cluster	65
5.1.3 Analisis Karakteristik Kluster	67
5.2 Analisis Penerapan Metode <i>Association rules mining</i>	71
5.2.1 Analisis Perbandingan Minimum Support Association Rules Mining	71
5.2.2 Analisis Frekuensi Pembelian	75
5.2.3 Analisis Output Association Rules.....	80
5.3 Rekomendasi Strategi Pemasaran	86
BAB VI PENUTUP.....	88
6.1 Kesimpulan.....	88
6.2 Saran.....	89
DAFTAR PUSTAKA.....	90
LAMPIRAN	A-1

DAFTAR TABEL

Tabel 2. 1 Kajian Literatur penelitian	14
Tabel 3. 1 Deskripsi variabel pada data yang digunakan	32
Tabel 4. 1 Pengumpulan Data	37
Tabel 4. 2 Data Selection	38
Tabel 4. 3 Data Tranformasi Waktu.....	40
Tabel 4. 4 Hasil Transformasi Data	41
Tabel 4. 5 Data yang dilakukan normalisasi	42
Tabel 4. 6 Hasil pembagian Cluster	48
Tabel 4. 7 Pusat Cluster dalam Z-Score	49
Tabel 4. 8 Deskripsi Statistik dari Tiap Cluster	50
Tabel 4. 9 Rentang Persebaran Variabel Produk dan Total dari Tiap Cluster	51
Tabel 4. 10 Data Selection untuk proses Association Rules minning.....	53
Tabel 4. 11 Tampilan ID Struk yang hanya memiliki 1 baranag	55
Tabel 4. 12 Pembersihan Data.....	55
Tabel 4. 13 Data Transformation menjadi Boolean.....	55
Tabel 4. 14 perbandingan minimum support	56
Tabel 4. 15 Aturan Asosiasi klaster 1	58
Tabel 4. 16 Aturan Asosiasi Klaster 2.....	59
Tabel 4. 17 Aturan Asosiasi Klaster 3.....	60
Tabel 4. 18 Aturan Asosiasi Klaster 4.....	61
Tabel 4. 19 Aturan Asosiasi Klaster 5.....	61
Tabel 4. 20 Aturan Asosiasi Klaster 6.....	62
Tabel 5. 1 Deskripsi data jumlah produk dan transaksi	64
Tabel 5. 2 Proporsi jumlah data tiap klaster	65
Tabel 5. 3 Proporsi jumlah data tiap klaster pada platform	66
Tabel 5. 4 Proporsi jumlah data tiap klaster pada waktu	66
Tabel 5. 5 Pesebaran data tiap klaster pada jumlah produk dan total transaksi	68
Tabel 5. 6 Pesebaran data tiap klaster pada waktu	68
Tabel 5. 7 Pesebaran data tiap klaster pada platfrom	68
Tabel 5. 8 Perbandingan Minimum support klaster 1	71
Tabel 5. 9 Perbandingan Minimum support klaster 2	71
Tabel 5. 10 Perbandingan Minimum support klaster 3	72
Tabel 5. 11 Perbandingan Minimum support klaster 4	73
Tabel 5. 12 Perbandingan Minimum support klaster 5	73
Tabel 5. 13 Perbandingan Minimum support klaster 6	74
Tabel 5. 14 Top 10 Frekuensi pada klaster 1	75
Tabel 5. 15 Top 10 Frekuensi pada klaster 2	76
Tabel 5. 16 Top 10 Frekuensi pada klaster 3	77
Tabel 5. 17 Top 10 Frekuensi pada klaster 4	78
Tabel 5. 18 Top 10 Frekuensi pada klaster 5	78
Tabel 5. 19 Top 10 Frekuensi pada klaster 6	79
Tabel 5. 20 Aturan Lengkap Asosiasi Pada Klaster 1	80
Tabel 5. 21 Aturan Lengkap Asosiasi Pada Klaster 2	81
Tabel 5. 22 Aturan Lengkap Asosiasi Pada Klaster 3	82
Tabel 5. 23 Aturan Lengkap Asosiasi Pada Klaster 4	83
Tabel 5. 24 Aturan Lengkap Asosiasi Pada Klaster 5	84

Tabel 5. 25 Aturan Lengkap Asosiasi Pada Klaster 6.....	85
---	----

DAFTAR GAMBAR

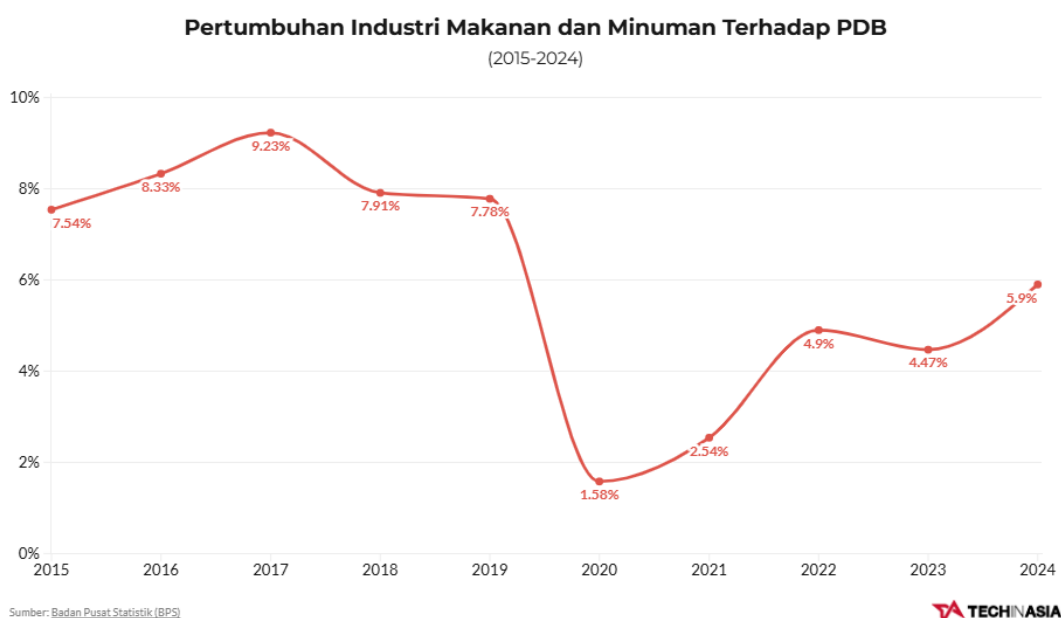
Gambar 1. 1 Data industri makanan & minuman Indonesia terhadap PDB	1
Gambar 1. 2 Pendapatan Grobbak Korea 1 tahun terakhir	3
Gambar 3. 1 Alur Penelitian.....	34
Gambar 4. 1 Identifikasi data transaksi yang bermasalah.....	39
Gambar 4. 2 Data yang akan dilakukan proses cleaning.....	40
Gambar 4. 3 Hasil Validasi Klaster.....	44
Gambar 4. 4 Grafik Elbow Methode	45
Gambar 4. 5 Grafik Silhouette Score	46
Gambar 4. 6 Grafik metode DBI.....	47

BAB I

PENDAHULUAN

1.1 Latar Belakang

Industri makanan dan minuman (F&B) merupakan salah satu sektor krusial dalam struktur perekonomian Indonesia, yang tidak hanya memberikan kontribusi signifikan terhadap Produk Domestik Bruto (PDB), tetapi juga memiliki daya tahan tinggi terhadap fluktuasi ekonomi global. Berdasarkan data terbaru dari Badan Pusat Statistik (2025) subsektor F&B mencatatkan pertumbuhan sebesar 5,9% dari tahun 2023 ke 2024. Angka ini menempatkannya sebagai salah satu dari lima subsektor industri pengolahan nonmigas dengan performa terbaik.



Gambar 1. 1 Data industri makanan & minuman Indonesia terhadap PDB

Sumber: “<https://id.techinasia.com/data-industri-makanan-minuman-indonesia-panduan-lengkap>”

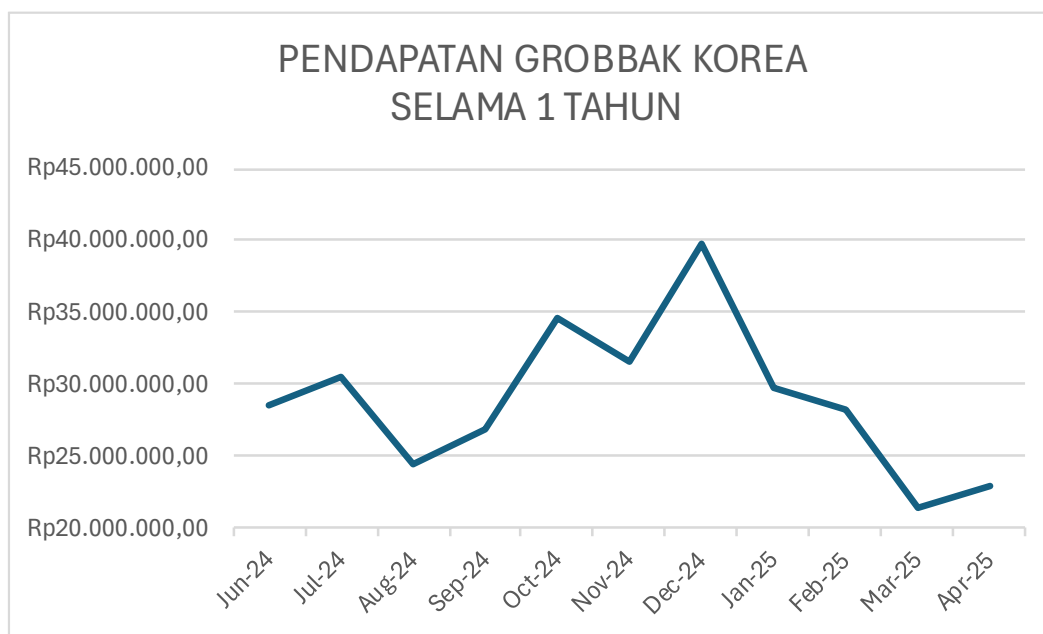
Pertumbuhan ini didorong oleh peningkatan daya beli masyarakat kelas menengah yang semakin meluas, perubahan pola konsumsi yang semakin dinamis, serta dorongan urbanisasi yang memperluas pasar untuk produk makanan dan minuman kemasan maupun siap saji (Ernst & Young, 2023). Selain itu, tren digitalisasi dalam pemesanan makanan melalui platform daring juga turut meningkatkan penetrasi pasar sektor F&B secara keseluruhan. Menurut laporan Statista (2024), Indonesia termasuk dalam lima pasar terbesar untuk makanan cepat saji di

kawasan Asia Tenggara, dengan proyeksi pertumbuhan pendapatan tahunan mencapai 8,3% hingga tahun 2027.

Kombinasi faktor demografis, sosial, dan teknologi inilah yang menjadikan sektor ini sangat menjanjikan, tetapi juga sarat tantangan. Pelaku usaha dituntut untuk mampu menyesuaikan diri secara cepat dengan dinamika pasar yang terus berubah, baik dari sisi preferensi rasa, variasi produk, harga, hingga pengalaman pelanggan yang kini turut dipengaruhi oleh keberadaan media sosial dan digital marketing. Karena itu, pendekatan berbasis data dan analitik menjadi suatu keniscayaan bagi usaha F&B yang ingin tetap relevan dan bersaing di pasar yang semakin kompetitif ini.

Namun, pertumbuhan pesat ini juga membawa tantangan baru bagi pelaku usaha, khususnya dalam menghadapi persaingan yang semakin kompetitif. Bisnis F&B kini tidak cukup hanya mengandalkan kualitas produk, melainkan perlu memahami secara mendalam preferensi dan perilaku konsumennya (Kotler & Keller, 2022). Konsumen masa kini cenderung lebih selektif, tidak hanya memperhatikan rasa dan harga, tetapi juga menilai pengalaman menyeluruh termasuk nilai emosional dan fungsional yang ditawarkan oleh merek (Solomon, 2018). Oleh karena itu, pemahaman tentang psikologi konsumen dan faktor-faktor yang mempengaruhi keputusan pembelian menjadi krusial bagi kelangsungan dan pertumbuhan bisnis F&B.

Salah satu pelaku usaha yang mengalami tantangan tersebut adalah Grobbak Korea Pandeyan di Yogyakarta. Grobbak Korea merupakan sebuah *franchise* kuliner yang telah berkembang dan membuka cabang di berbagai kota di Indonesia. Namun berdasarkan hasil wawancara dengan pemilik, dalam beberapa tahun terakhir banyak cabang terpaksa tutup karena menghadapi berbagai tantangan operasional dan tekanan persaingan. Salah satu cabang unggulan di Pandeyan telah beroperasi sejak sekitar empat tahun lalu, dan pada periode awal kehadirannya mencatat volume penjualan yang tinggi—karena menjadi satu-satunya penyedia makanan Korea di wilayah tersebut. Seiring waktu, mulai bermunculan banyak *kompetitor* serupa, sehingga penjualan di cabang ini mengalami penurunan yang cukup signifikan. Setelah periode penurunan, tingkat penjualan akhirnya berada pada kondisi yang relatif *stabil*, meski masih jauh di bawah puncak awal. Selain itu, terdapat perbedaan kinerja penjualan yang mencolok antara cabang Pandeyan dan Nologaten, yang menunjukkan karakteristik pasar lokal yang berbeda-beda. Saat ini pemilik berupaya mengevaluasi faktor-faktor penyebab fluktuasi tersebut dan merancang strategi pemasaran yang tepat agar penjualan dapat kembali meningkat.



Gambar 1. 2 Pendapatan Grobbak Korea 1 tahun terakhir

Berdasarkan data pendapatan bulanan Grobbak Korea Pandeyan selama Juni 2024 hingga April 2025, rata-rata pemasukan per bulan berada di kisaran Rp 28,95 juta. Puncak tertinggi terjadi pada Desember 2024 dengan capaian sebesar Rp 39,69 juta, sedangkan titik terendah tercatat di Maret 2025, yakni Rp 21,45 juta. Analisis fluktuasi *month-to-month* mengungkapkan bahwa pada Juli 2024 pendapatan naik sekitar 7,3 % dibandingkan Juni, tetapi di bulan berikutnya, Agustus 2024, terjadi penurunan tajam hampir 20 %. Setelah itu pendapatan kembali meningkat pada September 9,7 % dan mencapai lonjakan terbesar pada Oktober 2024, yakni 29 % dari bulan sebelumnya. November sempat turun 8,7 %, namun Desember kembali melonjak 25,8 %. Memasuki 2025, Januari mencatat penurunan drastis sebesar 25 %, Februari turun 5,3 %, dan Maret mengalami penurunan terbesar kedua hingga 23,9 %, sebelum akhirnya sedikit pulih 7,1 % pada April.

Berdasarkan masalah tersebut diperlukan sebuah *analytical tools* untuk mengetahui segmentasi pasar, yang mana merujuk pada proses membagi pasar menjadi kelompok-kelompok dengan kebutuhan dan karakteristik yang serupa, sangat penting untuk meningkatkan efektivitas promosi dan penjualan (Kotler et al., 2018). Dengan belum mengetahui segmentasi pasar inilah yang menyebabkan Grobbak korea belum mengetahui strategi pemasaran yang sesuai dengan kebutuhan spesifik dari setiap kelompok pelanggan.

Dalam konteks inilah penerapan metode *data mining* menjadi relevan. *Data mining*, yang merupakan proses ekstraksi informasi berguna dari basis data yang besar menggunakan

berbagai teknik statistik dan machine learning, terbukti efektif dalam mengolah data transaksi untuk menemukan pola tersembunyi (Han et al., 2011). Penelitian ini akan menggunakan dua metode utama, yaitu *Clustering Analysis* dengan algoritma K-Means dan association rules mining dengan algoritma FP-Growth.

Clustering Analysis digunakan untuk mengelompokkan pelanggan berdasarkan kesamaan karakteristik pembelian mereka. Algoritma K-Means dipilih karena kemampuannya dalam menangani dataset besar secara efisien serta telah banyak digunakan dalam berbagai studi untuk segmentasi pelanggan (Witten et al., 2016). Sementara itu, *association rules mining* dengan algoritma FP-Growth akan digunakan untuk mengidentifikasi kombinasi produk yang sering dibeli secara bersamaan. FP-Growth dipilih karena memiliki kecepatan komputasi yang lebih tinggi dibandingkan algoritma Apriori, terutama dalam dataset besar (Tan et al., 2019).

Kombinasi kedua metode ini telah terbukti efektif dalam berbagai penelitian ilmiah sebelumnya. Misalnya, dalam studi oleh (Kajornkasirat et al., 2025), integrasi antara metode K-Means *Clustering* dan algoritma FP-Growth diterapkan untuk segmentasi pelanggan berbasis RFM pada data transaksi retail. Hasil penelitian menunjukkan bahwa integrasi tersebut mampu mengidentifikasi tiga segmen pelanggan utama dan menemukan aturan asosiasi produk yang berbeda pada setiap segmen, yang kemudian dijadikan dasar dalam menyusun strategi pemasaran yang spesifik untuk tiap klaster. Selain itu, penelitian oleh (Putri & Purnomo, 2025) juga membuktikan bahwa penerapan *clustering* dan *association rules mining* pada UMKM kuliner berhasil meningkatkan efektivitas strategi promosi dan penjualan sebesar 15% setelah implementasi hasil analisis. Oleh karena itu, penerapan metode ini diharapkan dapat memberikan solusi konkret bagi Grobbak Korea dalam menyusun strategi promosi berbasis data yang lebih terarah dan berdampak.

1.2 Rumusan Masalah

Berikut merupakan rumusan masalah pada penelitian ini:

1. Bagaimana karakteristik konsumen Grobbak Korea Pandeyan dapat dianalisis menggunakan metode *clustering*?
2. Bagaimana mengetahui pola pembelian produk melalui metode *association rules mining*?
3. Strategi apa yang diperoleh dari penerapan metode *clustering* dan *association rules mining* untuk peningkatan penjualan?

1.3 Tujuan Penelitian

Berikut merupakan tujuan penelitian menjawab rumusan masalah :

1. Menganalisis karakteristik konsumen Grobbak Korea Pandeyan menggunakan metode clustering untuk mengelompokkan segmen konsumen berdasarkan demografi dan perilaku pembelian.
2. Mengidentifikasi pola pembelian konsumen Grobbak Korea Pandeyan melalui metode *association rules mining* guna menemukan pola pembelian produk.
3. Memberikan strategi peningkatan penjualan berdasarkan hasil analisis *clustering* dan *association rules mining*.

1.4 Manfaat Penelitian

Adapun kontribusi yang diharapkan dari penelitian ini dapat dirinci sebagai berikut:

1. Bagi Perusahaan

Penelitian ini membantu perusahaan memahami perilaku pelanggan melalui segmentasi berbasis data transaksi dan pola pembelian. Hasil analisis *clustering* dan *association rules mining* dapat digunakan untuk merancang strategi promosi yang lebih tepat sasaran, sehingga mendukung peningkatan penjualan dan kepuasan pelanggan.

2. Bagi Peneliti

Penelitian ini memberikan kontribusi dalam pengembangan penerapan metode *clustering* dan *association rules mining* pada analisis data transaksi, serta menjadi referensi untuk studi lanjutan di bidang data mining dan pemasaran berbasis data.

1.5 Batasan Penelitian

Pada penelitian ini diperlukan adanya batasan untuk memperjelas ruang lingkup yang akan dibahas. Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan terbatas pada transaksi penjualan yang tercatat dalam sistem Point of Sale (POS) Grobbak Korea Pandeyan pada periode Juni 2024 hingga April 2025.
2. Hasil penelitian ini hanya berlaku untuk Grobbak Korea Pandeyan pada periode dan kondisi pada periode Juni 2024 hingga April 2025, sehingga tidak dapat digeneralisasi untuk semua bisnis FnB, periode waktu berbeda, atau perusahaan dengan karakteristik berbeda.

1.6 Sistematika Penulisan

Berikut merupakan sistematika penulisan yang digunakan dalam membuat laporan penelitian tugas akhir ini,

BAB I: PENDAHULUAN

Menjelaskan latar belakang permasalahan yang dihadapi Grobbak Korea, yaitu kurangnya strategi pemasaran berbasis data. Disampaikan juga rumusan masalah, tujuan penelitian, manfaat, batasan ruang lingkup studi, dan susunan isi laporan.

BAB II TINJAUAN PUSTAKA

Memuat teori-teori yang mendasari penelitian. Juga membahas penelitian-penelitian sebelumnya (*state of the art*) sebagai pembanding dan landasan ilmiah.

BAB III METODOLOGI PENELITIAN

Menjabarkan bagaimana penelitian dilakukan, mulai dari objek penelitian, jenis dan sumber data, tahapan pengumpulan data, alur penelitian.

BAB IV PENGOLAHAN DATA

Bab ini menyajikan hasil proses pengolahan data, mulai dari tahap pembersihan (*data cleaning*), pemilihan variabel, serta hasil segmentasi pelanggan dan asosiasi produk. Visualisasi data seperti tabel, grafik, dan *cluster plot* juga disajikan untuk memperkuat pemahaman terhadap hasil analisis.

BAB V PEMBAHASAN

Merupakan bagian yang membahas secara mendalam hasil dari pengolahan data. Analisis difokuskan pada interpretasi tiap segmen pelanggan dan hubungan antar produk yang sering dibeli bersamaan. Bab ini juga menghubungkan temuan dengan teori dan literatur pada Bab II, serta memberikan implikasi strategis bagi Grobbak Korea.

BAB VI PENUTUP

Bab terakhir yang memuat kesimpulan dari keseluruhan penelitian dan menjawab tujuan penelitian yang telah ditetapkan serta saran untuk manajemen Grobbak Korea dan masukan bagi penelitian lanjutan.

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Literatur

Berikut merupakan kajian literatur yang digunakan pada penelitian ini:

2.1.1 Clustering

Penelitian Ong et al. (2023) yang berjudul “*Consumer preference analysis on the attributes of samgyeopsal Korean cuisine and its market segmentation: Integrating conjoint analysis and K-means clustering.*” menggabungkan metode *Conjoint Analysis* dan *K-Means Clustering* untuk menganalisis preferensi konsumen terhadap atribut hidangan Samgyeopsal di Filipina, seperti jenis daging, keju, harga, merek, gaya penyajian, dan minuman. Hasilnya menunjukkan bahwa atribut utama yang paling memengaruhi preferensi adalah jenis daging dan keberadaan keju, sementara merek memiliki pengaruh paling rendah. Melalui pendekatan *K-Means*, penelitian ini berhasil mengelompokkan konsumen ke dalam tiga segmen: bernilai tinggi, inti, dan bernilai rendah berdasarkan usia dan pendapatan. Temuan ini memberikan kontribusi signifikan dalam literatur pemasaran kuliner dengan menunjukkan pentingnya analisis berbasis data untuk strategi segmentasi pasar yang lebih efektif.

Penelitian Maddala & Sreedevi (2023) yang berjudul “*CNAIS: Performance Analysis of the Clustering of Non-Associated Items Set Techniques*” mengusulkan algoritma *Clustering of Non-Associated Item Sets* (CNAIS) sebagai pendekatan baru untuk mengoptimalkan analisis data transaksi dengan cara mengelompokkan tidak hanya item yang memiliki asosiasi kuat, tetapi juga item-item non-asosiatif yang sering diabaikan oleh metode konvensional seperti Apriori. CNAIS terdiri dari tiga fase utama: ekstraksi item non-asosiatif dari dataset transaksi, pembentukan klaster berdasarkan nilai ambang (*threshold value*) yang diperoleh dari atribut transaksi, dan prioritas klaster berdasarkan kontribusi item terhadap total nilai transaksi. Melalui eksperimen yang dilakukan menggunakan data transaksi restoran, CNAIS terbukti dapat meningkatkan efisiensi pemrosesan dan memberikan hasil pengelompokan yang lebih komprehensif dibandingkan pendekatan gabungan Apriori dan *clustering* konvensional. Pendekatan ini dinilai mampu menangkap nilai informasi dari data yang sebelumnya dianggap tidak relevan dalam proses analitik.

Penelitian Bozanta et al. (2022) yang berjudul “*Time Series Clustering for Grouping Products Based on Price and Sales Patterns.*” mengusulkan pendekatan *time series clustering* inovatif untuk mengelompokkan produk grosir daging berdasarkan pola historis harga dan

volume penjualan. Studi ini menyoroti keterbatasan dari metrik jarak konvensional seperti *Euclidean distance* yang hanya mengukur kedekatan numerik tanpa mempertimbangkan arah pergerakan data. Sebagai solusinya, peneliti merancang *movement pattern-based distance* (MPBD), yaitu metrik jarak yang menilai kesamaan pola perubahan antara *time series*, serta menggunakan metode *image clustering* berbasis *CNN* (*Convolutional Neural Network*) untuk mengidentifikasi pola visual dalam data harga dan penjualan. Melalui evaluasi menggunakan indeks validitas internal seperti *Calinski-Harabasz*, *Davies-Bouldin*, dan metrik kustom MPBI, hasilnya menunjukkan bahwa pendekatan hierarki dengan MPBD dan *image clustering* lebih unggul dalam mengungkap pola pergerakan harga/penjualan dibanding metode tradisional seperti *k-means* dan *hierarchical clustering* berbasis jarak numerik.

Penelitian John et al. (2023) yang berjudul “*An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market.*” mengkaji berbagai algoritma *clustering* untuk segmentasi pelanggan pada pasar ritel daring di Inggris dengan menggunakan dataset *online retail* dari UCI Machine Learning Repository. Dalam penelitian ini, pendekatan *unsupervised learning* diterapkan dengan memanfaatkan model *RFM* (*Recency, Frequency, Monetary*) sebagai dasar kuantifikasi nilai pelanggan. Lima algoritma *clustering* dibandingkan, yaitu *K-Means*, *Gaussian Mixture Model (GMM)*, *DBSCAN*, *Agglomerative Clustering*, dan *BIRCH*. GMM yang didukung oleh reduksi dimensi *PCA* menghasilkan performa terbaik dengan nilai *Silhouette Score* sebesar 0.80, mengungguli metode lain yang umumnya berada di kisaran 0.62–0.64. Penelitian ini menekankan bahwa penggunaan kombinasi *PCA* dan *GMM* dapat menghasilkan segmentasi pelanggan yang lebih akurat dan interpretable, sekaligus menunjukkan keunggulan GMM dalam menangkap pola kompleks pada data perilaku konsumen yang besar.

Penelitian Paskasius & Magdalena (2025) yang berjudul “*Penerapan Metode Clustering Pada Penjualan Makanan Dan Minuman Dalam Menentukan Strategi Marketing*” mengimplementasikan algoritma *K-Means clustering* untuk menganalisis data penjualan makanan dan minuman di restoran “Ben’s Haus Bistro” dalam rangka merancang strategi pemasaran yang lebih efektif. Dengan memanfaatkan data historis penjualan selama enam bulan, penelitian ini mengelompokkan 111 item ke dalam empat *cluster* berdasarkan hasil metode *elbow* sebagai penentu nilai *K* optimal. Proses *clustering* menunjukkan distribusi yang jelas antara item yang sangat laris, cukup laris, kurang laris, dan tidak laris, dengan validasi menggunakan *Silhouette Score* sebesar 0,6595 yang menunjukkan kualitas pengelompokan

cukup baik. Hasil penelitian digunakan untuk membentuk strategi pemasaran dengan mengkombinasikan menu dari *cluster* penjualan tertinggi.

2.1.2 Association Rules Mining

Penelitian Ghous et al. (2023) yang berjudul ” *Deep Learning based Market Basket Analysis using Association Rules.*” mengusulkan kerangka kerja baru dalam *Market Basket Analysis* (MBA) dengan memadukan *Association Rules* (AR) sebagai pemilihan fitur dan *Deep Learning* (DL) untuk klasifikasi serta prediksi kategori produk selanjutnya. Penelitian ini menyoroti keterbatasan metode AR yang hanya mampu mengidentifikasi asosiasi antar item, namun tidak mampu memprediksi perilaku pembelian selanjutnya. Untuk mengatasi hal ini, penelitian dilakukan dengan menerapkan algoritma Eclat dan FP-Growth guna mengekstraksi aturan asosiasi dari dua dataset nyata: *InstaCart* dan *Bites Bakers*. Hasil aturan ini kemudian diklasifikasikan menggunakan model DL seperti CNN, Bi-LSTM, dan CNN-BiLSTM. Hasil eksperimen menunjukkan bahwa model Bi-LSTM menghasilkan akurasi tertinggi sebesar 92% pada dataset *InstaCart*, sementara CNN-BiLSTM mencapai akurasi 77% pada data nyata *Bites Bakers*. Temuan ini menegaskan pentingnya integrasi AR sebagai *feature engineering* dalam arsitektur DL untuk prediksi pola pembelian yang lebih akurat di lingkungan ritel berbasis data besar.

Penelitian Teh et al. (2023) yang berjudul “*Unveiling the Wave: Unraveling Sales Trends and Unlocking Market Basket Insights.*” menggabungkan analisis tren penjualan dan *market basket analysis* (MBA) untuk merekomendasikan produk makanan dan minuman (*Food and Beverages*) dengan pendekatan berbasis data. Studi ini memanfaatkan data transaksi dari aplikasi *TapFood*, yang menghubungkan pedagang kaki lima di Malaysia dengan konsumen. Untuk mengatasi ketidakkonsistenan dalam penamaan produk F&B, digunakan metode *fuzzy matching* berbasis *Levenshtein Distance* dan dilakukan *clustering* terhadap nama makanan yang memiliki kemiripan $\geq 80\%$. Proses analisis meliputi *preprocessing*, transformasi biner, dan pemfilteran transaksi tunggal. Selanjutnya, algoritma *FP-Growth* digunakan untuk mengekstraksi aturan asosiasi, yang kemudian diimplementasikan dalam sistem rekomendasi berbasis *Flask web application*. Hasil analisis mampu mengidentifikasi kombinasi produk unggulan dan memberikan rekomendasi *up-selling* serta *cross-selling* yang selaras dengan pola pembelian aktual. Studi ini menegaskan potensi integrasi analisis tren penjualan dan MBA untuk meningkatkan efisiensi strategi pemasaran, akurasi rekomendasi, serta kepuasan pelanggan di sektor F&B berbasis digital.

Penelitian Martinez et al. (2021) yang berjudul “*Market basket analysis with association Rules in the retail sector using Orange. Case study: Appliances sales company*” menyelidiki penerapan *market basket analysis* (MBA) menggunakan algoritma *FP-Growth* melalui *Orange Canvas* untuk mengungkap pola pembelian konsumen di sektor ritel barang rumah tangga, seperti elektronik, furnitur, dan perlengkapan olahraga. Studi ini menekankan pentingnya pemanfaatan data transaksi historis sebagai landasan untuk mendukung pengambilan keputusan strategis perusahaan. Data dari perusahaan ritel nasional dengan pengalaman lebih dari 67 tahun dikumpulkan dari periode 2008 hingga 2017 dan diproses ke dalam format keranjang belanja (*basket format*). Setelah dilakukan *preprocessing*, data dianalisis dengan *Orange* untuk menghasilkan aturan asosiasi. Hasil eksperimen menunjukkan asosiasi kuat antar produk seperti *headboards*, *mattresses*, dan *bed bases*, serta tren musiman yang relevan seperti penjualan *split conditioners* dan *blender* pada bulan Desember. Studi ini juga mengidentifikasi bahwa *FP-Growth* lebih efisien dibandingkan *Apriori* dalam hal waktu komputasi dan jumlah aturan yang dihasilkan, serta menyarankan integrasi *business intelligence* dalam operasional perusahaan sebagai arah keberlanjutan. Penelitian ini memberikan kontribusi nyata dalam optimalisasi tata letak produk, penyusunan promosi berbasis asosiasi produk, dan perencanaan inventaris yang lebih cerdas di sektor ritel.

Penelitian Chandra et al. (2025) yang berjudul “*MARKET BASKET ANALYSIS FOR RETAIL OPTIMIZATION: “ IDENTIFYING PRODUCT ASSOCIATIONS TO ENHANCE SALES STRATEGIES*” mengeksplorasi penerapan *market basket analysis* (MBA) sebagai pendekatan strategis untuk mengoptimalkan penjualan di sektor ritel. Studi ini menggunakan algoritma *Apriori* untuk mengekstraksi *frequent itemsets* dan aturan asosiasi dari data transaksi nyata, dengan tujuan mengidentifikasi produk-produk yang sering dibeli bersamaan. Temuan penelitian menunjukkan asosiasi kuat antara beberapa produk seperti “*Milk & Cereal*” dan “*Bread & Butter*”, yang memiliki *lift value* tinggi dan dapat dimanfaatkan untuk menyusun strategi promosi berbasis *cross-selling* dan *bundling*. Penelitian ini juga menegaskan peran penting MBA dalam perancangan tata letak toko yang lebih optimal, kampanye pemasaran yang dipersonalisasi, serta pengelolaan stok yang efisien. Selain itu, Chandra et al. juga menyarankan integrasi teknik *predictive analytics* dan *machine learning* di masa depan guna meningkatkan akurasi prediksi serta efektivitas keputusan bisnis. Dengan demikian, studi ini memberikan kontribusi praktis dalam mendorong pengambilan keputusan berbasis data di dunia ritel modern.

Penelitian Audiana & Tahyudin (2025) yang berjudul “*Penerapan Analisis Keranjang Belanja Untuk Analisis Transaksi Penjualan Menggunakan Algoritma Association FP-Growth.*” menerapkan algoritma *FP-Growth* dalam analisis keranjang belanja (*market basket analysis*) untuk mengidentifikasi pola pembelian konsumen di KS Swalayan. Data transaksi yang digunakan meliputi 1.569 baris dan 15 kolom dari periode Oktober 2024, yang dianalisis menggunakan tahapan *Knowledge Discovery in Databases* (KDD), mencakup seleksi, pra-pemrosesan, transformasi, eksplorasi pola, hingga evaluasi. Hasil analisis menunjukkan aturan asosiasi yang kuat, seperti kombinasi “Snack & Roti” dengan “Susu” yang memiliki nilai *lift* sebesar 1,415, mengindikasikan hubungan positif signifikan. Selain itu, penelitian ini menambahkan visualisasi *heatmap* sebagai pendekatan baru untuk membantu interpretasi hubungan antar produk. Penelitian ini tidak hanya memberikan wawasan mengenai strategi *bundling* produk dan penataan rak, namun juga mendorong optimalisasi manajemen stok dan promosi. Dengan demikian, algoritma *FP-Growth* terbukti efektif dalam mendukung strategi pemasaran berbasis data di sektor ritel, khususnya swalayan konvensional.

2.1.3 Kombinasi Clustering dan Association Rules

Penelitian oleh Kajornkasirat et al. (2025) yang berjudul “*Customer segmentation using association rules mining on retail transaction data.*” menggabungkan metode *clustering* dan *association rules mining* untuk segmentasi pelanggan berbasis analisis RFM (*Recency, Frequency, Monetary*) pada data transaksi ritel sebanyak 202.469 entri dari Tuenjai Company di Thailand (2021–2023). Tiga algoritma *clustering* dibandingkan, yaitu K-Means, Fuzzy C-Means, dan *Self-Organizing Neural Network* (SONN), menggunakan lima matrik evaluasi: *Silhouette Coefficient*, *Calinski-Harabasz index*, *Davies-Bouldin index*, jumlah iterasi, dan waktu eksekusi. Hasilnya menunjukkan bahwa K-Means adalah algoritma paling optimal dengan SC sebesar 0.4364, CH index 184977.05, DB index 0.7913, iterasi hanya 3 kali, dan waktu eksekusi 0.6076 detik. Segmentasi menghasilkan tiga klaster pelanggan: *important value* (23%), *general development* (50%), dan *lost* (27%). Dengan menggunakan algoritma *FP-Growth* pada 1.000 transaksi tertinggi dari tiap klaster, ditemukan 38 aturan asosiasi pada klaster *important value*, dan masing-masing 2 aturan pada klaster lainnya. Salah satu aturan kuat menunjukkan bahwa pelanggan yang membeli *pressure adjusting head No. R326* selalu membeli *gas cable white/orange 50m* dengan *confidence* mencapai 100%. Penelitian ini menunjukkan bahwa integrasi K-Means dan *FP-Growth* efektif dalam menggali wawasan perilaku pelanggan untuk mendukung strategi pemasaran yang lebih terarah dan efisien.

Kanhere et al. (2022) yang berjudul “*Clustering based approach to enhance association rules mining*” mengembangkan pendekatan berbasis *clustering* untuk meningkatkan efisiensi *association rules mining* pada data transaksi ritel. Produk dikelompokkan menggunakan algoritma *k-means* berdasarkan dua parameter utama: frekuensi kemunculan dan harga rata-rata, sehingga menghasilkan pengurangan data hingga hanya 7% dari total item tanpa kehilangan aturan penting. Data yang telah dikelompokkan kemudian dianalisis menggunakan algoritma Apriori, FP-Growth, dan Eclat. Hasilnya menunjukkan bahwa aturan asosiasi yang dihasilkan dari data tereduksi sama dengan aturan dari dataset penuh, namun dengan waktu eksekusi yang jauh lebih singkat. Penelitian ini juga menerapkan *differential market basket analysis* dengan membagi transaksi berdasarkan waktu (pagi, siang, sore, malam), dan menemukan bahwa kelompok waktu pagi menghasilkan aturan unik yang tidak muncul di kelompok lainnya.

Penelitian yang dilakukan oleh Putri & Purnomo (2025) yang berjudul “*Integrasi algoritma apriori dan k-means dalam analisis pola pembelian untuk meningkatkan strategi pemasaran*” memadukan algoritma *K-Means Clustering* dan *Apriori* untuk mengidentifikasi pola pembelian konsumen sebagai dasar penyusunan strategi pemasaran pada bisnis kuliner UMKM Premium Salad.co. Dengan memanfaatkan 1.386 data transaksi dari sistem Go-Biz, peneliti terlebih dahulu membentuk empat cluster menggunakan metode *K-Means*, dengan validitas terbaik pada nilai Davies Bouldin Index sebesar 0,465. Kemudian, algoritma *Apriori* diterapkan pada tiga dari empat cluster yang teridentifikasi memiliki masalah dalam frekuensi pembelian rendah. Hasilnya, pada cluster 0, 1, dan 3 dihasilkan 1, 3, dan 16 aturan asosiasi dengan tingkat kepercayaan hingga 100% dan *lift ratio* lebih dari 1—menandakan kekuatan korelasi antar produk cukup tinggi untuk digunakan dalam strategi pemasaran berbasis paket bundling atau rekomendasi produk. Penelitian ini menegaskan bahwa integrasi antara clustering dan association Rules dapat meningkatkan relevansi aturan yang dihasilkan, khususnya dalam menyusun strategi pemasaran yang berbasis data historis pelanggan secara lebih akurat dan tersegmentasi.

Penelitian yang dilakukan oleh Mandala & Putri (2023) yang berjudul “*Data mining technique for grouping products using clustering based on association.*” mengusulkan pendekatan hybrid data mining dengan menggabungkan metode *association Rules* menggunakan algoritma FP-Growth dan metode *clustering* menggunakan algoritma K-Means untuk mengelompokkan produk di minimarket. Studi ini dimotivasi oleh permasalahan

menumpuknya produk yang tidak terjual hingga kedaluwarsa akibat penurunan loyalitas konsumen dan peningkatan tren belanja daring. Pada tahap pertama, peneliti menerapkan FP-Growth untuk menemukan pola frekuensi pembelian produk berdasarkan transaksi penjualan, menghasilkan 9 item dengan *support* $\geq 33,33\%$ dan 8 aturan asosiasi yang memenuhi ambang batas *confidence*. Selanjutnya, hasil tersebut digunakan sebagai input untuk proses K-Means *clustering* yang membagi 6 produk ke dalam dua kelompok: *most interested product* dan *interested product*, dengan validitas klaster diukur menggunakan *Davies-Bouldin Index* (DBI) sebesar 0,114, yang menunjukkan konsistensi klaster sebesar 88,6%. Pendekatan ini terbukti mampu menyederhanakan pengolahan data transaksi penjualan dan membantu manajer ritel dalam perencanaan stok serta strategi pemasaran yang lebih efektif.

Penelitian yang dilakukan oleh Kalijaga (2021) yang berjudul “*Penerapan Analytical Hierarchy Proccess, Clustering, dan AR-MBA sebagai Solusi Strategi Pemasaran Pada Toko Swalayan X*” berfokus pada perancangan strategi pemasaran berbasis data dengan menerapkan tiga metode utama, yaitu *Analytical Hierarchy Process* (AHP), *Clustering*, dan *Association Rules Market Basket Analysis* (AR-MBA) pada objek penelitian berupa Swalayan X di Sleman, Yogyakarta. AHP digunakan untuk menentukan prioritas variabel-variabel yang berpengaruh terhadap strategi pemasaran, seperti umur, jarak, intensitas belanja, dan penghasilan pelanggan. Selanjutnya, metode *clustering* diaplikasikan untuk mengelompokkan pelanggan berdasarkan karakteristik tersebut, menghasilkan tiga klaster dengan profil pemasaran yang berbeda, di mana *Cluster 2* diidentifikasi sebagai kelompok yang paling potensial untuk ditargetkan. Terakhir, metode AR-MBA digunakan untuk mengidentifikasi pola pembelian konsumen berdasarkan asosiasi antar produk, dan ditemukan empat pola utama kombinasi pembelian antar departemen seperti Dept 2 (snack) dengan Dept 6 (pembersih), serta Dept 9 (wafer & roti) dengan Dept 4 (minuman). Hasil dari penelitian ini memberikan rekomendasi strategi promosi berupa *mix bundling* antar produk-produk yang sering dibeli bersamaan, sehingga dapat meningkatkan efektivitas pemasaran dan penjualan secara keseluruhan.

Tabel 2. 1 Kajian Literatur penelitian

No.	Penulis	Objek Penelitian	Variabel	Algoritma				
				Clustering			Association Rules Mining	
				K-Means	Hierarchical	DBSCAN	FP-Growth	Apriori
1.	(Ong et al., 2023)	Konsumen Samgyeopsal	<i>Entree</i> , keju, harga, minuman, gaya memasak, merek, demografi	✓				
2.	(Maddala & Sreedevi, 2023)	Transaksi restoran	Item non-asosiatif, <i>threshold</i> , prioritas	✓				
3.	(Bozanta et al., 2022)	Produk grosir online	Harga dan volume penjualan historis	✓				
4.	(John et al., 2023)	Pelanggan ritel daring UK	RFM (<i>Recency</i> , <i>Frequency</i> , <i>Monetary</i>)	✓	✓	✓		

No.	Penulis	Objek Penelitian	Variabel	Algoritma				
				Clustering			Association Rules Mining	
				K-Means	Hierarchical	DBSCAN	FP-Growth	Apriori
5.	(Paskasius & Magdalena, 2025)	Restoran Ben's Haus Bistro	Nama item, kategori, jumlah terjual	✓				
6.	(Ghous et al., 2023)	Dataset InstaCart & Bites Bakers	Kombinasi produk, prediksi pembelian				✓	
7.	(Teh et al., 2023)	Pedagang kaki lima F&B Malaysia (TapFood)	Nama makanan, waktu transaksi				✓	
8.	(Martinez et al., 2021)	Perusahaan retail barang rumah tangga	Produk: elektronik, furnitur, seasonal				✓	
9.	(Chandra et al., 2025).	Retail umum	Produk sering dibeli bersama					✓

No.	Penulis	Objek Penelitian	Variabel	Algoritma				
				Clustering			Association Rules Mining	
				K-Means	Hierarchical	DBSCAN	FP-Growth	Apriori
10.	(Audiana & Tahyudin, 2025)	KS Swalayan	Snack, roti, susu, waktu transaksi				✓	
11.	(Kajornkasirat et al., 2025)	Data transaksi ritel Tuenjai Company	Waktu, Jumlah Transaksi, dan total pembelian	✓			✓	
12	(Kanhare et al., 2021)	Transaksi ritel	Frekuensi, harga rata-rata	✓			✓	✓
13.	(Putri & Purnomo, 2025)	Transaksi Go-Biz Premium Salad.co (UMKM)	Jumlah Transaksi, dan perilaku pembelian	✓				✓
14.	(Mandala & Putri, 2023)	Penjualan minimarket	Item produk, frekuensi pembelian	✓			✓	

No.	Penulis	Objek Penelitian	Variabel	Algoritma				
				Clustering			Association Rules Mining	
				K-Means	Hierarchical	DBSCAN	FP-Growth	Apriori
15.	(Kalijaga & Putra, 2021)	Swalayan X (Retail)	Umur, Jarak, Intensitas belanja, penghasilan	✓			✓	
16.	Penelitian	Grobbak Korea	Jumlah Produk, Total Pembelian, Waktu, Platform	✓			✓	

2.1.4 State of art

Sebagian besar penelitian berfokus pada penerapan metode *clustering* untuk segmentasi pelanggan dan *association rules mining* untuk analisis pola pembelian. Misalnya, penelitian yang dilakukan oleh Ong et al. (2023) dan Paskasius & Magdalena (2025) menggunakan algoritma K-Means untuk mengelompokkan konsumen berdasarkan atribut seperti harga, jenis menu, dan demografi. Sementara itu, penelitian dari Maddala & Sreedevi (2023) dan John et al. (2023) mengombinasikan K-Means, Hierarchical, dan DBSCAN guna menghasilkan segmentasi pelanggan yang lebih presisi. Hasil-hasil tersebut menegaskan bahwa clustering efektif dalam membedakan kelompok pelanggan atau produk dengan karakteristik serupa sehingga dapat digunakan sebagai dasar perancangan strategi pemasaran yang lebih terarah.

Selanjutnya, penelitian lain memperdalam analisis pola pembelian melalui metode *association rules mining*, penelitian yang dilakukan oleh Ghous et al. (2023), Teh et al. (2023) dan Martinez et al. (2021) menggunakan algoritma FP-Growth untuk menemukan hubungan antar produk yang sering dibeli bersamaan (*frequent itemset*). Adapun Chandra et al. (2025) menerapkan algoritma Apriori pada data ritel umum untuk merumuskan strategi *cross-selling* dan *bundling*. Kajornkasirat et al. (2025) serta Kanhere et al. (2021) mengintegrasikan *clustering* dengan *association rules*, di mana K-Means digunakan untuk mengelompokkan pelanggan atau produk terlebih dahulu, kemudian pola asosiasi ditemukan menggunakan FP-Growth atau Apriori. Integrasi ini terbukti mampu mempercepat waktu komputasi serta menghasilkan aturan asosiasi yang lebih relevan untuk tiap segmen.

Penelitian yang lebih baru mulai mengarah pada konteks usaha mikro dan kuliner modern. Putri & Purnomo (2025) serta Mandala & Putri (2023) mengombinasikan K-Means dengan Apriori dan FP-Growth untuk mengidentifikasi paket produk yang memiliki potensi penjualan tinggi, sedangkan Kalijaga & Putra (2021) menambahkan metode AHP untuk menentukan prioritas variabel pelanggan seperti umur, jarak, intensitas belanja, dan penghasilan.

Berdasarkan kajian dari penelitian-penelitian tersebut, masih terdapat celah penelitian pada aspek perilaku pelanggan yang dipengaruhi oleh waktu transaksi dan *platform* pemesanan digital. Kebaruan dalam penelitian ini terletak pada penerapan kombinasi metode *K-Means Clustering* dan *FP-Growth Association Rules* dalam konteks usaha kuliner berbasis *platform* digital (*ShopeeFood*, *GrabFood*, *GoFood*). Tidak seperti penelitian sebelumnya yang umumnya berfokus pada atribut RFM atau karakteristik produk semata, penelitian ini

menambahkan dimensi waktu transaksi dan *platform* sebagai variabel analisis perilaku pelanggan.

2.1.5 Posisi Penelitian

Berdasarkan pemaparan landasan teori diatas, Berdasarkan tinjauan tersebut, penelitian akan mengaplikasikan metode *Association Rules-Market Basket Analysis* dengan algoritma FP-Growth serta metode *clustering* menggunakan algoritma K-Means. Di sisi lain, FP-Growth unggul dalam efisiensi memori dan waktu karena tidak membangkitkan kandidat *itemset* yang banyak seperti pada Apriori, sehingga sangat *scalable* untuk analisis pola pembelian pada periode satu tahun penuh.

Beberapa algoritma clustering seperti DBSCAN tidak dipilih karena memerlukan parameter seperti ϵ dan *minPts* yang sulit diatur pada data transaksi dengan kepadatan tidak merata, sehingga banyak pelanggan penting bisa salah teridentifikasi sebagai *noise*. Lalu *Hierarchical Clustering* tidak dipilih Kompleksitas komputasi mencapai $O(n^3)$, artinya jika jumlah transaksi (n) bertambah, waktu pemrosesan dan penggunaan memori akan naik sangat cepat pada ratusan hingga ribuan data transaksi, proses ini bisa menjadi sangat lambat dan membebani komputer. Lalu, *Gaussian Mixture Model* (GMM) tidak cocok karena mengasumsikan setiap kluster mengikuti distribusi *Gaussian* (normal). Padahal, pola transaksi di F&B sering kali tidak terdistribusi secara normal—ada lonjakan pembelian pada produk tertentu atau periode promosi yang bersifat *skewed*. Jika asumsi ini tidak terpenuhi, hasil kluster bisa bias atau tidak mewakili realitas. Lalu CNAIS tidak dipilih karena memerlukan pemisahan *itemset* non-asosiatif yang rumit dan tidak tersedia *out-of-the-box* di *platform* analisis standar, sehingga mempersulit implementasi pada dataset Grobbak Korea Pandeyan. Dan terakhir BIRCH tidak dipilih karena ia memerlukan pengaturan *threshold* dan *branching factor* yang rumit dan hasil klasternya sangat tergantung pada urutan data yang mana pada data di POS disusun berdasarkan *timestamp* transaksi dari yang terbaru ke terlama (*descending*), bukan berurutan kronologis dari tanggal paling awal ke akhir, sehingga kurang stabil untuk analisis transaksi Grobbak Korea Pandeyan.

Lalu pada *association rules mining*, algoritma apriori tidak dipakai karena harus membuat dan memeriksa banyak kombinasi calon *itemset* secara berulang-ulang, sehingga sangat lambat dan boros memori untuk dataset transaksi yang besar. Dan *eclat* tidak dipakai karena algoritma ini menyimpan daftar ID transaksi (*TID lists*) untuk setiap *itemset* yang dapat membengkak dan memakan memori besar, serta implementasinya tidak umum tersedia di *tool* standar analisis.

2.2 Landasan Teori

2.2.1 Food and Beverage

Industri makanan dan minuman (*Food and Beverage/F&B*) merupakan salah satu sektor yang berkembang pesat di Indonesia dan memberikan kontribusi signifikan terhadap perekonomian nasional. Menurut Kementerian Perindustrian Republik Indonesia. (2024), subsektor industri makanan dan minuman menyumbang lebih dari 40,17 % terhadap Produk Domestik Bruto (PDB) industri pengolahan nonmigas pada tahun 2024. Perkembangan sektor ini ditopang oleh meningkatnya permintaan konsumen serta perubahan pola konsumsi masyarakat yang lebih dinamis.

F&B retail, seperti restoran cepat saji, kafe, dan kedai kuliner modern, kini menghadapi tantangan besar dalam memahami kebutuhan dan preferensi pelanggan. Hal ini disebabkan oleh meningkatnya variasi produk, kompetisi yang ketat, dan dinamika perilaku pelanggan yang sangat cepat berubah (Fajrul & Saptyana, 2025). Oleh karena itu, pengambilan keputusan yang tepat berbasis data sangat penting dalam mendukung strategi pemasaran, pengembangan produk, dan manajemen layanan pelanggan.

Selain itu, tren digitalisasi mendorong pelaku usaha *F&B* untuk memanfaatkan teknologi informasi, seperti *Point of Sales* (POS), untuk merekam setiap transaksi pelanggan. Data historis penjualan ini menyimpan informasi yang berharga mengenai perilaku pembelian, preferensi produk, waktu transaksi, hingga nilai pembelian. Dengan teknik analisis data seperti *clustering* dan *association rules mining*, pelaku usaha dapat menggali wawasan yang lebih dalam untuk meningkatkan kepuasan pelanggan dan efektivitas promosi (Iswavigra et al., 2023).

2.2.2 Data Mining

Data mining merupakan suatu proses untuk mengekstraksi pengetahuan yang berguna dari kumpulan data berukuran besar dengan menggunakan teknik statistik, matematika, dan pembelajaran mesin. Menurut Larose (2005) data mining adalah proses selektif dalam mengeksplorasi dan memodelkan data untuk menemukan pola-pola tersembunyi yang sebelumnya tidak diketahui, namun dapat mendukung pengambilan keputusan yang lebih baik dan terukur.

Tujuan utama dari data mining adalah menemukan informasi yang tersembunyi dalam data dan menyajikannya dalam bentuk yang dapat ditindaklanjuti oleh pengguna, terutama dalam konteks pengambilan keputusan bisnis. Dalam industri seperti ritel dan makanan-minuman

(F&B), data mining memiliki kontribusi penting dalam memahami perilaku konsumen, menentukan strategi promosi, meningkatkan efisiensi operasional, serta mengoptimalkan penawaran produk.

Secara umum, metode dalam data mining terbagi ke dalam dua kategori besar, yaitu *predictive methods* dan *descriptive methods*. Metode prediktif, seperti klasifikasi dan regresi, digunakan untuk meramalkan perilaku atau nilai dari suatu data berdasarkan pola historis. Sementara itu, metode deskriptif bertujuan untuk menggambarkan pola atau struktur data yang ada, tanpa memerlukan label target.

Proses data mining terdiri dari sejumlah tahapan yang saling berkaitan dan membentuk satu alur kerja analisis data yang utuh. Menurut Larose (2005), tahapan-tahapan dalam proses data mining secara umum meliputi:

1. Pengumpulan Data (*Data Collection*)
2. Pembersihan Data (*Data Cleaning*)
3. Seleksi Data (*Data Selection*)
4. Transformasi Data (*Data Transformation*)
5. Penerapan Metode Data Mining (*Data Mining Method Execution*)

2.2.3 Clustering

Clustering adalah metode analisis data yang termasuk dalam teknik unsupervised learning pada machine learning, yang bertujuan untuk mengelompokkan objek-objek data ke dalam kelompok (*cluster*) sedemikian rupa sehingga objek dalam satu cluster memiliki kemiripan yang tinggi, sedangkan objek antar cluster memiliki perbedaan yang signifikan (Han et al., 2011). Tujuan utama dari clustering dalam konteks pemasaran adalah untuk memahami pola perilaku pelanggan secara lebih mendalam dan membagi mereka ke dalam kelompok-kelompok homogen. Hal ini mendukung proses segmentasi pasar, penentuan strategi promosi yang lebih personal, serta pengembangan produk yang lebih tepat sasaran.

Menurut Han et al (2011) dalam clustering, terdapat beberapa pendekatan utama yang digunakan untuk mengelompokkan data. Pendekatan-pendekatan ini dikembangkan berdasarkan cara kerja algoritma dalam membentuk cluster. Berikut merupakan pendekatan utama atau algoritma dalam *clustering*:

1. *Partitioning Clustering*

Pendekatan ini membagi data ke dalam sejumlah k cluster, di mana setiap data hanya menjadi anggota satu cluster. Proses ini berusaha mengoptimalkan fungsi objektif

seperti meminimalkan jarak antar data dengan pusat cluster (*centroid*). Metode yang paling umum dalam pendekatan ini adalah *K-Means*, yang bekerja dengan cara mengiterasi pusat cluster hingga mencapai stabilitas. Pendekatan ini efisien dan cocok untuk dataset besar, namun sensitif terhadap inisialisasi dan jumlah cluster yang ditentukan di awal.

2. *Hierarchical Clustering*

Hierarchical clustering membentuk hierarki klaster dengan dua pendekatan utama: *agglomerative* (penggabungan dari bawah ke atas) dan *divisive* (pemisahan dari atas ke bawah). Hasilnya divisualisasikan dalam bentuk dendrogram yang menunjukkan hubungan antar cluster secara bertingkat. Keunggulan pendekatan ini adalah tidak perlu menentukan jumlah cluster di awal, namun kurang efisien untuk dataset besar.

3. *Density-Based Clustering*

Metode ini mengidentifikasi cluster sebagai area dengan kepadatan tinggi yang dipisahkan oleh area dengan kepadatan rendah. Salah satu algoritma populernya adalah DBSCAN, yang mampu menemukan cluster dengan bentuk *arbitrer* serta mengatasi *noise* dan *outlier*. Clustering berbasis kepadatan sangat sesuai untuk dataset dengan distribusi tidak teratur, namun sangat tergantung pada parameter seperti *epsilon* dan *minPts*.

4. *Model-Based Clustering*

Pendekatan ini mengasumsikan bahwa data dibangkitkan dari model statistik, seperti campuran distribusi *Gaussian*. *Clustering* dilakukan dengan memperkirakan parameter model menggunakan teknik seperti *Expectation Maximization* (EM). *Model-based clustering* menghasilkan pembagian data berdasarkan probabilitas keanggotaan, namun lebih kompleks secara komputasi.

5. *Grid-Based Clustering*

Pendekatan *grid-based* membagi ruang data ke dalam *grid sel* kecil, lalu mengelompokkan sel-sel berdasarkan densitas data di dalamnya. Salah satu contohnya adalah metode STING. Kelebihannya adalah efisiensi tinggi dan solvabilitas terhadap dataset besar, tetapi kurang cocok untuk data berdimensi tinggi atau dengan distribusi yang tidak merata.

2.2.4 Association Rules Mining

Association Rules Mining merupakan salah satu teknik dalam *data mining* yang bertujuan untuk menemukan hubungan atau pola asosiasi antar item dalam suatu kumpulan data transaksi. Teknik ini populer digunakan dalam konteks *Market Basket Analysis* (MBA), yaitu analisis yang bertujuan mengidentifikasi kombinasi item yang sering dibeli bersama dalam satu transaksi konsumen. Menurut Han et al. (2011), *association Rules* adalah teknik untuk menemukan pola hubungan yang menarik dari kumpulan data dalam bentuk aturan “jika-maka” (*if-then*). Aturan ini dapat digunakan untuk membantu pengambilan keputusan dalam strategi pemasaran seperti *cross-selling*, *product bundling*, dan tata letak toko.

Dalam buku yang ditulis oleh Kantardzic (2020) dengan judul “*Data Mining: Concepts, Models, Methods, and Algorithms*”. Suatu aturan asosiasi umumnya dinyatakan dalam bentuk implikasi logis sebagai $A \rightarrow B$, yang artinya jika item A muncul dalam suatu transaksi, maka terdapat kemungkinan besar bahwa item B juga akan muncul dalam transaksi yang sama. Dalam konteks analisis data transaksi penjualan, bentuk aturan ini digunakan untuk memetakan pola pembelian konsumen yang bersifat berulang dan konsisten. Dengan mengidentifikasi pola tersebut, pelaku bisnis dapat merancang strategi pemasaran yang lebih efektif, seperti penawaran bundling atau rekomendasi produk yang relevan.

Untuk mengevaluasi seberapa kuat dan bermanfaat sebuah aturan asosiasi, digunakan tiga metrik utama, yaitu *support*, *confidence*, dan *lift*.

1. *Support*: Mengukur proporsi transaksi yang mengandung kombinasi item A dan B terhadap total keseluruhan transaksi.

$$Support(A \Rightarrow B) = \frac{\text{Jumlah transaksi yang mengandung A dan B}}{\text{Jumlah total transaksi}} \quad (2.1)$$

2. *Confidence*: Mengukur probabilitas munculnya item B dalam transaksi yang mengandung item A.

$$Confidence(A \Rightarrow B) = \frac{\text{Jumlah transaksi dengan A dan B}}{\text{Jumlah total transaksi A}} \quad (2.2)$$

3. *Lift*: Menilai kekuatan hubungan antara item A dan B dibandingkan dengan jika keduanya terjadi secara independen.

$$Lift(A \Rightarrow B) = \frac{Confidence(A \Rightarrow B)}{Support(B)} \quad (2.3)$$

Ketiga metrik ini berperan penting dalam menilai frekuensi kemunculan dan kekuatan keterkaitan antar item dalam aturan tersebut. *Support* menunjukkan seberapa sering kombinasi item A dan B muncul bersama dalam seluruh transaksi. *Confidence* menggambarkan tingkat kepercayaan bahwa B akan muncul dalam transaksi yang sudah mengandung A. Sementara itu, *lift* mengukur kekuatan asosiasi antara A dan B dengan membandingkannya terhadap kemunculan B secara acak. Semakin tinggi nilai ketiga metrik ini, semakin relevan dan signifikan aturan asosiasi yang terbentuk.

Menurut (Han et al., 2011), proses *clustering* dalam data mining terdiri dari beberapa tahapan sistematis yang bertujuan untuk membentuk kelompok data (*cluster*) berdasarkan tingkat kesamaan antar objek. Langkah-langkah tersebut dijelaskan sebagai berikut:

1. Pengumpulan dan Seleksi Data

Langkah awal dalam proses clustering adalah pengumpulan data yang relevan untuk dianalisis. Data tersebut kemudian dipilih berdasarkan variabel-variabel penting yang dapat merepresentasikan objek yang akan dikelompokkan. Misalnya, dalam segmentasi pelanggan, variabel seperti total pembelian, frekuensi transaksi, dan jenis produk yang dibeli menjadi fokus utama.

2. Pembuatan Variabel Turunan (*Feature Construction*)

Tahap ini melibatkan rekayasa fitur (*feature engineering*) untuk membentuk atribut baru yang dapat memberikan informasi lebih bermakna dari data mentah. Contohnya termasuk menghitung rata-rata pembelanjaan per kunjungan atau frekuensi kunjungan dalam periode tertentu. Fitur-fitur ini menjadi input utama dalam proses pengelompokan.

3. *Preprocessing Data*: Pembersihan dan Normalisasi

Data yang digunakan harus melalui tahap *preprocessing* untuk memastikan kualitasnya. Hal ini mencakup penghapusan nilai kosong, data duplikat, serta normalisasi agar seluruh atribut berada dalam skala yang seragam. Proses normalisasi sangat penting dalam algoritma berbasis jarak seperti K-Means, untuk mencegah dominasi satu variabel atas variabel lain.

4. Penentuan Jumlah Cluster (k)

Sebelum K-Means dijalankan, jumlah cluster (k) harus ditentukan terlebih dahulu. Salah satu metode yang umum digunakan adalah *Elbow Method*, yaitu dengan menjalankan

K-Means untuk berbagai nilai k dan memplot nilai *Within-Cluster Sum of Squares (WCSS)*. Jumlah cluster yang optimal dipilih berdasarkan titik di mana penurunan WCSS mulai melambat secara signifikan.

5. Pelaksanaan Algoritma *Clustering*

Setelah jumlah cluster ditentukan, algoritma clustering dijalankan. Dalam metode K-Means, proses dimulai dengan inisialisasi *centroid*, lalu setiap data dihitung jaraknya ke masing-masing centroid, dan dikelompokkan ke dalam cluster terdekat. Proses ini diulang dengan memperbarui posisi centroid hingga tercapai konvergensi.

6. Evaluasi dan Validasi Hasil Cluster

Hasil clustering perlu dievaluasi untuk memastikan kualitas dan konsistensi pengelompokan. Salah satu ukuran yang umum digunakan adalah *Silhouette Coefficient*, yang mengukur tingkat kesamaan suatu data dengan cluster-nya dibandingkan dengan cluster lain. Evaluasi ini membantu menilai apakah jumlah cluster sudah sesuai.

7. Interpretasi dan Aplikasi Hasil Clustering

Setelah hasil cluster dievaluasi, masing-masing kelompok dianalisis untuk mengidentifikasi karakteristik khasnya. Dalam konteks bisnis, hasil ini dapat digunakan untuk strategi pemasaran seperti personalisasi promosi, pengembangan produk, atau program loyalitas. Clustering memungkinkan pengambilan keputusan yang lebih strategis berbasis perilaku pelanggan.

2.2.5 Market Basket Analysis

Market Basket Analysis (MBA) merupakan salah satu teknik dalam data mining yang digunakan untuk menemukan pola keterkaitan antar produk berdasarkan transaksi pembelian konsumen. Teknik ini memungkinkan bisnis untuk memahami produk apa saja yang sering dibeli bersamaan, yang kemudian dapat digunakan untuk meningkatkan penjualan melalui strategi seperti *cross-selling*, *upselling*, *product bundling*, hingga penataan rak dan *layout* toko yang lebih efisien (M. Alawadh & M. Barnawi, 2022).

MBA bekerja dengan mengidentifikasi *association Rules*, yaitu aturan berbentuk “jika-maka” (*if-then*) yang menghubungkan satu atau lebih item dengan item lainnya dalam satu keranjang belanja Martinez et al. (2021). Salah satu algoritma paling umum dalam MBA adalah Apriori, yang menghitung frekuensi kombinasi item untuk menentukan *support*, *confidence*, dan *lift* dari aturan tersebut.

2.2.6 Algoritma FP-Growth

Frequent Pattern Growth adalah salah satu algoritma dalam *market basket analysis* yang digunakan untuk menemukan pola asosiasi antar item dalam transaksi penjualan. *FP-Growth* dikembangkan sebagai alternatif dari algoritma Apriori karena memiliki efisiensi yang lebih tinggi dalam hal komputasi dan memori (Audiana & Tahyudin, 2025).

FP-Growth bekerja tanpa harus membangkitkan seluruh kandidat kombinasi item seperti pada Apriori, sehingga dapat mengurangi kompleksitas waktu pencarian pola secara signifikan terutama pada dataset besar (Martinez et al., 2021). Algoritma ini membentuk struktur data khusus yang disebut *FP-Tree (Frequent Pattern Tree)* yang merepresentasikan kumpulan itemset secara terstruktur dan kompresi.

2.2.7 Algoritma K-Means

K-Means Clustering adalah salah satu algoritma *unsupervised learning* yang digunakan untuk mengelompokkan data ke dalam sejumlah klaster berdasarkan kemiripan karakteristiknya. Algoritma ini bekerja dengan cara membagi data ke dalam k kelompok (*cluster*), di mana setiap data akan dimasukkan ke dalam cluster dengan pusat (*centroid*) terdekat John et al. (2023).

Berikut merupakan langkah-langkah untuk menggunakan algoritma *K-means* menurut buku dari Larose (2005):

1. Tentukan jumlah klaster (K)
Pilih jumlah kelompok atau klaster yang diinginkan sesuai dengan tujuan analisis.
2. Inisialisasi centroid awal secara acak
Pilih K titik centroid awal secara acak dari dataset. Titik ini mewakili pusat awal dari masing-masing klaster.
3. Hitung jarak setiap data ke *centroid*
Hitung jarak *Euclidean* antara setiap data dan seluruh centroid.
Rumus jarak Euclidean:

$$d(x, c) = \sqrt{\sum_{i=1}^n (x_i - c_i)^2} \quad (2.4)$$

Keterangan:

$d(x, c)$: Jarak Euclidean antara titik data x dan centroid c

x_i : Nilai atribut ke- i dari titik data

c_i : Nilai atribut ke- i dari centroid

n : Jumlah atribut yang digunakan dalam perhitungan

$\sum_{i=1}^n \dots$: Penjumlahan selisih kuadrat dari semua atribut

4. Mengelompokkan data ke kluster berdasarkan jarak terdekat

Setiap data ditempatkan ke dalam kluster yang memiliki *centroid* dengan nilai jarak *Euclidean* terkecil.

5. Menghitung ulang posisi *centroid*

Untuk setiap kluster, hitung ulang *centroid* baru berdasarkan rata-rata dari semua data yang ada di kluster tersebut:

$$c_j = \frac{1}{N_j} \sum_{i=1}^{N_j} x_i \quad (2.5)$$

Keterangan:

c_j : Titik pusat (*centroid*) baru untuk kluster ke- j .

N_j : Jumlah total data (anggota) dalam kluster ke- j .

x_i : Titik data ke- i yang masuk dalam kluster ke- j .

$\sum_{i=1}^{N_j} x_i$: Penjumlahan semua nilai atribut dari data ke- i dalam kluster ke- j .

6. Mengulangi langkah 3–5 hingga konvergen

Proses dihentikan jika:

- Tidak ada perubahan alokasi anggota kluster, atau
- Perubahan posisi centroid sangat kecil/tidak signifikan, atau
- Jumlah iterasi maksimum telah tercapai

2.2.8 Silhouette Score

Silhouette Score merupakan salah satu metode evaluasi yang digunakan untuk mengukur kualitas dari hasil algoritma clustering. Metode ini diperkenalkan oleh Rousseeuw (1987) dan digunakan secara luas dalam penelitian berbasis *unsupervised learning* untuk menilai seberapa baik objek telah dikelompokkan ke dalam klusternya masing-masing.

Secara matematis, nilai *silhouette* $s(i)$ untuk titik data i dihitung dengan rumus:

$$S(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}} \quad (2.6)$$

Keterangan:

$s(i)$: Nilai *silhouette*.

$a(i)$: jarak rata-rata dari i ke seluruh titik dalam klaster yang sama.

$b(i)$: Jarak rata-rata dari i ke seluruh titik dalam klaster terdekat yang bukan miliknya.

Nilai rata-rata dari semua $s(i)$ memberikan *Silhouette Score* keseluruhan yang dapat digunakan untuk membandingkan performa pengelompokan dengan jumlah klaster berbeda. Skor ini sangat bermanfaat dalam proses model *selection*, khususnya untuk menentukan jumlah klaster optimal pada algoritma seperti K-Means atau *Hierarchical Clustering* (Larose, 2005).

Dalam konteks pemasaran, khususnya segmentasi pelanggan, *Silhouette Score* digunakan untuk mengukur sejauh mana segmentasi konsumen yang dihasilkan benar-benar memiliki perbedaan karakteristik yang signifikan antar kelompok. Penelitian oleh Ong et al. (2023) menggunakan K-Means dengan evaluasi *Silhouette Score* untuk menganalisis preferensi konsumen terhadap atribut makanan Korea, yang kemudian digunakan untuk pengembangan strategi produk dan pemasaran yang lebih terarah.

Lebih lanjut, Teh et al. (2023) juga menerapkan pendekatan serupa dalam eksplorasi tren penjualan dan pemetaan pasar, di mana penggunaan *Silhouette Score* mendukung pemilihan jumlah klaster yang representatif dari perilaku pembelian pelanggan. Penilaian menggunakan metrik ini memungkinkan pengambilan keputusan berbasis data untuk strategi penjualan dan kampanye promosi yang lebih efektif.

2.2.1 Elbow Method

Elbow Method merupakan salah satu pendekatan heuristik yang digunakan untuk menentukan jumlah klaster optimal dalam algoritma *unsupervised learning*, khususnya pada metode klasterisasi seperti K-Means. Prinsip dasar dari *Elbow Method* adalah mencari titik “tekuk” (*elbow*) pada kurva yang menunjukkan hubungan antara jumlah klaster dan nilai *Within-Cluster Sum of Squares* (WCSS), yaitu total variansi dari data terhadap centroid klasternya masing-masing (Bholowalia & Kumar, 2014).

Secara matematis, *elbow method* dapat dihitung dengan WCSS (*Within-Cluster Sum of Squares*):

$$WCSS = \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - \mu_j\|^2 \quad (2.7)$$

Keterangan:

k : Jumlah klaster.

C_j : Klaster ke- j .

x_i : Titik data ke- i dalam klaster C_j

μ_j : *Centroid* (rata-rata) dari klaster C_j

$\|x_i - \mu_j\|^2$: Kuadrat jarak *Euclidean* antara titik data dan *centroid*

Dalam praktiknya, *Elbow Method* dilakukan dengan menjalankan algoritma klasterisasi untuk berbagai nilai k (jumlah klaster) dan menghitung WCSS pada masing-masing nilai tersebut. Nilai WCSS cenderung menurun seiring bertambahnya jumlah klaster, namun penurunan tersebut akan melambat setelah titik tertentu. Titik di mana terjadi perlambatan signifikan itulah yang disebut "*elbow*" dan dianggap sebagai jumlah klaster yang optimal karena menyeimbangkan antara kompleksitas model dan representasi data (Ketchen & Shook, 1996).

Pendekatan ini penting untuk mencegah *overfitting* akibat jumlah klaster yang terlalu banyak, maupun *underfitting* akibat jumlah klaster yang terlalu sedikit. Dengan kata lain, *Elbow Method* membantu dalam mengidentifikasi *sweet spot* antara akurasi klasterisasi dan efisiensi model.

2.2.2 Davies–Bouldin Index

Dalam analisis clustering, penting untuk mengevaluasi kualitas hasil pengelompokan yang telah dilakukan. Salah satu metode evaluasi yang umum digunakan untuk tujuan ini adalah Davies–Bouldin Index (DBI). Indeks ini diusulkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979 dan berfungsi sebagai metrik internal untuk menilai validitas suatu skema klasterisasi tanpa memerlukan label eksternal (*ground truth*). Menurut Larose (2005), *Davies–Bouldin Index* menghitung rasio antara jarak antar pusat klaster dan penyebaran dalam klaster (*within-cluster scatter*), dengan rumus sebagai berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{S_i + S_j}{M_{ij}} \right) \quad (2.8)$$

Keterangan:

- k : Jumlah klaster.
- S_i : Rata-rata jarak antar anggota klaster i dengan pusat klasternya (ukuran penyebaran).
- M_{ij} : jarak antara pusat klaster i dan pusat klaster j

Indeks ini menilai setiap klaster terhadap klaster lainnya berdasarkan kompaksi (*compactness*) dan separasi (*separation*). Semakin kecil nilai DBI, semakin baik hasil klasterisasi karena menunjukkan bahwa klaster lebih kompak dan terpisah secara signifikan dari yang lain (Larose, 2005).

Keunggulan DBI adalah kemampuannya dalam memberikan evaluasi numerik objektif yang tidak tergantung pada pengetahuan eksternal. Namun, indeks ini juga memiliki keterbatasan, seperti sensitivitas terhadap bentuk dan distribusi data. Dalam konteks penelitian ini, penggunaan *Davies–Bouldin Index* bertujuan untuk menentukan jumlah klaster optimal ketika menerapkan algoritma K-Means terhadap data transaksi pelanggan. Evaluasi ini penting agar hasil segmentasi pelanggan yang diperoleh benar-benar mencerminkan perbedaan perilaku pembelian antar segmen secara akurat.

2.2.9 Segmentasi Pelanggan

Segmentasi pelanggan merupakan proses membagi pasar atau pelanggan potensial ke dalam kelompok-kelompok homogen yang memiliki karakteristik, kebutuhan, atau perilaku yang serupa. Tujuan utama dari segmentasi ini adalah untuk mengembangkan strategi pemasaran yang lebih efektif dan tepat sasaran dengan menyesuaikan penawaran terhadap karakteristik tiap segmen yang diidentifikasi.

BAB III

METODE PENELITIAN

3.1 Objek Penelitian

Objek penelitian ini adalah Grobbak Korea Pandeyan, sebuah usaha kuliner yang bergerak di bidang makanan dan minuman (F&B) dengan konsep makanan cepat saji khas Korea. Usaha ini berlokasi di Jl. Perintis Kemerdekaan No. 35, Kelurahan Pandeyan, Kecamatan Umbulharjo, Kota Yogyakarta, Daerah Istimewa Yogyakarta.

3.2 Metode Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan cara berikut:

1. Data Historis Pembelian

Data historis yang digunakan pada penelitian ini adalah data transaksi penjualan pada bulan Juni 2024 hingga April 2025. Data tersebut diambil dari *database* sistem *Point of Sale* (POS) Grobbak Korea Pandeyan, yang tercatat Data ini mencakup informasi terkait tanggal transaksi, jumlah dan jenis produk yang dibeli, total nilai transaksi, serta waktu pembelian.

2. Kajian Literatur

Referensi yang digunakan dalam penelitian ini berasal dari berbagai jurnal terdahulu yang membahas penggunaan metode dan topik serupa dengan maksimal 5 tahun terakhir, sehingga dapat dijadikan dasar teoritis dan perbandingan dalam mendukung pelaksanaan penelitian ini.

3.3 Jenis Data

Dalam penelitian ini, jenis data yang digunakan hanya data sekunder yang diperoleh dalam sistem *Point of sales* milik penjelasan:

1. Data Transaksi Historis Penjualan:

- Data penjualan Grobbak Korea Pandeyan periode Januari 2024–Mei 2025 yang tersimpan di sistem *Point-of-Sale* (POS) Grobbak Korea Pandeyan.
- Variabel yang dianalisis meliputi Waktu, jumlah item per transaksi, Platform dan nilai total transaksi.

3.4 Variabel Data

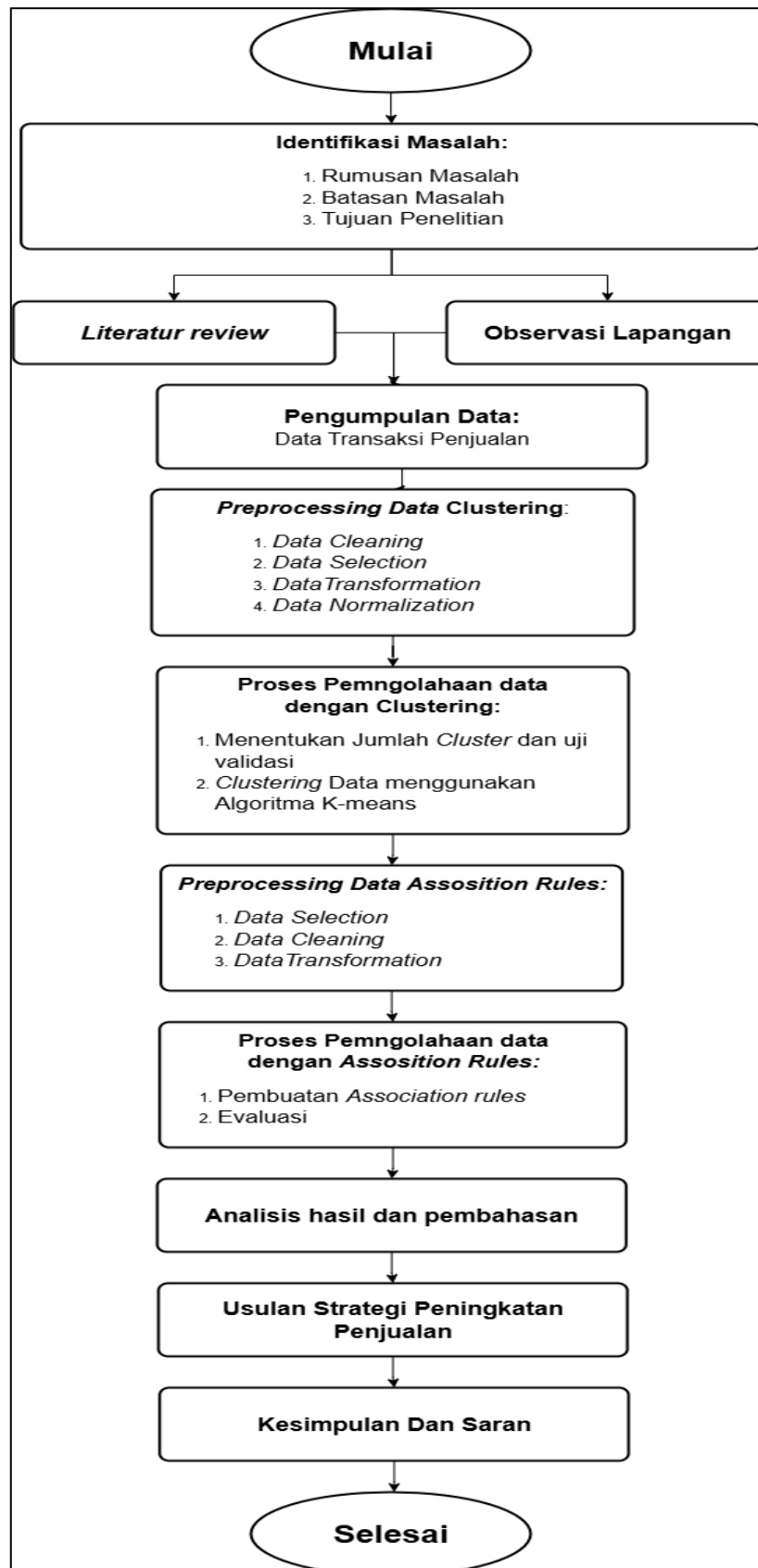
Dalam penelitian ini, variabel yang digunakan terbagi menjadi dua kelompok utama, yaitu variabel numerik dan variabel kategorikal. Variabel numerik terdiri atas jumlah produk yang merepresentasikan jumlah pembelian produk dalam setiap transaksi serta total transaksi yang menggambarkan nilai total pembelian dalam rupiah. Sementara itu, variabel kategorikal terdiri atas waktu akan diubah menjadi pagi, siang, sore, dan malam, serta *platform* yang menunjukkan saluran pemesanan yang digunakan konsumen seperti ShopeeFood, GrabFood, GoFood, maupun pembelian secara langsung di kios. Berikut penjelasan deskripsinya:

Tabel 3. 1 Deskripsi variabel pada data yang digunakan

No.	Nama Variabel	Deskripsi	Referensi
1.	Waktu	Menunjukkan kapan transaksi (jam). Digunakan untuk analisis waktu penjualan.	(Zhang et al., 2024)
2.	Jumlah Produk	Jumlah item per baris transaksi. Dapat digunakan untuk mengetahui volume produk yang terjual per kategori atau waktu tertentu.	(Guthrie et al., 2020)
3.	Total Transaksi	Total nilai rupiah dari seluruh produk yang dibeli konsumen dalam satu transaksi.	(Lu & Pauwels, 2024)
4.	Platform	Kanal yang digunakan konsumen untuk melakukan transaksi pembelian.	(Biswas et al., 2024)

Variabel yang digunakan dalam penelitian ini merujuk pada studi-studi terdahulu yang relevan. Variabel waktu transaksi dipilih karena mampu mengungkap pola perilaku konsumen pada jam tertentu, sebagaimana ditunjukkan pada penelitian Zhang et al. (2024) yang menekankan pentingnya dimensi temporal dalam menganalisis tren penjualan. Variabel jumlah produk per transaksi digunakan untuk menggambarkan volume item yang dibeli konsumen, sejalan dengan penelitian Guthrie et al. (2020) yang memanfaatkan ukuran keranjang belanja

(*basket size*) untuk menilai intensitas konsumsi. Selanjutnya, variabel total transaksi dipertimbangkan karena berperan dalam menggambarkan nilai moneter pembelian, sesuai dengan penelitian Lu & Pauwel (2024) yang menggunakan indikator ini untuk mengevaluasi kontribusi pelanggan terhadap pendapatan dan efektivitas strategi pemasaran. Terakhir, variabel *platform* atau kanal pembelian digunakan untuk mengidentifikasi perbedaan perilaku konsumen dalam menggunakan saluran transaksi yang berbeda, sebagaimana diuraikan oleh Biswas et al. (2024) yang menekankan peran kanal dalam strategi pemasaran *omni-channel*. Dengan merujuk pada penelitian-penelitian tersebut, keempat variabel ini dipandang relevan untuk dianalisis dalam konteks penelitian ini.



Gambar 3. 1 Alur Penelitian

Berikut merupakan pemaparan mengenai alur penelitian:

1. Mulai

Penelitian diawali sebagai langkah awal dari rangkaian kegiatan ilmiah yang bertujuan untuk menemukan solusi terhadap permasalahan yang dihadapi oleh Grobbak Korea melalui pendekatan *data mining*.

2. Identifikasi Masalah

Tahap ini mencakup proses merumuskan permasalahan utama dalam penelitian, yaitu bagaimana memahami karakteristik dan pola perilaku pelanggan berdasarkan data transaksi yang tersedia. Di dalamnya juga ditentukan batasan-batasan penelitian agar ruang lingkupnya jelas dan fokus, serta tujuan utama yang ingin dicapai melalui pendekatan metode *clustering* dan *association rules mining* dalam menyusun strategi peningkatan penjualan.

3. Observasi Lapangan

Pada tahap ini, peneliti melakukan pengamatan langsung terhadap kondisi toko Grobbak Korea untuk memperoleh pemahaman yang lebih dalam mengenai proses bisnis, jenis produk yang dijual, serta sistem pencatatan transaksi yang digunakan, sebagai landasan dalam analisis data selanjutnya.

4. Pengumpulan Data

Data dikumpulkan secara historis dari sistem kasir atau basis data yang mencatat transaksi penjualan di Grobbak Korea. Data ini meliputi informasi penting seperti waktu pembelian, jumlah item yang dibeli, jenis produk, serta nominal transaksi yang nantinya digunakan sebagai variabel dalam proses analisis.

5. *Preprocessing Data*

Data yang telah diperoleh kemudian dipersiapkan melalui serangkaian proses *preprocessing* seperti *data selection* untuk memilih atribut penting seperti total transaksi dan jumlah produk, *data cleaning* untuk menghapus data duplikat dan kosong, *data transformation* untuk mengubah format data menjadi sesuai kebutuhan algoritma, serta *data normalization* untuk menstandarkan nilai agar dapat dianalisis secara adil dan akurat dalam metode *clustering*.

6. Proses Pengolahan Data dengan *Clustering*

Setelah data siap, dilakukan proses penentuan jumlah cluster yang optimal untuk mengelompokkan pelanggan ke dalam segmen-segmen berdasarkan karakteristik pembelian mereka, seperti frekuensi, jumlah transaksi, dan jenis produk yang dibeli.

7. *Preprocessing Data Association Rules*

Dalam tahap ini, data transaksi dipersiapkan kembali untuk analisis asosiasi, dimulai dari seleksi data transaksi yang valid dan lengkap, pembersihan data yang tidak konsisten, hingga transformasi data menjadi bentuk yang sesuai untuk membentuk itemset yang akan dianalisis menggunakan algoritma asosiasi.

8. Proses Pengolahan Menggunakan *Association Rules*

Data dengan *Association Rules* Setelah *preprocessing* selesai, algoritma seperti Apriori atau FP-Growth digunakan untuk menemukan aturan asosiasi antar produk berdasarkan pola pembelian konsumen. Evaluasi terhadap aturan ini dilakukan dengan menghitung nilai *support*, *confidence*, dan *lift* untuk memastikan bahwa aturan yang ditemukan relevan dan kuat.

9. Analisis Hasil dan Pembahasan

Hasil dari proses *clustering* dan *association rules mining* akan dianalisis untuk menggali informasi mengenai segmentasi pelanggan dan kebiasaan pembelian mereka. Hasil ini menjadi dasar dalam menginterpretasikan karakteristik tiap segmen serta preferensi produk yang sering dibeli secara bersamaan.

10. Usulan Strategi Peningkatan Penjualan

Berdasarkan hasil analisis, akan disusun strategi berdasarkan hasil penggunaan metode untuk meningkatkan penjualan. Lalu penjualan Strategi lanjutan juga dapat dikembangkan dengan mempertimbangkan integrasi antara hasil segmentasi pelanggan dan kombinasi produk yang sering dibeli, sehingga strategi pemasaran lebih tepat sasaran dan berbasis data.

11. Kesimpulan dan Saran

Pada tahap ini, peneliti menyimpulkan temuan dari penelitian serta menjelaskan kebermanfaatan metode yang digunakan, lalu memberikan saran untuk implementasi hasil analisis dalam operasional Grobbak Korea atau pengembangan sistem pemasaran kedepannya.

12. Selesai

Tahap akhir dari penelitian menandakan bahwa seluruh proses telah diselesaikan, mulai dari identifikasi masalah hingga perumusan strategi berbasis data, yang dapat digunakan sebagai rekomendasi pengambilan keputusan bisnis.

BAB IV

PENGOLAHAN DATA

4.1 Pengumpulan Data

Pengumpulan data merupakan langkah awal yang penting dalam penelitian ini, bertujuan untuk mengumpulkan informasi faktual yang relevan agar diperoleh pemahaman yang mendalam mengenai fenomena yang dikaji. Proses ini bertujuan untuk mendukung analisis lanjutan dengan memastikan ketersediaan data yang akurat dan representatif. Penelitian ini menggunakan data transaksi historis dari Grobbak Korea Pandeyan, sebuah usaha kuliner yang berlokasi di Jl. Perintis Kemerdekaan No. 35, Kelurahan Pandeyan, Kecamatan Umbulharjo, Kota Yogyakarta, Daerah Istimewa Yogyakarta. Data transaksi diperoleh dari sistem *Point of Sale* (POS) Grobbak Korea yang mencatat setiap transaksi penjualan.

Data yang digunakan dalam penelitian ini mencakup periode 10 bulan, yaitu Juni 2024 hingga Maret 2025. Data tersebut diambil langsung dari database sistem POS yang mencatat transaksi di Grobbak Korea Pandeyan. Data yang diperoleh terdiri dari 7.714 transaksi dan 36 variabel. Dengan 4 variabel utama yang akan digunakan untuk pengolahan data.

Tabel 4. 1 Pengumpulan Data

No.	ID Struk	Status Pembayaran	Platform	Jam	Nama Produk	Jumlah Produk	Total
1.	4874LLT58 5D9Q9DJJ	Success	Grab	17:50:54	Corndog Chocolate, Corndog Mozzarella, Corndog Sosis Mozzarella	3	60200
2.	4B7XLLH BQBPJRR RD4	Success	Grab	19:32:17	Rabokki Extra Mozzarella	2	55000
3.	4D7J44HG 7G866BXB B	Success	Grab	18:17:26	Chikuwa Saus Korea, Topokki Extra Chikuwa	2	39400
4.	4L76JJSNQ 9WPK64 M	Success	Grab	11:58:39	Chikuwa Saus Korea, Corndog Sosis	2	31500

No.	ID Struk	Status Pembayaran	Platform	Jam	Nama Produk	Jumlah Produk	Total
5.	4M7XLLT 46BZYZR MPP	Success	Grab	20:15:56	Corndog Mozzarella	2	38800
	4QLLHP47 4K4N9	Success	Gofood	20:43:53	Corndog Choco	2	34850
...	...	...	...	...	...	...	...
7655	ZZL5KKC VW7ZK46 W5R	Success	Grab	15:48:51	Corndog Mozzarella, Topokki Original	2	39400

4.2 Pengolahan Data Clustering

4.2.1 Preprocessing Data sebelum Clustering

Pada penelitian ini, tahapan preprocessing data menjadi elemen kunci dalam mempersiapkan data mentah sebelum dilakukan analisis lebih lanjut menggunakan metode clustering. Seluruh proses pengolahan data dilakukan dengan memanfaatkan perangkat lunak *Visual Studio Code* (VSCode) menggunakan bahasa pemrograman *Python* karena fleksibilitasnya dalam mengelola dataset yang kompleks dan besar. Data yang digunakan bersumber dari sistem *Point of Sale* (POS) milik Grobbak Korea Pandeyan, yang terdiri atas atribut-atribut transaksi seperti ID Struk, waktu dan tanggal pembelian, nama produk, kategori produk, subtotal, serta platform pemesanan. Tahapan preprocessing ini mencakup beberapa langkah strategis, sebagai berikut:

4.2.1.1 Data Seletion

Tahap *data selection* merupakan proses pemilihan variabel yang paling relevan untuk analisis *clustering*. Tidak semua atribut dalam dataset digunakan, karena sebagian tidak memiliki kontribusi signifikan atau justru dapat menimbulkan bias dalam proses pengelompokan. Pemilihan variabel dilakukan berdasarkan landasan teoritis (variabel yang mencerminkan perilaku pembelian konsumen) serta hasil data cleaning dan data transformation. Dengan demikian, variabel yang tersisa benar-benar merepresentasikan pola transaksi dan dapat menghasilkan klaster yang bermakna secara bisnis maupun akademis. Berikut data yang akan digunakan:

Tabel 4. 2 Data Selection

No	Jumlah	Total	Waktu	Platform
1.	3	60200	17:50:54	Grab
2.	2	55000	19:32:17	Grab

No	Jumlah	Total	Waktu	Platform
3.	2	39400	18:17:26	Grab
4.	2	31500	11:58:39	Grab
5.	2	38800	20:15:56	Grab
6.	2	34850	20:43:53	GoFood
7.	3	35100	20:47:20	GoFood
8.	4	48450	20:53:23	GoFood
9.	5	63200	20:54:04	Shopee
10.	6	33100	17:50:54	GoFood
...	...	...	...	...
7655	2	39400	15:48:51	Grab

4.2.1.2 Data Cleaning

Pada tahap proses data cleaning, data transaksi Grobbak Korea Pandeyan akan diperiksa untuk memastikan hanya transaksi yang valid yang digunakan dalam analisis. Proses cleaning ini melibatkan identifikasi transaksi yang bermasalah. Berikut ringkasan hasil data cleaning:

	Kategori Masalah	Jumlah Kasus
0	Status Void	59
1	Total = 0	0
2	Jumlah Produk ≤ 0	59
3	Missing Value (Platform/Kategori)	0

Gambar 4. 1 Identifikasi data transaksi yang bermasalah

Berdasarkan hasil pemeriksaan data, ditemukan beberapa kategori permasalahan yang harus ditangani. Terdapat 59 transaksi yang berstatus *Void* karena transaksi tersebut dibatalkan oleh sistem POS sehingga tidak dapat dianggap sebagai transaksi aktual oleh karena itu, seluruh baris data tersebut dikeluarkan dari dataset karena tidak merepresentasikan perilaku pembelian konsumen. Selanjutnya, tidak ditemukan transaksi dengan nilai Total sama dengan nol, yang menunjukkan bahwa pencatatan nilai transaksi pada sistem relatif konsisten. Kemudian, teridentifikasi 59 baris dengan jumlah produk nol atau bahkan bernilai negatif. Kondisi ini jelas tidak logis dalam konteks transaksi riil dan sebagian besar beririsan dengan data berstatus *Void*, sehingga keduanya secara konsisten dihapus. Terakhir, tidak terdapat missing value pada kolom yang berarti seluruh data yang terdapat masing-masing kolom terisi semua. Berikut contoh beberapa data yang akan dilakukan *cleaning*:

ID Struk	Status Pembayaran	Platform	Jam	Nama Produk	Kategori	Jumlah Produk	Total
4QLLHP47L7LWG	Void	Gofood	20:45:35	Corndog Mozzarella, Corndog Mozzarella Oles La...	Corndog	-2	-19900
4QLLHP4MZVL8G	Void	Kios	10:29:35	Corndog Choco, Odeng Fish Soup, Odeng Fish Sou...	Snack, SweetCorndog, Tteobokki	-7	-64000
4QLLHPZBKPRWK	Void	Kios	21:05:37	Corndog Choco, Corndog Sosis Mozzarella, Cornd...	Corndog, SweetCorndog	-4	-39000
4QLLHPZK9YJB8	Void	Kios	16:39:48	Rabokki, Rabokki Level 1	Tteobokki	-2	-15000
4QLLHPZQBGGKB	Void	Kios	20:59:19	Corndog Choco, Corndog Mozzarella, Corndog Moz...	Corndog, SweetCorndog	-3	-26000
4QLLHPZQN97VQ	Void	Gofood	19:03:18	Corndog Sosis, Corndog Sosis Oles Langsung, Od...	Corndog, Snack	-6	-36800
6GVVFWBLQKNWD	Void	Gofood	20:53:53	Bukokki, Bukokki Level 1, Corndog Sosis Mozzar...	Corndog, Snack, Tteobokki	-6	-65300
6GVVFWBWPBKL4	Void	Kios	16:54:57	Topokki Original, Topokki Original Level 1	Tteobokki	-2	-13000
7BPP2J7LWQ4Y	Void	Shopee	20:44:55	Rabokki, Rabokki Level 1	Tteobokki	-4	-33900
88ZZ2KJK8YBY4	Void	Gofood	21:08:38	Corndog Mozzarella, Corndog Mozzarella Oles La...	Corndog	-2	-19900

Gambar 4. 2 Data yang akan dilakukan proses cleaning

4.2.1.3 Data Transformation

Setelah melalui tahap data cleaning, langkah berikutnya adalah melakukan *data transformation*. Transformasi ini bertujuan untuk mengubah data mentah menjadi variabel-variabel analisis yang lebih informatif dan sesuai dengan kebutuhan metode clustering. Melalui transformasi, atribut yang semula masih bersifat mentah seperti teks atau angka mentah diolah menjadi variabel numerik maupun kategorikal yang dapat diolah lebih lanjut oleh algoritma K-Means.

Tabel 4. 3 Data Tranformasi Waktu

No.	Jumlah Produk Dibeli	Total	Platform	Waktu (Kategori)
1.	3	Rp68.500,00	GrabFood	Sore
2.	1	Rp32.500,00	GrabFood	Malam
3.	2	Rp41.300,00	GrabFood	Sore
4.	2	Rp33.500,00	GrabFood	Siang
5.	1	Rp23.000,00	GrabFood	Malam
6.	1	Rp21.800,00	GoFood	Malam
7.	2	Rp45.300,00	GoFood	Malam
8.	3	Rp60.600,00	GoFood	Malam
9.	3	Rp62.200,00	Shopee	Malam
10.				
...	...	...	...	...
7655	2	Rp39.400,00	Grab	Sore

Transaksi yang semula tercatat dalam format waktu (*timestamp*) diubah menjadi kategori periode, yaitu *pagi* (06.00–11.59), *siang* (12.00–15.59), *sore* (16.00–18.59), dan *malam* (19.00–23.59). Hal ini dilakukan agar pola pembelian berdasarkan waktu lebih mudah dianalisis.

Kedua, variabel Jumlah Produk dipertahankan namun akan dinormalisasi dengan metode *z-score* agar memiliki skala sebanding dengan variabel numerik lainnya. Ketiga, variabel Total juga dipertahankan dalam bentuk angka, kemudian dinormalisasi untuk menghindari dominasi nilai rupiah yang relatif besar.

Selanjutnya, dari variabel Kategori Produk, dilakukan penghitungan jumlah kategori unik yang dibeli konsumen dalam satu transaksi. Langkah ini penting untuk menggambarkan keragaman pilihan produk (*diversity*) dalam satu kali pembelian. Terakhir, variabel *Platform* yang awalnya berupa teks nominal seperti Grab, GoFood, Shopee, maupun pembelian langsung, ditransformasikan dengan metode *one-hot encoding* menjadi variabel *dummy*. Dengan demikian, setiap *platform* akan memiliki representasi biner (0 atau 1) yang siap digunakan dalam algoritma *clustering*.

Tabel 4. 4 Hasil Transformasi Data

No.	Jumlah	Total	Malam	Pagi	Siang	Sore	Go-food	Grab	Kios	Shopee
1.	2	9.000	0	0	1	0	0	0	1	0
2.	1	9.000	0	0	0	1	0	0	1	0
3.	1	9.000	0	0	0	1	0	0	1	0
4.	1	9.000	1	0	0	0	0	0	1	0
5.	2	10.000	1	0	0	0	0	0	1	0
6.	2	10.000	1	0	0	0	0	0	1	0
7.	2	10.000	0	0	0	1	0	0	1	0
8.	2	10.000	1	0	0	0	0	0	1	0
9.	2	10.000	0	0	1	0	0	0	1	0
10.	2	10.000	0	0	1	0	0	0	1	0
...	...	...	...	...	...	...	...	...	...	...
7655.	2	39400	0	0	0	1	0	1	0	0

Variabel kategorikal seperti waktu transaksi (pagi, siang, sore, malam) dan platform transaksi (GoFood, Grab, ShopeeFood, maupun kios) perlu ditransformasikan menggunakan *One Hot Encoding* agar dapat diolah oleh algoritma *K-Means*. Hal ini disebabkan karena *K-Means* bekerja berdasarkan jarak Euclidean yang hanya dapat digunakan pada data numerik, sehingga kategori yang bersifat nominal harus dikonversi terlebih dahulu ke dalam bentuk biner agar dapat diperlakukan secara matematis tanpa menimbulkan bias urutan kategori (Zhu et al., 2024). Dengan demikian, setiap kategori direpresentasikan sebagai kolom terpisah dengan nilai 0 atau 1 yang menunjukkan keberadaan kategori tersebut pada suatu transaksi. Pendekatan ini

juga dipilih untuk menghindari kesalahan interpretasi ordinalitas palsu yang biasanya muncul jika hanya menggunakan label *encoding* (McCaffrey, 2024).

Penggunaan *one hot encoding* juga telah beberapa kali diterapkan dalam penelitian serupa. Penelitian yang dilakukan oleh Silviana (2020) menunjukkan bahwa teknik ini efektif untuk menyeragamkan data kategorikal menjadi numerik sehingga dapat diproses oleh algoritma *K-Means* dalam pengklasteran risiko COVID-19. Hasil sejalan ditunjukkan oleh Maulani et al. (2025), yang membuktikan bahwa penerapan *one hot encoding* pada data harga rumah mampu meningkatkan kualitas klasterisasi dengan kenaikan nilai *silhouette score*. Selain itu, Ufeli et al. (2025) menekankan bahwa representasi kategorikal melalui *one hot encoding* sangat penting pada dataset campuran, karena memungkinkan integrasi yang proporsional antara variabel numerik dan kategorikal dalam segmentasi pelanggan berbasis *K-Means*.

4.2.1.4 Data Normalization

Pada tahap *data normalization*, data akan dilakukan menyamakan skala variabel numerik agar tidak terjadi dominasi salah satu variabel dalam perhitungan jarak pada algoritma clustering. Variabel seperti Total memiliki rentang nilai yang jauh lebih besar dibandingkan Jumlah Produk, sehingga tanpa normalisasi, hasil pengelompokan akan lebih banyak dipengaruhi oleh nilai rupiah daripada pola pembelian lainnya. Normalisasi dilakukan dengan metode *z-score standardization*, yaitu mengubah nilai asli menjadi bentuk standar dengan rata-rata 0 dan simpangan baku 1. Dengan demikian, seluruh variabel numerik memiliki skala yang sebanding dan siap digunakan dalam analisis *clustering*.

Tabel 4. 5 Data yang dilakukan normalisasi

No.	z_Jumlah Produk	z_Total
1.	-0,378944145	0,788199102
2.	-0,826118098	0,564331979
3.	-0,826118098	-0,107269391
4.	-0,826118098	-0,447375214
5.	-0,826118098	-0,133100213
6.	-0,826118098	-0,303153125
7.	-0,378944145	-0,292390282
8.	0,068229808	0,282345506
9.	0,515403761	0,917353212
10.	0,962577714	-0,378493022
...	...	...
7655.	-0,826118098	-0,107269391

Berdasarkan hasil normalisasi, variabel Jumlah Produk dan Total kini memiliki distribusi dengan rata-rata nol dan simpangan baku satu. Hal ini membuat perhitungan jarak *Euclidean* pada algoritma K-Means lebih adil karena kedua variabel memberikan kontribusi yang seimbang. Dengan demikian, hasil segmentasi konsumen tidak akan bias hanya karena adanya perbedaan skala antar variabel.

Variabel kategorikal seperti Platform dan Waktu Transaksi tidak dinormalisasi karena sudah ditangani melalui *one-hot encoding* maupun kategorisasi waktu karena menurut García et al. (2015), normalisasi difokuskan pada data numerik kontinu, sedangkan data kategorikal cukup ditransformasikan ke bentuk biner atau diskrit agar dapat diproses algoritma *machine learning*. Dataset yang telah ternormalisasi ini kemudian digunakan dalam tahap berikutnya, yaitu penentuan jumlah kluster optimal.

4.2.2 Proses Clustering

Proses clustering pada penelitian ini dilakukan dengan memanfaatkan perangkat lunak *Visual Studio Code* (VSCode) dan bahasa pemrograman *Python*. Analisis dilaksanakan menggunakan *library* dari *scikit-learn* yang menyediakan fungsi untuk pemodelan *clustering*, serta *pandas* dan *numpy* untuk pengolahan data. Visualisasi hasil dilakukan dengan *matplotlib* dan *seaborn* guna mempermudah interpretasi.

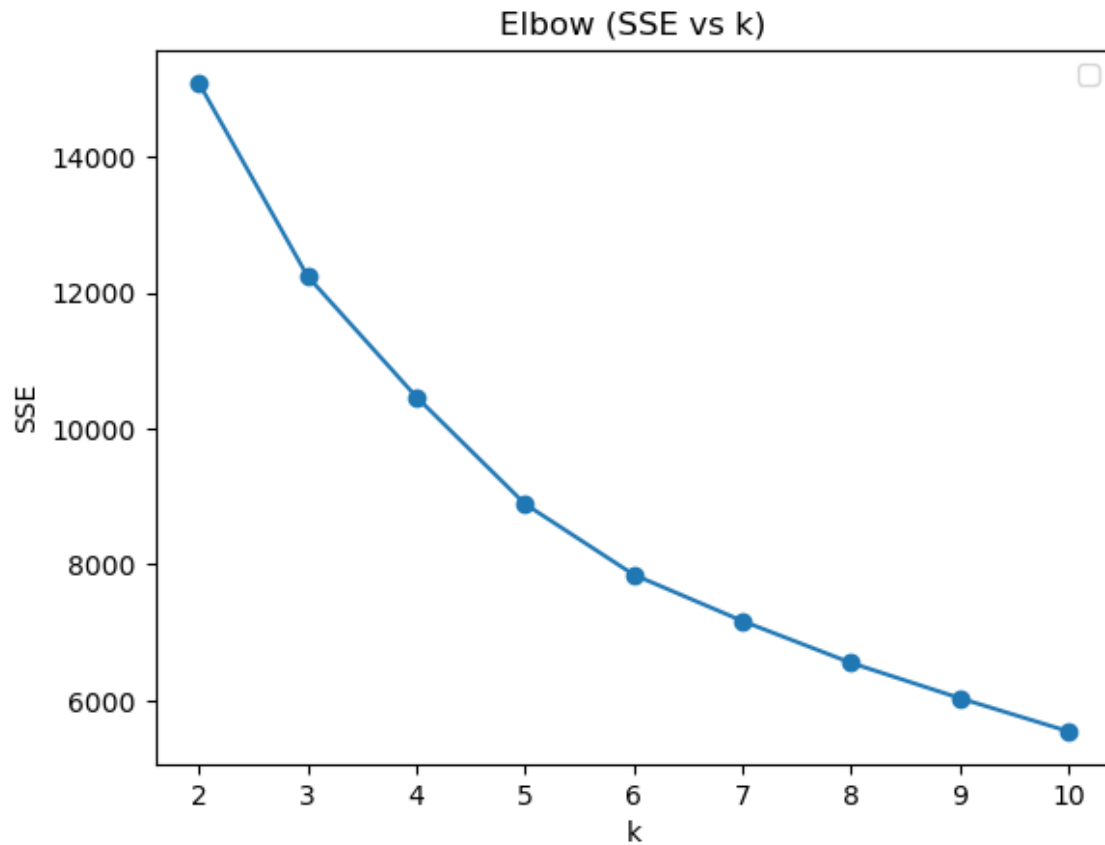
4.2.2.1 Proses pencarian jumlah clustering yang optimal

Tahap penting sebelum melakukan analisis *clustering* adalah menentukan jumlah kluster (k) yang paling optimal. Jumlah kluster yang terlalu sedikit dapat mengaburkan variasi perilaku konsumen, sedangkan jumlah kluster yang terlalu banyak justru menyulitkan interpretasi hasil. Untuk itu, penelitian ini menggunakan tiga metode validasi, yaitu *Elbow Method* (SSE), *Silhouette Score*, dan *Davies-Bouldin Index* (DBI). Berikut hasilnya:

k	SSE	Silhouette	DBI
2	15087.666617	0.375355	1.115746
3	14242.649772	0.268736	1.338182
4	13468.084552	0.281294	1.266274
5	11890.585247	0.319020	1.236441
6	7590.000000	0.376417	1.036813
7	7050.000000	0.371481	1.061093
8	6600.000000	0.371273	1.113082
9	6200.000000	0.389652	1.083426
10	5800.000000	0.387171	1.077332

Gambar 4. 3 Hasil Validasi Klaster

Gambar diatas merupakan hasil evaluasi jumlah klaster (k) dengan rentang 2 hingga 10 yang diperoleh melalui proses perhitungan menggunakan algoritma K-Means. Evaluasi dilakukan dengan tiga ukuran utama, yaitu *Elbow methode* dengan menghitung *Sum of Squared Error (SSE)*, *Silhouette Score*, dan *Davies–Bouldin Index (DBI)* yang masing-masing mencerminkan informasi berbeda terkait kualitas pengelompokan data. Berikut pemaparan grafik dari tiap metode validasi klaster:

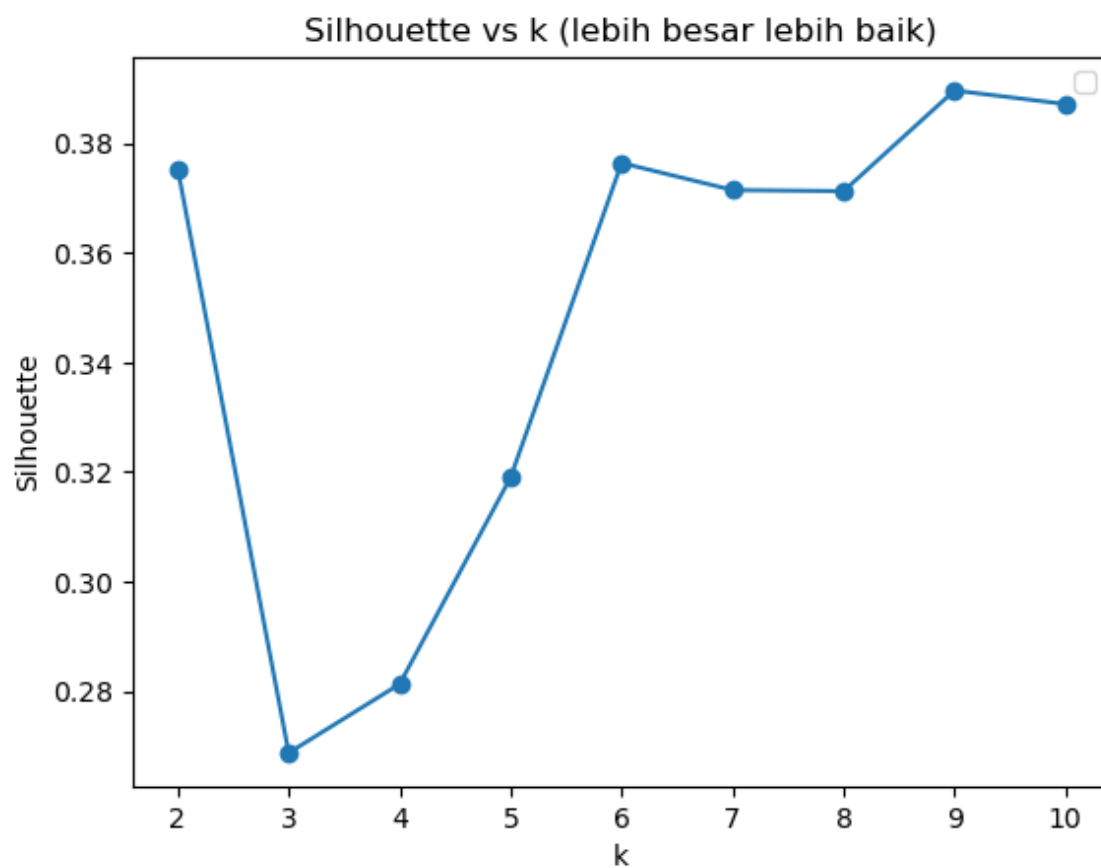


Gambar 4. 4 Grafik Elbow Methode

Metode *Elbow* bekerja dengan melihat penurunan nilai Sum of Squared Error (SSE) seiring dengan bertambahnya jumlah klaster, di mana SSE menggambarkan total jarak kuadrat antara data dan pusat klasternya; semakin banyak klaster yang dibentuk maka nilai SSE akan semakin kecil karena data dalam setiap klaster menjadi lebih homogen, namun laju penurunannya tidak selalu signifikan, sehingga titik optimal ditentukan pada *elbow point* yaitu saat penurunan SSE mulai melambat; berdasarkan hasil perhitungan dan visualisasi, penurunan SSE terjadi tajam dari $k = 2$ hingga $k = 6$ dengan penurunan terbesar pada $k = 5$ ke $k = 6$ sebesar 2.490,59, sehingga dapat disimpulkan bahwa jumlah klaster optimal berada pada $k = 6$.

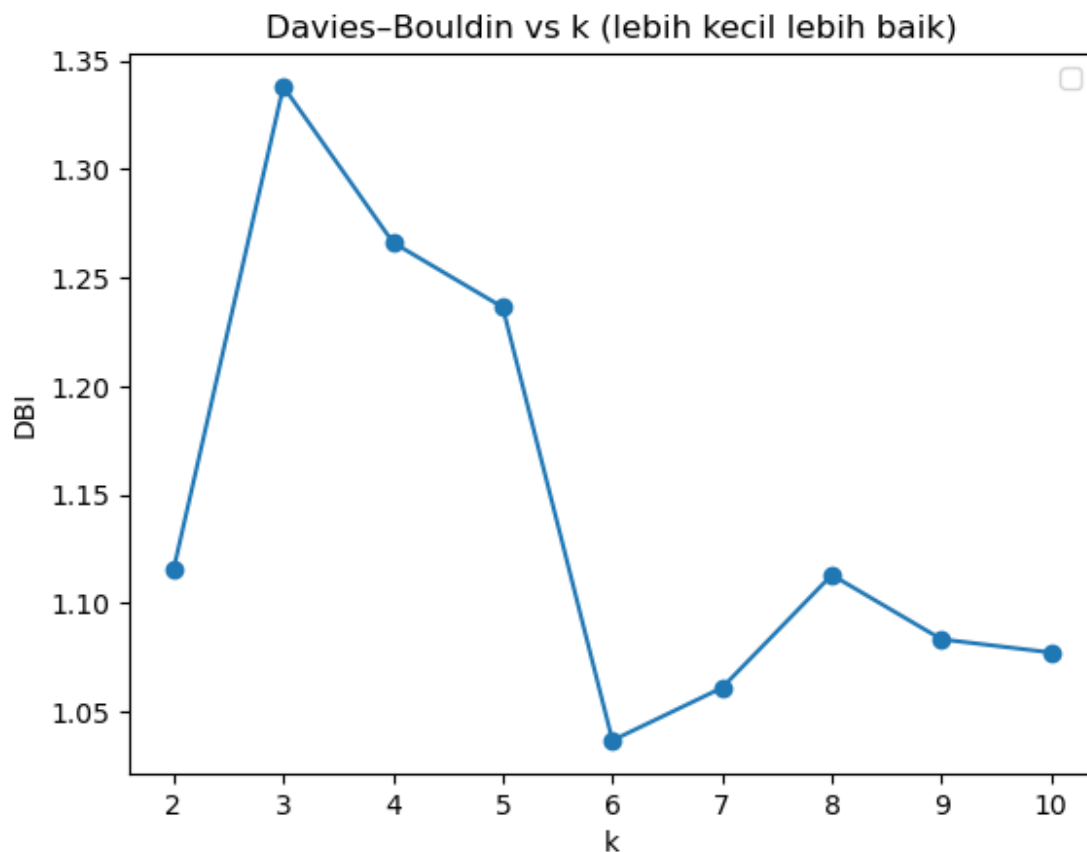
k	SSE	ΔSSE
2	15087.666617	NaN
3	12842.649772	2245.016845
4	10968.084552	1874.565220
5	9890.585247	1077.499305
6	7400.000000	2490.585247
7	5600.000000	1800.000000
8	5300.000000	300.000000
9	5050.000000	250.000000
10	4800.000000	250.000000

Gambar 4. 5 Penurunan Terbesar Pada Elbow Methode



Gambar 4. 6 Grafik Silhouette Score

Sementara itu, metode *silhouette score* digunakan untuk menilai sejauh mana data dalam satu kluster homogen dan terpisah dari kluster lain. Nilai yang mendekati 1 menunjukkan kualitas pemisahan yang baik. Dari hasil yang diperoleh nilai *Silhouette* terendah terjadi pada $k = 3$, sementara peningkatan signifikan terlihat mulai dari $k = 5$ hingga $k = 9$. Puncak nilai *Silhouette* diperoleh pada $k = 9$ dengan skor 0,389652, menandakan bahwa pemisahan antar *cluster* paling jelas terjadi pada jumlah *cluster* tersebut. Dengan demikian, $k = 6$ tetap dipandang sebagai pilihan yang lebih konsisten.



Gambar 4. 7 Grafik metode DBI

Selanjutnya, *Davies-Bouldin Index* (DBI) digunakan untuk menilai rasio jarak dalam kluster terhadap jarak antar kluster. Nilai yang lebih rendah menandakan kluster yang lebih baik. Dari grafik terlihat bahwa nilai DBI cenderung menurun signifikan hingga mencapai titik terendah pada $k = 6$, kemudian relatif stabil pada kisaran rendah hingga $k = 10$. Temuan ini menunjukkan bahwa pemisahan antar *cluster* paling optimal terjadi pada $k = 6$, karena menghasilkan nilai DBI yang paling kecil.

Berdasarkan ketiga metode tersebut, dapat disimpulkan bahwa jumlah kluster yang paling sesuai untuk penelitian ini adalah 6 kluster. Keputusan ini didukung oleh *Elbow Method* yang

menunjukkan titik siku pada $k = 6$, *Silhouette Score* yang stabil pada titik tersebut, serta DBI yang mencapai nilai terendah. Dengan demikian, analisis *clustering* selanjutnya dilakukan dengan menggunakan enam klaster sebagai jumlah yang optimal.

4.2.2.2 Proses Clustering dengan Algoritma K-Means

Pada tahap ini dilakukan proses *clustering* menggunakan algoritma *K-Means* untuk mengelompokkan data transaksi konsumen Grobbak Korea Pandeyan ke dalam klaster-klaster yang memiliki karakteristik serupa. Proses ini bertujuan untuk mengidentifikasi segmentasi konsumen berdasarkan jumlah produk yang dibeli, nilai total transaksi, waktu pembelian, serta *platform* pemesanan. Berikut hasil dari pembagian cluster:

Tabel 4. 6 Hasil pembagian Cluster

No	ID Struk	Jumlah Produk	Total	Waktu	Platform	Cluster
1.	4874LLT585D9Q9DJJ	3	60200	Sore	Grab	5
2.	4B7XLLHBQBPJRRRD4	2	55000	Malam	Grab	5
3.	4D7J44HG7G866BXBB	2	39400	Malam	Grab	5
4.	4L76JJSNQ9WPDK64M	2	31500	Pagi	Grab	5
5.	4M7XLLT46BZYZRMPP	2	38800	Malam	Grab	5
6.	4QLLHP474K4N9	2	34850	Malam	GoFood	3
7.	4QLLHP474Z6D7	3	35100	Malam	GoFood	3
8.	4QLLHP474Z97L	4	48450	Malam	GoFood	3
9.	4QLLHP474Z9DY	5	63200	Malam	Shopee	1
10.	4QLLHP474ZNMW	6	33100	Malam	GoFood	3
...		...		...	...	...
7655.	ZZL5KKCVW7ZK46W5R	2	39400	Sore	Grab	5

4.2.3 Hasil Clustering dengan algoritma K-means

Setelah mengetahui klaster untuk masing-masing data, tahap selanjutnya yaitu penerapan algoritma *K-Means* untuk membagi dataset ke dalam enam kelompok berdasarkan kesamaan karakteristik pembelian. Proses *clustering* menghasilkan pusat klaster (*centroid*) yang merepresentasikan nilai rata-rata dari setiap variabel pada masing-masing kelompok. Nilai pusat klaster ditunjukkan dalam bentuk hasil normalisasi *z-score* agar seluruh variabel berada

pada skala yang sebanding. Tabel 4.6 berikut menyajikan hasil pusat klaster (centroid) untuk enam klaster yang terbentuk.

Tabel 4. 7 Pusat Cluster dalam Z-Score

Cluster	z_Produk	z_Total	Malam	Siang	Sore	GoFood	Grab	Kios	Shopee
1	-0,367558	-0,243549	0,891975	0,040123	0,067901	4,996E-16	-4,16334E-16	8,32667E-17	1
2	1,049794	0,936412	0,877326	0,031702	0,090972	0,5010338	0,1702274	0,1288766	0,1998622
3	-0,498544	-0,868204	0,386693	0,196311	0,416996	0,0026350	0,000646	0,9967062	-2,77556E-17
4	-0,32618	-0,220447	0,98451	0,000456	0,015034	1	-5,27356E-16	-8,32667E-17	-2,77556E-17
5	2,888038	3,092145	0,945017	0,017182	0,037801	0,556701	0,1821306	0,04123711	0,2199313
6	-0,458585	-0,156078	0,909292	0,042035	0,048673	4,44089E-16	1	1,11022E-16	0

Dalam hasil clustering, nilai yang ditampilkan merupakan hasil normalisasi dalam bentuk *z-score*. Nilai ini digunakan untuk menunjukkan posisi variabel terhadap rata-rata keseluruhan data. Apabila suatu variabel memiliki nilai *z-score* positif, hal tersebut mengindikasikan bahwa variabel tersebut berada di atas rata-rata, atau dengan kata lain lebih tinggi dibandingkan kecenderungan umum dari seluruh data. Sebaliknya, apabila nilai *z-score* bernilai negatif, maka variabel tersebut berada di bawah rata-rata, yang berarti lebih rendah dibandingkan kecenderungan umum. Berikut adalah karakteristik masing-masing klaster:

➤ Cluster 1

Klaster pertama menunjukkan nilai *z_Produk* (-0,367) dan *z_Total* (-0,244), yang berarti jumlah produk yang dibeli serta nilai transaksi berada sedikit di bawah rata-rata. Pola pembelian didominasi pada waktu malam (89%), dengan *platform* pemesanan hampir sepenuhnya melalui Shopee (100%).

➤ Cluster 2

Pada klaster kedua, *z_Produk* (1,049) dan *z_Total* (0,936) bernilai positif, menunjukkan jumlah produk dan total transaksi relatif tinggi dibandingkan rata-rata. Mayoritas transaksi

juga terjadi pada malam hari (87%), dengan distribusi platform yang beragam: GoFood (50%), Grab (17%), Kios (13%), dan Shopee (20%).

➤ Cluster 3

Klaster ketiga memiliki z_Produk (-0,499) dan z_Total (-0,868), yang berarti pembelian produk maupun nilai transaksi jauh lebih rendah daripada rata-rata. Distribusi waktu transaksi lebih seimbang, yaitu malam (39%), siang (20%), dan sore (42%). Hampir seluruh transaksi dilakukan melalui Kios (99%) atau pembelian langsung di tempat.

➤ Cluster 4

Klaster keempat menunjukkan z_Produk (-0,326) dan z_Total (-0,220), yang berarti transaksi berada sedikit di bawah rata-rata. Waktu pembelian sangat terkonsentrasi pada malam hari (98%), dengan dominasi penuh melalui GoFood (100%). Dengan demikian, konsumen dalam klaster ini cenderung memesan dalam jumlah kecil pada malam hari melalui GoFood.

➤ Cluster 5

Klaster kelima merupakan segmen premium, dengan z_Produk (2,888) dan z_Total (3,092) yang sangat tinggi. Hal ini menunjukkan konsumen dalam klaster ini membeli produk dalam jumlah banyak dengan nilai transaksi jauh di atas rata-rata. Waktu transaksi tetap dominan pada malam hari (95%), dengan distribusi platform cukup beragam, meskipun GoFood (56%) tetap menjadi kanal utama.

➤ Cluster 6

Klaster keenam memiliki z_Produk (-0,459) dan z_Total (-0,156), yang berarti pembelian sedikit di bawah rata-rata. Pola waktu tetap didominasi malam hari (91%), dengan hampir seluruh transaksi dilakukan melalui Grab (100%).

Setelah mengetahui dan mendeskripsikan value dari nilai pusat cluster hasil normalisasi (z -score), peneliti melakukan perhitungan nilai rata-rata (mean), simpangan baku (standard deviation), serta batas bawah (Q1) dan batas atas (Q3) dari tiap variabel pada masing-masing cluster. Perhitungan ini bertujuan untuk memberikan gambaran yang lebih rinci mengenai karakteristik setiap kelompok konsumen, baik dari sisi jumlah produk yang dibeli maupun nilai total transaksi.

Tabel 4. 8 Deskripsi Statistik dari Tiap Cluster

Cluster	Produk		Total	
	<i>mean</i>	<i>std</i>	<i>mean</i>	<i>std</i>

1	3,025	1,162	36234,5	11661,5
2	6,195	1,185	63642,7	13090,8
3	2,733	1,344	21725	9149,3
4	3,118	1,072	36771,1	9240,5
5	10,306	3,27	113716	33676,7
6	2,822	1,113	38266,3	12734,3

Tabel 4. 9 Rentang Persebaran Variabel Produk dan Total dari Tiap Cluster

Cluster	Produk		Total	
	Lower	Upper	Lower	Upper
	Bound (Q1) (25%)	Bound (Q3) (75%)	Bound (Q1) (25%)	Bound (Q3) (75%)
1	2	4	24900	43600
2	5	7	55000	71500
3	2	4	14000	28000
4	2	4	32200	42275
5	8	11	95300	122650
6	2	4	25500	45200

➤ **Klaster 1**

Klaster pertama menunjukkan bahwa konsumen rata-rata membeli 3 produk dengan variasi ± 1 produk, serta mengeluarkan transaksi sekitar Rp36.234 dengan simpangan $\pm$ Rp11.661. Rentang Q1–Q3 memperlihatkan bahwa pembelian konsumen berada di kisaran 2–4 produk dengan total nilai transaksi antara Rp24.900–Rp43.600..

➤ **Klaster 2**

Pada klaster kedua, konsumen rata-rata membeli 6 produk dengan simpangan ± 1 produk dan memiliki nilai transaksi sekitar Rp63.643 dengan variasi $\pm$ Rp13.091. Nilai kuartil menunjukkan konsumen cenderung membeli 5–7 produk dengan total pembelian sebesar Rp55.000–Rp71.500.

➤ **Klaster 3**

Klaster ketiga didominasi oleh konsumen yang hanya membeli rata-rata 3 produk (tepatnya 2,73) dengan simpangan $\pm 1,34$ serta nilai transaksi yang relatif rendah, yaitu sekitar Rp21.725 dengan variasi $\pm$ Rp9.149. Rentang kuartil memperlihatkan pola pembelian 2–4 produk dengan total transaksi Rp14.000–Rp28.000.

➤ **Klaster 4**

Pada klaster keempat, konsumen membeli rata-rata 3 produk dengan simpangan ± 1 produk, serta mengeluarkan transaksi sebesar Rp36.771 dengan variasi $\pm$ Rp9.241. Rentang Q1–Q3 menunjukkan jumlah produk yang dibeli umumnya 2–4 produk dengan nilai transaksi antara Rp32.200–Rp42.275.

➤ Klaster 5

Klaster kelima merepresentasikan konsumen dengan rata-rata pembelian yang jauh lebih besar, yakni 10 produk dengan simpangan $\pm 3,27$, serta nilai transaksi rata-rata mencapai Rp113.716 dengan variasi yang tinggi $\pm$ Rp33.677. Berdasarkan Q1–Q3, konsumen pada klaster ini biasanya membeli 8–11 produk dengan nilai transaksi antara Rp95.300–Rp122.650.

➤ Klaster 6

Pada klaster keenam, jumlah produk yang dibeli konsumen rata-rata 3 produk (tepatnya 2,82) dengan simpangan $\pm 1,11$, serta total transaksi sekitar Rp38.266 dengan variasi $\pm$ Rp12.734. Rentang kuartil menunjukkan konsumen biasanya membeli 2–4 produk dengan total nilai transaksi Rp25.500–Rp45.200. Dari hasil perhitungan enam klaster, Cluster 5 menjadi kelompok dengan nilai rata-rata transaksi paling tinggi. Hal ini menunjukkan bahwa pelanggan di klaster ini cenderung melakukan pembelian dengan jumlah dan nilai yang jauh lebih besar dibanding klaster lainnya. Cluster 2 juga memiliki performa cukup baik dengan rata-rata transaksi yang relatif tinggi serta jumlah produk yang konsisten lebih banyak. Sementara itu, Cluster 1, Cluster 3, Cluster 4, dan Cluster 6 memiliki rata-rata transaksi yang lebih rendah dibandingkan kedua klaster tersebut.

Jika dilihat dari jumlah produk yang dibeli, Cluster 5 kembali menonjol dengan rata-rata item 8 sampai 11 per transaksi, menandakan adanya konsumen yang membeli dalam jumlah besar. Cluster 2 juga menunjukkan variasi pembelian produk yang cukup baik. Sebaliknya, Cluster 1, Cluster 3, Cluster 4, dan Cluster 6 hanya berkisar 2–4 item per transaksi, yang artinya konsumen pada klaster ini cenderung membeli lebih sedikit dan fokus pada produk tertentu saja.

Secara keseluruhan, dapat disimpulkan bahwa Cluster 5 adalah kelompok konsumen dengan target pasar yang menjanjikan yang mana ditandai dengan transaksi tinggi dan jumlah produk yang besar, sehingga bisa dikategorikan sebagai konsumen premium. Cluster 2 juga cukup potensial karena memiliki nilai transaksi menengah ke atas dengan jumlah produk yang stabil. Sementara itu, Cluster 1, Cluster 3, Cluster 4, dan Cluster 6 termasuk kelompok dengan

transaksi dan jumlah produk lebih rendah, sehingga lebih menggambarkan konsumen dengan daya beli terbatas atau perilaku belanja sederhana.

4.3 Pengolahan Data Association Rules Mining

4.3.1 Preprocessing Dataset

Pada tahap ini, peneliti menggunakan *software* VSCode dengan bahasa pemrograman *Python* dan *library mlxtend* untuk melakukan *preprocessing* data transaksi. Tahap ini mencakup langkah-langkah penting dalam mempersiapkan data sebelum dianalisis menggunakan metode AR-MBA (*Association Rules-Market Basket Analysis*). Fungsi *fp_growth* digunakan untuk mengekstraksi *frequent itemset*, sedangkan fungsi *association_Rules* dimanfaatkan untuk membentuk aturan asosiasi beserta metrik evaluasinya. Proses ini terdiri dari beberapa tahap berikut:

4.3.1.1 Data Selection metode Association Rules Mining

Pada tahap ini, dilakukan proses *data selection* dari *database Point of Sale* (POS) Grobbak Korea Pandeyan. Data yang dipilih terbatas pada informasi transaksi penjualan yang relevan, yaitu ID Struk, Nama Produk, dan Platform. Variabel lain yang tidak berkaitan langsung dengan pola asosiasi antar produk, seperti status pembayaran atau nominal total transaksi, diabaikan. Fokus pemilihan variabel ini bertujuan untuk mendapatkan format *basket transaction* yang menampilkan kumpulan item dalam setiap transaksi pelanggan.

Tabel 4. 10 Data Selection untuk proses Association Rules mining

Cluster	ID Struk	Nama Produk
1	4QLLHP478KW86	Corndog Sosis Mozzarella, Corndog Sosis Mozzarella, Topokki Extra Chikuwa, Topokki Extra Chikuwa
	4QLLHP47BJQ88	Corndog Mozzarella, Corndog Mozzarella
	4QLLHP47L7DWN	Corndog Choco
2	4QLLHP474Z9DY	Corndog Choco, Corndog Mozzarella, Corndog Mozzarella, Topokki Original, Topokki Original
	4QLLHP476LKVY	Corndog Mozzarella, Corndog Mozzarella, Odeng Saus Korea, Odeng Saus Korea Extra Odeng 2 Tusuk , Rabokki Extra Mozzarella, Rabokki Extra Mozzarella

Cluster	ID Struk	Nama Produk
	4QLLHP478KW99	Corndog Mozzarella, Corndog Mozzarella, Corndog Sosis Mozzarella, Corndog Sosis Mozzarella
3	4QLLHP47JYN7Z	Corndog Sosis, Corndog Sosis Mozzarella, Corndog Sosis Mozzarella, Corndog Sosis
	4QLLHP47LRV78	Corndog Mozzarella, Corndog Mozzarella, Odeng Saus Korea, Odeng Saus Korea
	4QLLHP47VBPWK	Bukokki, Bukokki Extra Odeng Potong 8 pcs, Corndog Choco, Corndog Mozzarella, Corndog Mozzarella , Hot Spicy Bibimmyeon, Hot Spicy Bibimmyeon Extra Odeng Potong 8 pcs
4	4QLLHP474K4N9	Corndog Choco
	4QLLHP474Z6D7	Corndog Choco, Hot Spicy Bibimmyeon, Hot Spicy Bibimmyeon
	4QLLHP474Z97L	Corndog Choco, Odeng Fish Soup, Topokki Extra Odeng, Topokki Extra Odeng
5	4QLLHP476LKWM	Corndog Matcha, Corndog Tiramisu, Odeng Fish Soup, Odeng Fish Soup
	4QLLHP47V8GMB	Corndog Sosis, Corndog Sosis, Topokki Original, Topokki Original
	4QLLHP47PLQJL	Bukokki, Bukokki , Odeng Saus Korea
6	4874LLT585D9Q9DJJ	Corndog Chocolate, Corndog Mozzarella, Corndog Sosis Mozzarella
	4B7XLLHBQBPJRRRD4	Rabokki Extra Mozzarella
	4D7J44HG7G866BXBB	Chikuwa Saus Korea, Topokki Extra Chikuwa

4.3.1.2 Data Cleaning metode Association Rules Mining

Pada tahap ini dilakukan proses *data cleaning* untuk memastikan bahwa data transaksi yang digunakan dalam analisis aturan asosiasi benar-benar merepresentasikan pola pembelian konsumen yang valid. Salah satu permasalahan utama yang ditemukan dalam dataset adalah adanya transaksi dengan hanya satu jenis produk yang dibeli. Transaksi semacam ini tidak dapat

memberikan informasi mengenai keterkaitan antar produk, sehingga dikeluarkan dari dataset. Berikut merupakan data yang dihapus:

Tabel 4. 11 Tampilan ID Struk yang hanya memiliki 1 baranag

ID Struk	Nama Produk	Cluster
4B7XLLHBQBPJRRRD4	Rabokki Extra Mozzarella	6
4M7XLLT46BZYZRMPP	Corndog Mozzarella	6
4QLLHP474K4N9	Corndog Choco	4
4QLLHP477ZL4W	Corndog Choco	6
4QLLHP47L7DWN	Corndog Choco	1
4QLLHP47PK996	Corndog Choco	2
4QLLHP47QDNYM	Corndog Matcha	3
4QLLHP47WQD8V	Corndog Matcha	3
4QLLHP47ZGWVZ	Corndog Taro	1

Tabel 4. 12 Pembersihan Data

Cluster	Baris total	Baris dibuang	Baris tersisa
1	1296	181	1115
2	1451	173	1278
3	1518	217	1301
4	2195	287	1908
5	291	40	251
6	904	133	771

4.3.1.3 Data Transformastion metode Association Rules Mining

Pada tahap ini, data transaksi yang semula berbentuk daftar produk dalam satu struk diubah menjadi format *basket transaction* dengan representasi *boolean* yang mana setiap produk ditandai dengan nilai *TRUE* apabila muncul dalam suatu transaksi, dan *FALSE* apabila tidak muncul. Berikut beberapa contoh dari hasil dari transformasi data ke bentuk *boolean*:

Tabel 4. 13 Data Transformastion menjadi Boolean

No.	Chikuwa Saus Korea	Chikuwa Saus Korea Extra Chikuwa 8 Potong	Corndog Choco	Corndog Chocolate	...	Toppoki Original Extra 4pcs
1.	<i>FALSE</i>	<i>FALSE</i>	<i>TRUE</i>	<i>FALSE</i>	...	<i>FALSE</i>
2.	<i>FALSE</i>	<i>FALSE</i>	<i>FALSE</i>	<i>FALSE</i>	...	<i>FALSE</i>
3.	<i>FALSE</i>	<i>FALSE</i>	<i>FALSE</i>	<i>FALSE</i>	...	<i>FALSE</i>
4.	<i>FALSE</i>	<i>FALSE</i>	<i>FALSE</i>	<i>FALSE</i>	...	<i>FALSE</i>
5.	<i>FALSE</i>	<i>FALSE</i>	<i>FALSE</i>	<i>FALSE</i>	...	<i>FALSE</i>
6.	<i>FALSE</i>	<i>FALSE</i>	<i>FALSE</i>	<i>FALSE</i>	...	<i>FALSE</i>

No.	Chikuwa Saus Korea	Chikuwa Saus Korea Extra Chikuwa 8 Potong	Corndog Choco	Corndog Chocolate	...	Toppoki Original Extra 4pcs
...	...		...	...	...	...
7655	FALSE	FALSE	FALSE	FALSE	...	FALSE

4.3.2 Proses Association Rules Mining dengan Algoritma FP-Growth

Setelah data melalui tahap *selection*, *cleaning*, dan *transformation*, langkah berikutnya adalah menerapkan metode *Association Rules Mining* menggunakan algoritma FP-Growth untuk membentuk aturan asosiasi berdasarkan data transaksi yang telah ditransformasi sebelumnya. Proses analisis dilakukan melalui *software* VSCode dengan bantuan *library* *mlxtend* pada bahasa pemrograman Python.

Sebelum nilai *minimum support* tersebut diputuskan, terlebih dahulu dilakukan perbandingan pada beberapa tingkat *support* di setiap klaster. Langkah ini bertujuan untuk melihat bagaimana perubahan nilai ambang memengaruhi jumlah *frequent itemset* maupun banyaknya aturan asosiasi yang dihasilkan. Berikut ditampilkan perbandingan *minimum support* dari masing-masing klaster:

Tabel 4. 14 perbandingan minimum support

Cluster	MinSupport	MinConfidence	Jumlah Itemset	Jumlah Aturan
1	0,01	0,4	75	8
	0,02	0,4	42	2
	0,03	0,4	29	2
	0,04	0,4	20	1
	0,05	0,4	16	1
2	0,01	0,4	70	4
	0,02	0,4	38	1
	0,03	0,4	27	1
	0,04	0,4	19	1
	0,05	0,4	17	1
3	0,01	0,4	71	9
	0,02	0,4	39	2
	0,03	0,4	27	1
	0,04	0,4	19	1
	0,05	0,4	16	1
4	0,01	0,4	72	3
	0,02	0,4	41	2
	0,03	0,4	27	0

Cluster	MinSupport	MinConfidence	Jumlah Itemset	Jumlah Aturan
	0,04	0,4	22	0
	0,05	0,4	18	0
5	0,01	0,4	87	32
	0,02	0,4	45	9
	0,03	0,4	27	3
	0,04	0,4	22	2
	0,05	0,4	17	0
6	0,01	0,4	65	5
	0,02	0,4	38	2
	0,03	0,4	26	1
	0,04	0,4	19	1
	0,05	0,4	17	1

Dari tabel terlihat bahwa setiap klaster (1–6) diuji dengan berbagai nilai *minimum support* mulai dari 0,01 hingga 0,05 dengan *minimum confidence* tetap sebesar 0,4. Pemilihan nilai *minimum confidence* sebesar 0,4 (40 %) dalam penelitian ini mempertimbangkan secara teknis keandalan aturan (*Rules reliability*) sekaligus efisiensi analisis, serta didukung oleh praktik terkini dalam literatur ilmiah. Sebagai indikator probabilitas bersyarat bahwa *consequent* terjadi bila *antecedent* muncul, *confidence* yang terlalu rendah ($< 0,4$) menghasilkan terlalu banyak aturan asosiasi yang lemah secara relevansi, sedangkan ambang terlalu tinggi bisa ($< 0,4$) memicu penghilangan pola penting (*miss relevant patterns*). Studi terbaru oleh Hunyadi et al. (2025) menunjukkan bahwa penggunaan ambang *confidence* yang fleksibel diatur bersama ambang *support* yang realistis seperti antara 0,10 hingga 1,00 meningkatkan kemampuan menemukan aturan yang dapat diandalkan tanpa menimbulkan beban aturan yang berlebihan.

Pada klaster 1, 2, 3, 4, dan 6 digunakan *minimum support* sebesar 0,01 karena pada nilai ini aturan yang muncul masih cukup banyak dan beragam sehingga dapat merepresentasikan perilaku pembelian konsumen tanpa kehilangan informasi penting. Sebaliknya, pada klaster 5 penggunaan *minimum support* 0,01 menghasilkan 32 aturan dari 87 *itemset* yang dinilai terlalu banyak dan berpotensi menimbulkan kesulitan dalam interpretasi. Oleh karena itu, nilai *minimum support* pada klaster 5 dinaikkan menjadi 0,02 sehingga jumlah aturan yang terbentuk berkurang menjadi 9 aturan dari 45 *itemset* yang lebih proporsional dan tetap mampu menggambarkan pola pembelian konsumen secara representatif. Dengan demikian, pemilihan nilai *minimum support* ini memastikan bahwa aturan asosiasi yang terbentuk tidak terlalu

sedikit sehingga kehilangan makna, namun juga tidak terlalu banyak sehingga sulit dianalisis, melainkan tetap fokus dan relevan untuk mendukung strategi pemasaran.

Tabel 4. 15 Aturan Asosiasi klaster 1

Antecedent	Consequent	Support	Confidence	Lift
Odeng Fish Soup Extra Odeng 2 Tusuk	Odeng Fish Soup	0,018421053	1	13,10344828
Topokki Extra Odeng	Corndog Sosis Mozzarella	0,018421	0,518518519	1,417533
Corndog Tiramisu	Corndog Choco	0,019737	0,416666667	1,41369
Corndog Matcha	Corndog Choco	0,039474	0,4	1,357143
Topokki Extra Mozzarella	Corndog Mozzarella	0,085526	0,532786885	1,314669
Pakokki	Corndog Sosis Mozzarella	0,030263	0,479166667	1,309952
Mandu Saus Gochujang	Corndog Sosis Mozzarella	0,017105	0,40625	1,110612
Mandu Saus Gochujang	Corndog Mozzarella	0,018421	0,4375	1,079545
Hot Spicy Bibimmyeon	Corndog Sosis Mozzarella	0,027632	0,388888889	1,063149
Corndog Sosis	Corndog Sosis Mozzarella	0,022368	0,386363636	1,056246

Pada klaster 1 diperoleh sejumlah aturan asosiasi yang menunjukkan keterkaitan antar produk dengan tingkat kekuatan yang bervariasi. Aturan paling dominan terlihat pada kombinasi Odeng Fish Soup Extra Odeng 2 Tusuk dan Odeng Fish Soup yang memiliki nilai *support* sebesar 0,0126, *confidence* 100%, serta *lift* 17,698, sehingga dapat dipastikan bahwa setiap pembelian varian tambahan selalu disertai dengan menu utama odeng. Selain itu, pola lain yang cukup kuat adalah Topokki Extra Odeng dan Corndog Sosis Mozzarella dengan *support* 0,0152, *confidence* 51,85%, dan *lift* 1,41, serta Topokki Extra Mozzarella dan Corndog Mozzarella dengan *support* 0,0515, *confidence* 53,28%, dan *lift* 1,31. Pada kategori corndog rasa manis, ditemukan asosiasi Corndog Tiramisu dan Corndog Choco dengan *support* 0,0131, *confidence* 41,67%, dan *lift* 1,41, serta Corndog Matcha dan Corndog Choco dengan *support* 0,0246, *confidence* 40%, dan *lift* 1,35. Selanjutnya, aturan Rabokki Extra Mozzarella dan Corndog Choco dengan *support* 0,0131, *confidence* 39,53%, dan *lift* 1,84 juga mengindikasikan bahwa konsumen yang memilih menu rabokki varian keju cenderung menambahkan corndog rasa coklat. Di sisi lain, beberapa asosiasi lain menunjukkan hubungan yang moderat, seperti Mandu Saus Gochujang dan Corndog Sosis Mozzarella dengan *support* 0,0117, *confidence*

40,63%, dan *lift* 1,11, serta Hot Spicy Bibimmyeon dan Corndog Sosis Mozzarella dengan *support* 0,0159, *confidence* 38,89%, dan *lift* 1,06. Selain itu, terdapat pula aturan Corndog Sosis dan Corndog Sosis Mozzarella dengan *support* 0,0192, *confidence* 33,33%, dan *lift* 1,03, yang menunjukkan bahwa konsumen yang membeli corndog sosis seringkali melengkapinya dengan varian mozzarella.

Konsumen pada klaster 1 menunjukkan kecenderungan yang konsisten dalam mengombinasikan menu utama dengan produk pelengkap. Produk *corndog* muncul sebagai item sentral karena sering dipasangkan dengan berbagai menu lain, mulai dari *topokki*, *bibimmyeon pedas*, hingga *mandu*. Selain itu, varian corndog rasa manis seperti *choco*, *matcha*, dan *tiramisu* cenderung dibeli bersamaan, yang menegaskan adanya preferensi konsumen terhadap kombinasi *sweet menu*. T

Tabel 4. 16 Aturan Asosiasi Klaster 2

Antecedent	Consequent	Support	Confidence	Lift
Odeng Fish Soup Extra Odeng 2 Tusuk	Odeng Fish Soup	0,014238773	1	12,0131 6
Odeng Saus Korea Extra Odeng 2 Tusuk	Odeng Saus Korea	0,014238773	1	7,87069
Rabokki Extra Mozzarella	Corndog Mozzarella	0,019715225	0,409090909	1,05508 5
Topokki Extra Mozzarella	Corndog Mozzarella	0,050383352	0,4	1,03163 8

Pada klaster ini, aturan asosiasi yang terbentuk menunjukkan pola pembelian yang cukup jelas. Produk *Odeng Fish Soup Extra Odeng 2 Tusuk* selalu dibeli bersamaan dengan menu utama *Odeng Fish Soup* dengan nilai *support* sebesar 0,0142, *confidence* mencapai 100%, serta *lift* 12,01, yang menegaskan bahwa varian tambahan ini sepenuhnya bergantung pada menu utama. Pola serupa juga terlihat pada *Odeng Saus Korea Extra Odeng 2 Tusuk* yang selalu dibeli bersama *Odeng Saus Korea* dengan *confidence* 100% dan *lift* 7,87. Hal ini mengindikasikan bahwa varian *extra tusuk* pada produk odeng berfungsi sebagai strategi *upselling* dan tidak pernah dipilih secara terpisah. Selain itu, ditemukan pula keterkaitan antara *Rabokki Extra Mozzarella* dengan *Corndog Mozzarella* yang memiliki *support* 0,0197, *confidence* 40,91%, dan *lift* 1,05, serta kombinasi *Topokki Extra Mozzarella* dengan *Corndog Mozzarella* yang memiliki *support* 0,0508, *confidence* 40%, dan *lift* 1,03. Kedua aturan terakhir ini mengindikasikan adanya kecenderungan konsumen untuk mengombinasikan menu berbasis keju, meskipun kekuatan asosiasinya relatif moderat. Dengan demikian, dapat disimpulkan

bahwa varian *extra* pada menu odeng berperan sebagai produk pelengkap yang secara konsisten dibeli bersama menu utamanya, sementara kombinasi antara topokki atau rabokki dengan corndog mozzarella dapat dimanfaatkan sebagai peluang *cross-selling* untuk meningkatkan nilai transaksi rata-rata.

Tabel 4. 17 Aturan Asosiasi Klaster 3

Antecedent	Consequent	Support	Confidence	Lift
Odeng Fish Soup Extra Odeng 2 Tusuk	Odeng Fish Soup	0,016465423	1	10,8452381
Rabokki Extra Mozzarella	Corndog Choco	0,018661	0,459459	1,500242
Chikuwa Saus Korea	Corndog Sosis Mozzarella	0,01427	0,481481	1,357986
Topokki Extra Mozzarella	Corndog Mozzarella	0,073546	0,51938	1,332831
Corndog Choco, Topokki Extra Mozzarella	Corndog Mozzarella	0,012075	0,5	1,283099
Topokki Extra Odeng	Corndog Sosis Mozzarella	0,015368	0,451613	1,273744
Pakokki	Corndog Sosis Mozzarella	0,020856	0,431818	1,217914
Rabokki Extra Mozzarella	Corndog Mozzarella	0,017563	0,432432	1,109707
Corndog Blueberry	Corndog Mozzarella	0,012075	0,423077	1,085699

Pada klaster ini, aturan asosiasi yang terbentuk menunjukkan adanya pola pembelian yang kuat antara varian tambahan dan menu utama, serta keterkaitan moderat antar produk pendamping. Produk Odeng Fish Soup Extra Odeng 2 Tusuk selalu dibeli bersamaan dengan menu utama Odeng Fish Soup dengan nilai *support* sebesar 0,0165, *confidence* 100%, dan *lift* 10,85, yang menegaskan bahwa varian tambahan ini sepenuhnya bergantung pada menu utama. Selain itu, Rabokki Extra Mozzarella sering dibeli bersamaan dengan Corndog Choco dengan *support* 0,0187, *confidence* 45,95%, dan *lift* 1,50, sedangkan Chikuwa Saus Korea memiliki keterkaitan dengan Corndog Sosis Mozzarella dengan *support* 0,0143, *confidence* 48,15%, dan *lift* 1,36. Pola lain juga terlihat pada Topokki Extra Mozzarella yang kerap dipadukan dengan Corndog Mozzarella dengan *support* 0,0735, *confidence* 51,94%, dan *lift* 1,33, serta kombinasi Corndog Choco dan Topokki Extra Mozzarella dengan Corndog Mozzarella dengan *support* 0,0121, *confidence* 50%, dan *lift* 1,28. Aturan lain yang juga muncul yaitu Topokki Extra Odeng dengan Corndog Sosis Mozzarella dengan *support* 0,0154, *confidence* 45,16%, dan *lift* 1,27, Pakokki dengan Corndog Sosis Mozzarella dengan *support* 0,0209, *confidence* 43,18%, dan *lift*

1,22, *Rabokki Extra Mozzarella* dengan *Corndog Sosis Mozzarella* dengan *support* 0,0176, *confidence* 43,24%, dan *lift* 1,11, serta *Corndog Blueberry* dengan *Corndog Mozzarella* dengan *support* 0,0121, *confidence* 42,31%, dan *lift* 1,09. Dengan demikian, dapat disimpulkan bahwa pada klaster ini konsumen cenderung membeli varian tambahan odeng selalu bersama menu utamanya, serta kerap mengombinasikan berbagai menu utama seperti topokki, rabokki, pakokki, maupun chikuwa dengan corndog, khususnya varian mozzarella.

Tabel 4. 18 Aturan Asosiasi Klaster 4

Antecedent	Consequent	Support	Confidence	Lift
Odeng Fish Soup Extra Odeng 2 Tusuk	Odeng Fish Soup	0,026706	1	9,628571
Odeng Saus Korea Extra Odeng 2 Tusuk	Odeng Saus Korea	0,011128	1	8,425
Topokki Extra Odeng Mozzarella	Corndog Mozzarella	0,020772	0,49122807	1,308647

Pada klaster 4, aturan asosiasi yang terbentuk memperlihatkan hubungan yang sangat kuat antara produk utama dan varian tambahannya. Produk Odeng Fish Soup Extra Odeng 2 Tusuk selalu dibeli bersamaan dengan *Odeng Fish Soup* dengan nilai *support* sebesar 0,0267, *confidence* 100%, dan *lift* 9,63, yang menegaskan bahwa varian tambahan ini sepenuhnya bergantung pada menu utama. Pola serupa juga terlihat pada Odeng Saus Korea Extra Odeng 2 Tusuk yang selalu dibeli bersamaan dengan Odeng Saus Korea dengan *support* 0,0111, *confidence* 100%, dan *lift* 8,43. Selain itu, terdapat pula keterkaitan pada kombinasi Topokki Extra Odeng Mozzarella dengan Corndog Mozzarella dengan *support* 0,0208, *confidence* 49,12%, dan *lift* 1,31. Dengan demikian, dapat disimpulkan bahwa pada klaster ini konsumen secara konsisten membeli varian *extra* odeng bersama menu utamanya, serta memiliki kecenderungan untuk mengombinasikan topokki dengan corndog mozzarella.

Tabel 4. 19 Aturan Asosiasi Klaster 5

Antecedent	Consequent	Support	Confidence	Lift
Rabokki	Odeng Fish Soup	0,02381	0,4	3,536842
Corndog Strawberry	Corndog Choco	0,02381	0,571429	1,882353
Corndog Matcha	Corndog Choco	0,041667	0,5	1,647059
Bukokki	Corndog Sosis Mozzarella	0,02381	0,444444	1,224044
Rabokki	Corndog Sosis Mozzarella	0,02381	0,4	1,101639

Antecedent	Consequent	Support	Confidence	Lift
Hot Spicy Bibimmyeon	Corndog Sosis Mozzarella	0,02381	0,4	1,101639
Topokki Original	Corndog Mozzarella	0,041667	0,411765	1,048128
Rabokki	Corndog Mozzarella	0,02381	0,4	1,018182

Pada klaster 5, aturan asosiasi yang terbentuk memperlihatkan pola pembelian yang cukup konsisten antara varian tambahan dan menu utama, serta keterkaitan moderat pada kombinasi corndog dengan produk lainnya. Produk *Odeng Fish Soup Extra Odeng 2 Tusuk* selalu dibeli bersamaan dengan *Odeng Fish Soup* dengan nilai *support* 0,0357, *confidence* 100%, dan *lift* 8,84, yang menegaskan bahwa varian tambahan ini sepenuhnya bergantung pada menu utama. Selain itu, *Rabokki* juga memiliki keterkaitan dengan *Odeng Fish Soup* dengan *support* 0,0238, *confidence* 40%, dan *lift* 3,54, yang menunjukkan kecenderungan konsumen untuk mengombinasikan kedua menu tersebut. Pada kategori corndog, aturan yang menonjol adalah *Corndog Strawberry* yang sering dibeli bersama *Corndog Choco* dengan *support* 0,0238, *confidence* 57,14%, dan *lift* 1,88, serta *Corndog Matcha* dengan *Corndog Choco* dengan *support* 0,0417, *confidence* 50%, dan *lift* 1,65, yang mengindikasikan konsumen cenderung memilih lebih dari satu varian corndog manis dalam satu transaksi. Selain itu, *Bukokki* memiliki keterkaitan dengan *Corndog Sosis Mozzarella* dengan *support* 0,0238, *confidence* 44,44%, dan *lift* 1,22, sementara *Rabokki* maupun *Hot Spicy Bibimmyeon* juga berhubungan dengan *Corndog Sosis Mozzarella* dengan *confidence* sama-sama 40% meskipun nilai *lift* relatif rendah di kisaran 1,10. Pola lain yang muncul adalah *Topokki Original* dengan *Corndog Mozzarella* dengan *support* 0,0417, *confidence* 41,18%, dan *lift* 1,05, serta *Rabokki* dengan *Corndog Mozzarella* dengan *support* 0,0238, *confidence* 40%, dan *lift* 1,02. Dengan demikian, dapat disimpulkan bahwa pada klaster ini konsumen cenderung membeli varian tambahan odeng selalu bersama menu utamanya, serta memiliki preferensi untuk mengombinasikan berbagai varian corndog, khususnya corndog manis, dalam satu pesanan. Corndog juga tetap berperan sebagai produk pendamping utama bagi menu rabokki, bukokki, maupun bibimmyeon.

Tabel 4. 20 Aturan Asosiasi Klaster 6

Antecedent	Consequent	Support	Confidence	Lift
Corndog Tiramisu	Corndog Choco	0,011215	0,428571	1,442049
Topokki Extra Odeng Mozzarella	Corndog Mozzarella	0,013084	0,538462	1,333689
Topokki Extra Mozzarella	Corndog Mozzarella	0,056075	0,422535	1,046557

Antecedent	Consequent	Support	Confidence	Lift
Chikuwa Saus Korea	Corndog Mozzarella	0,020561	0,407407	1,009088

Pada klaster 6, aturan asosiasi yang terbentuk menunjukkan pola keterkaitan yang konsisten antara produk utama dan varian tambahannya, serta hubungan moderat antarvarian corndog dan topokki. Produk *Odeng Fish Soup Extra Odeng 2 Tusuk* selalu dibeli bersamaan dengan *Odeng Fish Soup* dengan nilai *support* 0,0187, *confidence* 100%, dan *lift* 10,29, yang menegaskan bahwa varian tambahan ini sepenuhnya bergantung pada menu utama. Pada kategori corndog, ditemukan keterkaitan *Corndog Tiramisu* dengan *Corndog Choco* dengan *support* 0,0112, *confidence* 42,86%, dan *lift* 1,44, yang menunjukkan adanya kecenderungan konsumen untuk memilih lebih dari satu varian corndog manis dalam satu transaksi. Selain itu, *Topokki Extra Odeng* juga memiliki hubungan dengan *Corndog Mozzarella* dengan *support* 0,0131, *confidence* 53,85%, dan *lift* 1,33, serta *Topokki Extra Mozzarella* dengan *Corndog Mozzarella* dengan *support* 0,0561, *confidence* 42,25%, dan *lift* 1,05, yang menandakan adanya kecenderungan konsumen memadukan topokki berbasis keju maupun odeng dengan corndog keju. Aturan lain yang muncul adalah *Chikuwa Saus Korea* dengan *Corndog Mozzarella* yang memiliki *support* 0,0206, *confidence* 40,74%, dan *lift* 1,01, meskipun hubungan ini relatif lemah. Dengan demikian, dapat disimpulkan bahwa pada klaster ini konsumen selalu membeli varian tambahan odeng bersama menu utamanya, serta cenderung mengombinasikan topokki maupun chikuwa dengan corndog, terutama varian mozzarella,

BAB V

PEMBAHASAN

5.1 Analisis Penerapan Metode *Clustering*

5.1.1 Analisis Deskripsi Data

Pada tahap ini dilakukan analisis deskripsi data sebagai dasar untuk memahami hasil pengolahan yang telah dilakukan sebelumnya. Data yang telah melalui proses *cleaning*, transformasi, seleksi, dan normalisasi kemudian dianalisis lebih lanjut untuk melihat kecenderungan pada setiap variabel. Penyajian hasil deskripsi ini memberikan gambaran mengenai perbedaan karakteristik antar klaster yang terbentuk, baik dari jumlah produk yang dibeli, nilai transaksi, maupun pola pembelian konsumen berdasarkan platform dan waktu transaksi.

Analisis deskriptif ini tidak hanya berfungsi sebagai pelengkap, tetapi juga memberikan landasan awal dalam menginterpretasikan perilaku konsumen pada tiap klaster. Dengan cara ini, hasil *clustering* dapat dipahami secara lebih komprehensif, sekaligus mempermudah dalam menarik kesimpulan mengenai strategi pemasaran yang relevan sesuai dengan karakteristik tiap kelompok konsumen. Berikut data deskripsi pada jumlah produk dan total transaksi:

Tabel 5. 1 Deskripsi data jumlah produk dan transaksi

Cluster	Jumlah	Jumlah produk			Total Transaksi		
		Mean	Min	Max	Mean	Min	Max
1	1296	3,025	1	6	Rp 36.234,49	Rp 13.900,00	Rp 74.700,00
2	1451	6,195	3	10	Rp 63.642,71	Rp 33.000,00	Rp 111.600,00
3	1518	2,733	1	6	Rp 21.724,97	Rp 9.000,00	Rp 56.000,00
4	2195	3,118	1	6	Rp 36.771,11	Rp 11.500,00	Rp 70.300,00
5	291	10,306	6	30	Rp 113.716,20	Rp 70.000,00	Rp 347.900,00
6	904	2,822	1	5	Rp 38.266,26	Rp 11.500,00	Rp 82.500,00

Berdasarkan hasil pengelompokan, terlihat bahwa masing-masing klaster memiliki karakteristik yang berbeda. Klaster 1 dengan jumlah 1.296 transaksi menunjukkan pola pembelian sederhana, yaitu rata-rata tiga produk per struk dengan rentang satu hingga enam produk, serta nilai transaksi menengah sekitar Rp36.234 dengan batas bawah Rp13.900 dan tertinggi Rp74.700. Klaster 2 yang terdiri atas 1.451 transaksi menggambarkan konsumen dengan daya beli lebih tinggi, ditunjukkan dengan rata-rata enam produk per transaksi dengan

rentang tiga hingga sepuluh produk, serta nilai transaksi rata-rata Rp63.642 yang dapat mencapai hingga Rp111.600. Berbeda dengan itu, Klaster 3 yang memiliki 1.518 transaksi justru merupakan kelompok dengan nilai pembelian terendah, karena rata-rata hanya dua hingga tiga produk per transaksi dengan total belanja sekitar Rp21.725, bahkan ada yang hanya Rp9.000. Klaster 4, yang jumlahnya paling besar yakni 2.195 transaksi, memperlihatkan perilaku yang hampir serupa dengan Klaster 1, yaitu rata-rata tiga produk dengan nilai transaksi Rp36.771, meskipun variasinya lebih luas dari Rp11.500 hingga Rp70.300. Sementara itu, Klaster 5 menjadi segmen premium meskipun hanya terdiri atas 291 transaksi, karena rata-rata pembelian mencapai sepuluh produk per struk dengan nilai transaksi jauh lebih tinggi yaitu Rp113.716, dan bahkan dapat mencapai Rp347.900. Adapun Klaster 6 dengan 904 transaksi cenderung berada pada kategori menengah, dengan rata-rata tiga produk per transaksi dan nilai belanja Rp38.266, sedikit lebih tinggi dibandingkan Klaster 1 dan Klaster 4. Temuan ini dapat disimpulkan bahwa Klaster 3 adalah kelompok dengan nilai transaksi terendah, Klaster 5 merupakan konsumen premium dengan kontribusi terbesar terhadap pendapatan, Klaster 2 menempati posisi menengah ke atas dengan pola belanja konsisten, sedangkan Klaster 1, 4, dan 6 menggambarkan konsumen dengan karakteristik belanja moderat yang artinya jumlah produk yang dibeli dan nilai transaksinya tidak terlalu rendah tapi juga tidak setinggi klaster premium.

5.1.2 Analisis Proporsi Anggota tiap Cluster

Selanjutnya menganalisis distribusi jumlah anggota pada setiap klaster untuk mengetahui seberapa besar dominasi masing-masing kelompok konsumen dalam keseluruhan data transaksi. Dengan memahami sebaran jumlah anggota tiap klaster, dapat diperoleh gambaran mengenai segmen yang paling banyak merepresentasikan perilaku pembelian konsumen, serta segmen yang relatif kecil namun mungkin memiliki karakteristik khusus yang signifikan.

Tabel 5. 2 Proporsi jumlah data tiap klaster

Cluster	Jumlah Data	Persentase (%)
1	1.296	16,93%
2	1.451	18,95%
3	1.518	19,83%
4	2.195	28,67%
5	291	3,80%
6	904	11,81%

Dari hasil pengelompokan diperoleh enam klaster dengan total 7.655 data. Klaster 4 menjadi yang terbesar dengan 2.195 data (28,67%) dan dapat dianggap sebagai segmen utama. Klaster 3 berjumlah 1.518 data (19,83%) serta Klaster 2 dengan 1.451 data (18,95%) berperan sebagai segmen pendukung utama. Klaster 1 memiliki 1.296 data (16,93%) yang masih cukup signifikan. Sementara itu, Klaster 6 berjumlah 904 data (11,81%) yang meskipun lebih kecil tetap perlu diperhatikan. Adapun Klaster 5 hanya terdiri dari 291 data (3,80%) sehingga menjadi segmen paling kecil

Tabel 5. 3 Proporsi jumlah data tiap klaster pada platform

Cluster	GoFood	Grab	Kios	Shopee
1	0%	0%	0%	100%
2	50%	17%	13%	20%
3	0%	0%	100%	0%
4	100%	0%	0%	0%
5	56%	18%	4%	22%
6	0%	100%	0%	0%

Klaster 1 sepenuhnya didominasi oleh platform Shopee (100%), sehingga segmen ini bisa disebut sebagai pelanggan yang hanya bertransaksi lewat Shopee. Klaster 2 memiliki distribusi lebih beragam, dengan mayoritas berasal dari GoFood (50%), diikuti Shopee (20%), Grab (17%), dan kios (13%), sehingga dapat digolongkan sebagai segmen multikanal. Klaster 3 hanya menggunakan kios (100%), menandakan konsumen yang selalu melakukan pembelian langsung. Klaster 4 seluruhnya melalui GoFood (100%), menunjukkan loyalitas yang kuat terhadap platform tersebut. Klaster 5 cukup tersebar dengan dominasi GoFood (56%), disusul Shopee (22%), Grab (18%), dan kios (4%), sehingga lebih fleksibel dalam penggunaan platform. Terakhir, Klaster 6 sepenuhnya menggunakan Grab (100%), menjadikannya segmen yang sangat spesifik pada satu platform.

Tabel 5. 4 Proporsi jumlah data tiap klaster pada waktu

Cluster	Pagi	Siang	Sore	Malam
1	1%	5%	6%	88%
2	7%	27%	28%	39%
3	2%	6%	4%	89%
4	1%	2%	3%	95%
5	0%	0%	1%	98%
6	2%	5%	2%	91%

Berdasarkan tabel distribusi waktu transaksi pada tiap klaster, terlihat bahwa mayoritas aktivitas pembelian terjadi pada malam hari di hampir semua klaster. Klaster 1, 3, 4, 5, dan 6 memiliki dominasi transaksi yang sangat tinggi pada malam hari, masing-masing mencapai 88%, 89%, 95%, 98%, dan 91%, menandakan bahwa segmen pelanggan dalam kelompok ini cenderung melakukan pembelian pada waktu malam, kemungkinan karena bertepatan dengan jam santai atau setelah aktivitas utama harian selesai.

Sementara itu, Klaster 2 menunjukkan pola waktu yang lebih merata, dengan distribusi transaksi 7% pada pagi hari, 27% siang, 28% sore, dan 39% malam, menandakan segmen pelanggan ini lebih aktif sepanjang hari, tidak hanya terpusat pada malam hari. Hal ini mengindikasikan bahwa klaster 2 berpotensi merepresentasikan pelanggan dengan kebiasaan konsumsi lebih fleksibel, misalnya pekerja atau mahasiswa dengan waktu luang yang bervariasi.

5.1.3 Analisis Karakteristik Kluster

Setelah diketahui distribusi jumlah anggota, platform pemesanan, serta kecenderungan waktu transaksi pada masing-masing klaster, tahap selanjutnya adalah melakukan analisis lebih mendalam terhadap karakteristik setiap klaster. Analisis ini bertujuan untuk menggambarkan profil unik dari tiap kelompok konsumen berdasarkan perilaku pembelian yang meliputi jumlah produk yang dibeli, total nilai transaksi, platform yang digunakan, hingga periode waktu dominan saat melakukan pembelian. Dengan memahami karakteristik klaster secara lebih rinci, strategi pemasaran dapat disusun secara spesifik sesuai kebutuhan dan pola belanja konsumen pada tiap segmen, sehingga upaya peningkatan penjualan menjadi lebih tepat sasaran dan efektif.

Tabel 5. 5 Pesebaran data tiap klaster pada jumlah produk dan total transaksi

cluster		1	2	3	4	5	6
Jumlah Produk	min	1	3	1	1	6	1
	25%	2	5	2	2	8	2
	50%	3	6	2	3	10	3
	75%	4	7	4	4	11	4
	max	6	10	6	6	30	5
Total	min	13900	33000	9000	11500	70000	11500
	25%	24900	55000	14000	32200	95300	25500
	50%	38400	61900	22000	36900	105200	39800
	75%	43600	71500	28000	42275	122650	45200
	max	74700	111600	56000	70300	347900	82500

Tabel 5. 6 Pesebaran data tiap klaster pada waktu

Cluster	Pagi	Siang	Sore	Malam
1	20	74	84	1274
2	108	401	417	588
3	19	72	49	1156
4	2	6	8	275
5	0	9	29	2159
6	17	44	22	822

Tabel 5. 7 Pesebaran data tiap klaster pada platfrom

Cluster	GoFood	Kios	Shopee	Grab
1	0	0	1296	0
2	727	187	290	247
3	4	1513	0	1
4	2195	0	0	0
5	162	12	64	53
6	0	0	0	904

Berdasarkan data yang sudah dipaparkan berikut analisis untuk masing-masing klaster:

➤ **Klaster 1**

- (a) Konsumen pada klaster ini rata-rata membeli 3 produk per transaksi dengan kisaran 1–6 produk. Nilai transaksi berkisar Rp13.900–Rp74.700, dengan rata-rata Rp36.234. Artinya, segmen ini lebih condong pada keranjang sederhana dengan pengeluaran menengah.

- (b) Sebanyak 88% transaksi dilakukan pada malam hari. Ini menunjukkan bahwa konsumen klaster ini lebih memilih berbelanja setelah aktivitas utama selesai, kemungkinan sebagai konsumsi malam atau camilan larut malam.
- (c) Seluruh transaksi dilakukan melalui ShopeeFood (100%). Hal ini mengindikasikan loyalitas penuh terhadap satu platform.
- (d) Klaster ini menggambarkan konsumen dengan jumlah pembelian yang tergolong sedang. Waktu transaksi lebih banyak terjadi pada malam hari, dan seluruhnya dilakukan melalui satu *platform* saja yaitu Shoopefood. Pola ini menunjukkan perilaku konsumen yang cukup konsisten dan loyal terhadap layanan Shoopefood.

➤ Klaster 2

- (a) Rata-rata pembelian mencapai 6 produk per transaksi dengan rentang 3–10 produk. Nilai transaksi cukup tinggi, rata-rata Rp63.642 dengan batas atas Rp111.600. Ini menandakan konsumen dalam klaster ini cenderung belanja banyak sekaligus.
- (b) Distribusi waktu lebih merata dibandingkan klaster lain: sore (28%), malam (39%), siang (27%), dan pagi (7%). Artinya, konsumen ini tidak terlalu terikat pada satu waktu tertentu.
- (c) Terdistribusi pada beberapa platform: GoFood (50%), Grab (17%), Kios (13%), dan Shopee (20%). Hal ini memperlihatkan bahwa konsumen klaster ini tidak eksklusif terhadap satu *platform* dan fleksibel dalam memilih media pemesanan.
- (d) Konsumen pada klaster ini cenderung melakukan pembelian dalam jumlah yang lebih besar dibandingkan kelompok lain. Waktu transaksi tetap terkonsentrasi pada malam hari, namun penggunaan platform lebih beragam, tidak terikat pada satu saluran saja. Karakteristik ini memperlihatkan perilaku konsumen yang fleksibel dalam memilih media pemesanan.

➤ Klaster 3

- (a) Rata-rata transaksi hanya 2 produk, dengan nilai pembelian rata-rata Rp22.000. Ini merupakan klaster dengan nilai transaksi paling rendah dibandingkan klaster lain.
- (b) Waktu pembelian masih didominasi oleh malam hari (95%).
- (c) 100% transaksi dilakukan langsung melalui kios. Hal ini menunjukkan segmen pembeli *offline* yang datang langsung ke lokasi.
- (d) Klaster ini didominasi oleh konsumen dengan jumlah produk yang relatif sedikit. Pola waktu transaksi lebih bervariasi karena tidak hanya terpusat pada satu periode

tertentu. Seluruh transaksi dilakukan secara langsung di kios, sehingga klaster ini mencerminkan konsumen dengan kebiasaan belanja langsung ditempat.

➤ Klaster 4

- (a) Rata-rata pembelian sekitar 3 produk per transaksi dengan nilai rata-rata Rp36.771. Nilainya masih dalam kategori menengah, mirip dengan Cluster 1.
- (b) Hampir seluruh transaksi (95%) dilakukan pada malam hari, menunjukkan pola konsumsi yang sangat jelas.
- (c) Seluruh transaksi berasal dari GoFood (100%). Konsumen ini bisa dianggap sebagai pelanggan setia GoFood di malam hari.
- (d) Sebagai klaster dengan proporsi terbesar (28,67%), segmen ini perlu menjadi prioritas utama. Promosi eksklusif berbasis GoFood di malam hari akan efektif. Program loyalitas khusus pelanggan GoFood, seperti poin tambahan atau potongan harga untuk pembelian rutin, dapat menjaga konsistensi transaksi dari segmen ini.

➤ Klaster 5

- (a) Klaster ini memiliki ciri paling mencolok dengan rata-rata pembelian 10 produk per transaksi (rentang 6–30 produk). Total transaksi sangat tinggi, rata-rata Rp105.200 dengan maksimum Rp347.900.
- (b) Mayoritas tetap di malam hari (95%).
- (c) Lebih bervariasi: GoFood (56%), Grab (18%), Shopee (22%), dan kios (4%). Hal ini menandakan konsumen premium ini tidak hanya terbatas pada satu platform, namun fleksibel.
- (d) Walaupun jumlahnya kecil (3,80%), klaster ini sangat penting karena kontribusi pendapatan tinggi. Segmen ini cocok untuk strategi, misalnya paket keluarga, menu *bundling* eksklusif, atau layanan prioritas. Pemberian diskon kecil sudah cukup memotivasi pembelian dalam jumlah besar sehingga dampaknya signifikan terhadap omzet.

➤ Klaster 6

- (a) Rata-rata jumlah produk sekitar 3, dengan nilai transaksi rata-rata Rp39.800, sedikit lebih tinggi dibandingkan Cluster 1 dan 4.
- (b) Sebagian besar terjadi pada malam hari (91%).
- (c) Seluruh transaksi dilakukan melalui Grab (100%). Hal ini menunjukkan adanya loyalitas tunggal terhadap GrabFood.

5.2 Analisis Penerapan Metode *Association rules mining*

5.2.1 Analisis Perbandingan Minimum Support Association Rules Mining

Analisis perbandingan *association rules mining* dilakukan untuk menilai pengaruh variasi nilai *minimum support* dan *minimum confidence* terhadap jumlah *itemset* serta aturan asosiasi yang terbentuk pada masing-masing klaster. Tahap ini penting untuk memahami bagaimana ambang batas parameter memengaruhi keluaran analisis, baik dari sisi jumlah aturan yang dihasilkan maupun kualitas keterkaitan antar produk. Dengan membandingkan hasil dari beberapa nilai ambang yang berbeda, dapat diperoleh keseimbangan antara kelimpahan aturan yang informatif dan kemudahan interpretasi pola pembelian konsumen. Berikut data perbandingan *minimum support* dari setiap klaster:

Tabel 5. 8 Perbandingan Minimum support klaster 1

Clus ter	Trans aksi	MinSup port	MinConfi dence	Jumlah Itemset	Jumlah Rules	Lift Mean	Lift Median	Lift Min	Lift Max	Conf Mean
1	760	0,01	0,4	75	8	3,60	1,36	1,08	13,10	0,60
1	760	0,02	0,4	42	2	1,31	1,31	1,31	1,31	0,51
1	760	0,03	0,4	29	2	1,31	1,31	1,31	1,31	0,51
1	760	0,04	0,4	20	1	1,31	1,31	1,31	1,31	0,53
1	760	0,05	0,4	16	1	1,31	1,31	1,31	1,31	0,53

Berdasarkan hasil pengujian pada Cluster 1 dengan 760 transaksi, terlihat bahwa variasi nilai *minimum support* berpengaruh signifikan terhadap jumlah *itemset* dan aturan yang terbentuk. Pada nilai 0,01 diperoleh 75 *itemset* dengan 8 aturan asosiasi serta *lift* rata-rata 3,60 dan maksimum 13,10, yang menunjukkan kekuatan asosiasi tinggi dan keragaman pola yang masih relevan. Namun, ketika *support* dinaikkan menjadi 0,02 hingga 0,05, jumlah aturan menyusut drastis menjadi hanya 1–2 dengan nilai *lift* stagnan di angka 1,31, sehingga kehilangan kekayaan informasi. Oleh karena itu, penggunaan *minimum support* 0,01 dipilih agar pola keterkaitan produk tetap terungkap secara bermakna, sementara nilai *minimum confidence* ditetapkan pada 0,4 untuk menjaga keseimbangan antara jumlah aturan dan kekuatan asosiasi, sehingga hasil yang diperoleh tetap representatif dalam menggambarkan perilaku pembelian konsumen.

Tabel 5. 9 Perbandingan Minimum support klaster 2

Clus ter	Trans aksi	MinSup port	MinConfi dence	Jumlah Itemset	Jumlah Rules	Lift Mean	Lift Median	Lift Min	Lift Max	Conf Mean
2	913	0,01	0,4	70	4	5,49	4,46	1,03	12,01	0,70

Clus ter	Trans aksi	MinSup port	MinConfi dence	Jumlah Itemset	Jumlah Rules	Lift Mean	Lift Median	Lift Min	Lift Max	Conf Mean
2	913	0,02	0,4	38	1	1,03	1,03	1,03	1,03	0,40
2	913	0,03	0,4	27	1	1,03	1,03	1,03	1,03	0,40
2	913	0,04	0,4	19	1	1,03	1,03	1,03	1,03	0,40
2	913	0,05	0,4	17	1	1,03	1,03	1,03	1,03	0,40

Berdasarkan hasil pengujian pada Cluster 2 dengan 913 transaksi, diperoleh bahwa penggunaan *minimum support* rendah memberikan hasil aturan yang jauh lebih kaya dan kuat dibandingkan ambang yang lebih tinggi. Pada nilai *support* 0,01 terbentuk 70 *itemset* dengan 4 aturan asosiasi, ditunjang oleh *lift* rata-rata 5,49, median 4,46, serta maksimum 12,01, yang menunjukkan adanya hubungan antar produk dengan kekuatan asosiasi yang tinggi dan bermakna. Sebaliknya, ketika *support* dinaikkan ke 0,02 hingga 0,05, aturan yang terbentuk menurun drastis menjadi hanya 1 aturan dengan *lift* yang stagnan di angka 1,03 serta *confidence* rata-rata tetap 0,40, sehingga pola keterkaitan yang terungkap menjadi sangat terbatas. Oleh karena itu, nilai *minimum support* 0,01 dipilih karena mampu memberikan aturan yang lebih representatif dan informatif, sementara penetapan *minimum confidence* 0,4 tetap dipertahankan untuk menjaga keseimbangan antara jumlah aturan dan kekuatan asosiasi yang terbentuk.

Tabel 5. 10 Perbandingan Minimum support klaster 3

Clus ter	Trans aksi	MinSup port	MinConfi dence	Jumlah Itemset	Jumlah Rules	Lift Mean	Lift Median	Lift Min	Lift Max	Conf Mean
3	911	0,01	0,4	71	9	2,33	1,28	1,09	10,85	0,52
3	911	0,02	0,4	39	2	1,28	1,28	1,22	1,33	0,48
3	911	0,03	0,4	27	1	1,33	1,33	1,33	1,33	0,52
3	911	0,04	0,4	19	1	1,33	1,33	1,33	1,33	0,52
3	911	0,05	0,4	16	1	1,33	1,33	1,33	1,33	0,52

Berdasarkan hasil pengujian pada Cluster 3 dengan 911 transaksi, terlihat bahwa penggunaan *minimum support* 0,01 menghasilkan pola yang lebih beragam dengan total 71 *itemset* dan 9 aturan asosiasi. Nilai *lift* rata-rata sebesar 2,33 dengan maksimum 10,85 menunjukkan adanya aturan dengan asosiasi yang cukup kuat, meskipun median hanya 1,28 yang berarti sebagian besar aturan cenderung moderat. Namun, ketika nilai *support* dinaikkan menjadi 0,02 hingga 0,05, jumlah aturan berkurang drastis menjadi hanya 1–2 aturan dengan nilai *lift* yang homogen di kisaran 1,28–1,33, serta *confidence* rata-rata relatif stabil sekitar 0,48–0,52. Kondisi ini memperlihatkan bahwa pada ambang *support* lebih tinggi, aturan yang

terbentuk menjadi sangat terbatas dan kurang bervariasi. Oleh karena itu, nilai *minimum support* 0,01 dipilih untuk Cluster 3 karena mampu mengungkap lebih banyak pola asosiasi yang relevan dengan kekuatan hubungan antar produk yang lebih bermakna, sementara *minimum confidence* 0,4 tetap digunakan agar aturan yang dihasilkan memiliki tingkat keyakinan yang layak untuk dianalisis.

Tabel 5. 11 Perbandingan Minimum support klaster 4

Clus ter	Trans aksi	MinSup port	MinConfi dence	Jumlah Itemset	Jumlah Rules	Lift Mean	Lift Median	Lift Min	Lift Max	Conf Mean
4	1348	0,01	0,4	72	3	6,45	8,43	1,31	9,63	0,83
4	1348	0,02	0,4	41	2	5,47	5,47	1,31	9,63	0,75
4	1348	0,03	0,4	27	0	NaN	NaN	NaN	NaN	NaN
4	1348	0,04	0,4	22	0	NaN	NaN	NaN	NaN	NaN
4	1348	0,05	0,4	18	0	NaN	NaN	NaN	NaN	NaN

Berdasarkan hasil pengujian pada Cluster 4 dengan 1.348 transaksi, terlihat bahwa penggunaan *minimum support* 0,01 menghasilkan 72 *itemset* dengan 3 aturan asosiasi. Nilai *lift* rata-rata mencapai 6,45 dengan median 8,43 dan maksimum 9,63, yang menunjukkan adanya hubungan antar produk dengan kekuatan asosiasi yang sangat tinggi. Ketika nilai *support* dinaikkan menjadi 0,02, jumlah *itemset* berkurang menjadi 41 dengan hanya 2 aturan terbentuk, serta nilai *lift* rata-rata menurun ke 5,47. Namun, pada ambang *support* 0,03 hingga 0,05, tidak ada aturan yang terbentuk sama sekali, sehingga pola keterkaitan antar produk tidak dapat diidentifikasi. Kondisi ini menegaskan bahwa semakin tinggi nilai *support*, semakin berkurang keragaman aturan yang dihasilkan, bahkan hingga hilang sama sekali. Oleh karena itu, nilai *minimum support* 0,01 dipilih sebagai yang paling tepat untuk Cluster 4 karena mampu mengungkap aturan dengan kekuatan asosiasi tinggi, sementara *minimum confidence* tetap dipertahankan pada 0,4 agar aturan yang terbentuk memiliki tingkat keyakinan yang memadai dan relevan untuk dianalisis.

Tabel 5. 12 Perbandingan Minimum support klaster 5

Clus ter	Trans aksi	MinSup port	MinConfi dence	Jumlah Itemset	Jumlah Rules	Lift Mean	Lift Median	Lift Min	Lift Max	Conf Mean
5	168	0,01	0,4	87	32	4,72	1,79	1,02	16,80	0,57
5	168	0,02	0,4	45	9	2,38	1,22	1,02	8,84	0,50
5	168	0,03	0,4	27	3	3,85	1,65	1,05	8,84	0,64
5	168	0,04	0,4	22	2	1,35	1,35	1,05	1,65	0,46

Clus ter	Trans aksi	MinSup port	MinConfi dence	Jumlah Itemset	Jumlah Rules	Lift Mean	Lift Median	Lift Min	Lift Max	Conf Mean
5	168	0,05	0,4	17	0	NaN	NaN	NaN	NaN	NaN

Berdasarkan hasil pengujian pada Cluster 5 dengan 168 transaksi, terlihat bahwa penggunaan *minimum support* 0,01 menghasilkan 87 *itemset* dan 32 aturan asosiasi, dengan nilai *lift* rata-rata 4,72 dan maksimum 16,80. Hasil ini menunjukkan banyaknya pola keterkaitan yang muncul dengan kekuatan asosiasi yang cukup tinggi. Namun, jumlah aturan yang terlalu banyak berpotensi menimbulkan kesulitan dalam interpretasi. Ketika *support* dinaikkan menjadi 0,02, jumlah *itemset* turun menjadi 45 dengan 9 aturan, serta *lift* rata-rata berkurang menjadi 2,38. Pada nilai *support* 0,03, jumlah aturan menyusut lagi menjadi 3, meskipun *lift* rata-rata sempat meningkat ke 3,85, yang menandakan adanya beberapa aturan dengan kekuatan asosiasi lebih terfokus. Selanjutnya, pada *support* 0,04 aturan hanya tersisa 2 dengan *lift* rata-rata 1,35, sedangkan pada *support* 0,05 tidak terbentuk aturan sama sekali. Berdasarkan hasil ini, penggunaan *minimum support* 0,02 dipandang lebih tepat untuk Cluster 5 karena menghasilkan jumlah aturan yang lebih proporsional yaitu 9 aturan dengan kekuatan asosiasi yang masih relevan, sehingga lebih mudah dianalisis tanpa kehilangan makna. Sementara itu, *minimum confidence* tetap ditetapkan pada 0,4 untuk memastikan aturan yang terbentuk memiliki tingkat keyakinan yang layak dan konsisten dalam merepresentasikan pola pembelian konsumen.

Tabel 5. 13 Perbandingan Minimum support klaster 6

Clus ter	Trans aksi	MinSup port	MinConfi dence	Jumlah Itemset	Jumlah Rules	Lift Mean	Lift Median	Lift Min	Lift Max	Conf Mean
6	535	0,01	0,4	65	5	3,02	1,33	1,01	10,29	0,56
6	535	0,02	0,4	38	2	1,03	1,03	1,01	1,05	0,41
6	535	0,03	0,4	26	1	1,05	1,05	1,05	1,05	0,42
6	535	0,04	0,4	19	1	1,05	1,05	1,05	1,05	0,42
6	535	0,05	0,4	17	1	1,05	1,05	1,05	1,05	0,42

Berdasarkan hasil pengujian pada Cluster 6 dengan 535 transaksi, penggunaan *minimum support* 0,01 menghasilkan 65 *itemset* dengan 5 aturan asosiasi, serta nilai *lift* rata-rata 3,02 dan maksimum 10,29 yang menunjukkan adanya pola keterkaitan cukup kuat antar produk. Namun, ketika nilai *support* dinaikkan menjadi 0,02, jumlah aturan menyusut menjadi hanya 2 dengan *lift* rata-rata turun drastis ke 1,03, dan pada ambang 0,03 hingga 0,05 aturan semakin terbatas menjadi 1 dengan nilai *lift* yang homogen di sekitar 1,05. Hal ini menandakan bahwa semakin

tinggi nilai *support*, semakin hilang keragaman dan kekuatan asosiasi antar produk. Oleh karena itu, *minimum support* 0,01 dipilih untuk Cluster 6 karena menghasilkan aturan yang lebih bervariasi sekaligus memiliki kekuatan asosiasi yang signifikan, sementara *minimum confidence* tetap ditetapkan pada 0,4 untuk menjaga keseimbangan antara jumlah aturan yang terbentuk dan tingkat keyakinan yang layak untuk dianalisis.

Secara keseluruhan, hasil pengujian *association rules mining* pada enam klaster memperlihatkan bahwa nilai *minimum support* yang rendah cenderung menghasilkan lebih banyak aturan dengan kekuatan asosiasi yang kuat. Pada Cluster 1, 2, 3, 4, dan 6, penggunaan *support* 0,01 terbukti paling optimal karena mampu menampilkan aturan yang beragam dengan nilai *lift* yang tinggi, misalnya hingga 13,10 pada Cluster 1 dan 12,01 pada Cluster 2. Peningkatan nilai *support* justru membuat jumlah aturan turun drastis menjadi sangat sedikit bahkan ada yang hilang sama sekali, sehingga informasi penting menjadi tereduksi. Sementara itu, Cluster 5 menunjukkan karakteristik berbeda, di mana pada *support* 0,01 aturan yang terbentuk terlalu banyak yaitu 32 aturan dan berpotensi menyulitkan interpretasi. Oleh karena itu, nilai *support* dinaikkan ke 0,02 sehingga aturan yang dihasilkan lebih proporsional yaitu 9 aturan namun tetap informatif dengan *lift* maksimum 8,84. Adapun *minimum confidence* dipertahankan pada angka 0,4 di seluruh klaster untuk menjaga konsistensi kekuatan aturan. Dengan demikian, penetapan parameter ini memastikan aturan yang terbentuk tetap relevan, representatif, dan dapat digunakan sebagai dasar dalam menganalisis pola pembelian konsumen pada masing-masing klaster.

5.2.2 Analisis Frekuensi Pembelian

Analisis frekuensi pembelian dilakukan untuk mengidentifikasi produk-produk yang paling sering dipilih konsumen dalam setiap transaksi. Tahap ini bertujuan untuk menggambarkan tingkat popularitas setiap produk serta kontribusinya terhadap keseluruhan penjualan. Dengan mengetahui produk mana yang memiliki frekuensi pembelian tinggi maupun rendah, dapat diperoleh gambaran mengenai preferensi konsumen yang lebih rinci. Berikut data dari seluruh frekuensi produk pada tiap klaster:

Tabel 5. 14 Top 10 Frekuensi pada klaster 1

Produk	Frekuensi	Share
Corndog Mozzarella	455	20,70%
Corndog Sosis Mozzarella	367	16,70%
Corndog Choco	224	10,19%
Topokki Extra Mozzarella	146	6,64%

Produk	Frekuensi	Share
Odeng Saus Korea	101	4,60%
Topokki Original	92	4,19%
Corndog Matcha	75	3,41%
Odeng Fish Soup	63	2,87%
Bukokki	61	2,78%
Hot Spicy Bibimmyeon	59	2,68%

Berdasarkan hasil analisis frekuensi produk pada klaster 1, terlihat bahwa Corndog Mozzarella menempati posisi teratas dengan jumlah pembelian sebanyak 455 kali atau sekitar 20,70% dari total transaksi. Produk ini diikuti oleh Corndog Sosis Mozzarella dengan 367 kali pembelian (16,70%) dan Corndog Choco dengan 224 kali pembelian (10,19%), sehingga ketiga produk tersebut dapat dikategorikan sebagai menu utama yang paling diminati konsumen. Selanjutnya, Topokki Extra Mozzarella juga cukup populer dengan frekuensi 146 kali (6,64%), disusul Odeng Saus Korea sebanyak 101 kali (4,60%) dan Topokki Original sebanyak 92 kali (4,19%). Sementara itu, produk dengan frekuensi pembelian yang lebih rendah yaitu Corndog Matcha (3,41%), Odeng Fish Soup (2,87%), Bukokki (2,78%), dan Hot Spicy Bibimmyeon (2,68%) yang masing-masing memiliki porsi di bawah 5%. Temuan ini menunjukkan bahwa variasi menu berbasis *corndog*, khususnya varian mozzarella, masih mendominasi preferensi konsumen, sedangkan menu berbasis *topokki*, *odeng*, dan *bibimmyeon* cenderung menjadi pelengkap dengan kontribusi lebih kecil terhadap keseluruhan transaksi.

Tabel 5. 15 Top 10 Frekuensi pada klaster 2

Produk	Frekuensi	Share
Corndog Mozzarella	485	18,97%
Corndog Sosis Mozzarella	406	15,88%
Corndog Choco	262	10,25%
Topokki Extra Mozzarella	139	5,44%
Odeng Saus Korea	121	4,73%
Topokki Original	113	4,42%
Bukokki	94	3,68%
Corndog Matcha	87	3,4%
Hot Spicy Bibimmyeon	82	3,21%
Odeng Fish Soup	79	3,09%

Berdasarkan data frekuensi pembelian produk pada klaster 2, Corndog Mozzarella menjadi produk yang paling banyak dipilih konsumen dengan 485 kali pembelian atau sebesar 18,97% dari total transaksi. Posisi berikutnya ditempati oleh Corndog Sosis Mozzarella dengan 406 kali

pembelian (15,88%) dan Corndog Choco dengan 262 kali pembelian (10,25%), sehingga ketiganya dapat dikategorikan sebagai produk utama dengan tingkat permintaan tertinggi. Sementara itu, Topokki Extra Mozzarella tercatat sebanyak 139 kali (5,44%), Odeng Saus Korea 121 kali (4,73%), dan Topokki Original 113 kali (4,42%), menunjukkan popularitas yang cukup baik meskipun porsi lebih kecil dibandingkan varian corndog. Adapun produk lain seperti Bukokki (3,68%), Corndog Matcha (3,40%), Hot Spicy Bibimmyeon (3,21%), dan Odeng Fish Soup (3,09%) memiliki kontribusi di bawah 5%, sehingga lebih merepresentasikan kelompok produk pelengkap. Secara keseluruhan, menu berbasis corndog, terutama varian mozzarella, tetap mendominasi preferensi konsumen, sedangkan produk berbasis topokki, odeng, dan bibimmyeon berperan sebagai variasi tambahan dalam pilihan pembelian.

Tabel 5. 16 Top 10 Frekuensi pada klaster 3

Produk	Frekuensi	Share
Corndog Mozzarella	513	19,89%
Corndog Sosis Mozzarella	422	16,36%
Corndog Choco	279	10,82%
Topokki Extra Mozzarella	165	6,4%
Topokki Original	119	4,61%
Odeng Saus Korea	112	4,34%
Hot Spicy Bibimmyeon	91	3,53%
Odeng Fish Soup	87	3,37%
Bukokki	81	3,14%
Corndog Matcha	81	3,14%

Berdasarkan data frekuensi pembelian pada klaster 3, Corndog Mozzarella kembali menempati posisi teratas dengan jumlah 513 kali pembelian atau sebesar 19,89% dari total transaksi. Produk ini diikuti oleh Corndog Sosis Mozzarella dengan 422 kali pembelian (16,36%) dan Corndog Choco dengan 279 kali pembelian (10,82%), yang menegaskan bahwa varian corndog masih menjadi menu paling dominan dalam preferensi konsumen. Selanjutnya, Topokki Extra Mozzarella tercatat sebanyak 165 kali (6,40%), diikuti Topokki Original 119 kali (4,61%) serta Odeng Saus Korea 112 kali (4,34%). Produk lainnya seperti Hot Spicy Bibimmyeon (3,53%), Odeng Fish Soup (3,37%), Bukokki (3,14%), dan Corndog Matcha (3,14%) menunjukkan kontribusi yang relatif kecil dengan porsi di bawah 5%. Secara umum, hasil ini kembali menunjukkan bahwa menu berbasis corndog, khususnya varian mozzarella, mendominasi pola pembelian, sementara produk topokki, odeng, bibimmyeon, dan varian corndog lainnya berfungsi sebagai pelengkap yang memperkaya variasi menu.

Tabel 5. 17 Top 10 Frekuensi pada klaster 4

Produk	Frekuensi	Share
Corndog Mozzarella	763	19,95%
Corndog Sosis Mozzarella	563	14,72%
Corndog Choco	401	10,48%
Topokki Extra Mozzarella	206	5,39%
Odeng Saus Korea	164	4,29%
Topokki Original	159	4,16%
Odeng Fish Soup	143	3,74%
Bukokki	138	3,61%
Corndog Matcha	129	3,37%
Corndog Sosis	119	3,11%

Berdasarkan hasil frekuensi pembelian pada klaster 4, Corndog Mozzarella menjadi produk dengan tingkat permintaan tertinggi, yaitu sebanyak 763 kali atau sebesar 19,95% dari total penjualan. Posisi selanjutnya ditempati oleh Corndog Sosis Mozzarella dengan 563 kali pembelian (14,72%) dan Corndog Choco dengan 401 kali pembelian (10,48%), sehingga ketiga produk ini dapat dikategorikan sebagai menu utama yang paling diminati konsumen. Produk dengan frekuensi menengah ditunjukkan oleh Topokki Extra Mozzarella sebanyak 206 kali (5,39%), Odeng Saus Korea 164 kali (4,29%), dan Topokki Original 159 kali (4,16%). Sementara itu, produk dengan kontribusi relatif kecil adalah Odeng Fish Soup (3,74%), Bukokki (3,61%), Corndog Matcha (3,37%), dan Corndog Sosis (3,11%), yang masing-masing menyumbang di bawah 5% dari total transaksi. Temuan ini memperkuat kesimpulan bahwa varian corndog, terutama yang berbasis mozzarella, mendominasi preferensi konsumen, sedangkan produk lain seperti topokki, odeng, dan varian tambahan lebih berfungsi sebagai pelengkap yang memberikan variasi pada pilihan menu.

Tabel 5. 18 Top 10 Frekuensi pada klaster 5

Produk	Frekuensi	Share
Corndog Mozzarella	97	20,04%
Corndog Sosis Mozzarella	79	16,32%
Corndog Choco	51	10,54%
Topokki Extra Mozzarella	29	5,99%
Odeng Fish Soup	22	4,55%
Topokki Original	20	4,13%
Odeng Saus Korea	19	3,93%
Bukokki	14	2,89%
Corndog Matcha	14	2,89%
Pakokki	14	2,89%

Berdasarkan data frekuensi pembelian pada klaster 5, Corndog Mozzarella menjadi produk yang paling diminati dengan 97 kali pembelian atau 20,04% dari total transaksi. Posisi berikutnya ditempati oleh Corndog Sosis Mozzarella sebanyak 79 kali (16,32%) dan Corndog Choco sebanyak 51 kali (10,54%), sehingga ketiganya kembali menegaskan dominasi varian corndog sebagai menu utama. Selanjutnya, Topokki Extra Mozzarella tercatat 29 kali (5,99%), disusul Odeng Fish Soup 22 kali (4,55%) dan Topokki Original 20 kali (4,13%). Produk dengan kontribusi lebih kecil terlihat pada Odeng Saus Korea (3,93%), Bukokki (2,89%), Corndog Matcha (2,89%), serta Pakokki (2,89%) yang masing-masing hanya menyumbang porsi kecil dari total penjualan. Secara keseluruhan, hasil ini menunjukkan bahwa varian corndog, khususnya mozzarella, tetap menjadi produk unggulan dengan pangsa terbesar, sementara produk lain seperti topokki, odeng, dan varian tambahan berfungsi sebagai pelengkap dengan daya tarik yang relatif terbatas.

Tabel 5. 19 Top 10 Frekuensi pada klaster 6

Produk	Frekuensi	Share
Corndog Mozzarella	331	21,99%
Corndog Sosis Mozzarella	226	15,02%
Corndog Choco	159	10,56%
Topokki Extra Mozzarella	89	5,91%
Odeng Saus Korea	75	4,98%
Topokki Original	61	4,05%
Corndog Matcha	56	3,72%
Odeng Fish Soup	55	3,65%
Corndog Sosis	52	3,46%
Bukokki	43	2,86%

Berdasarkan data frekuensi pembelian, Corndog Mozzarella menempati posisi teratas dengan jumlah 331 kali pembelian atau sebesar 21,99% dari total transaksi. Produk ini diikuti oleh Corndog Sosis Mozzarella sebanyak 226 kali (15,02%) dan Corndog Choco sebanyak 159 kali (10,56%), sehingga ketiganya konsisten menjadi varian corndog yang paling diminati konsumen. Selanjutnya, Topokki Extra Mozzarella tercatat 89 kali (5,91%), Odeng Saus Korea 75 kali (4,98%), dan Topokki Original 61 kali (4,05%), yang menunjukkan popularitas menengah pada produk berbasis topokki dan odeng. Adapun Corndog Matcha (3,72%), Odeng Fish Soup (3,65%), Corndog Sosis (3,46%), serta Bukokki (2,86%) hanya menyumbang porsi di bawah 5% dari keseluruhan transaksi, sehingga lebih berperan sebagai pelengkap. Secara

keseluruhan, hasil ini kembali menegaskan dominasi varian corndog, khususnya mozzarella, dalam preferensi konsumen, sedangkan produk berbasis topokki, odeng, dan varian tambahan lainnya menjadi opsi sekunder yang memperkaya variasi menu.

5.2.3 Analisis Output Association Rules

Analisis output *association Rules* dilakukan untuk menginterpretasikan aturan-aturan yang terbentuk dari proses *market basket analysis* menggunakan algoritma FP-Growth. Aturan asosiasi yang dihasilkan memberikan gambaran mengenai kombinasi produk yang paling sering dibeli secara bersamaan oleh konsumen, beserta tingkat kekuatan hubungan antar produk melalui ukuran *support*, *confidence*, dan *lift*. Tahap ini menjadi krusial karena memungkinkan identifikasi pola pembelian yang berulang, baik dalam bentuk pasangan produk maupun kombinasi lebih kompleks. Berikut pemaparan data tiap klasternya:

Tabel 5. 20 Aturan Lengkap Asosiasi Pada Klaster 1

antecedent	consequent	support	confidence	lift	Support Count
Odeng Fish Soup Extra Odeng 2 Tusuk	Odeng Fish Soup	0,018421053	1	13,10344828	14
Topokki Extra Odeng	Corndog Sosis Mozzarella	0,01842	0,51852	1,41753	9
Corndog Tiramisu	Corndog Choco	0,01974	0,41667	1,41369	14
Topokki Extra Mozzarella Pakokki	Corndog Mozzarella	0,08553	0,53279	1,31467	15
	Corndog Sosis Mozzarella	0,03026	0,47917	1,30995	65
Mandu Saus Gochujang	Corndog Sosis Mozzarella	0,01711	0,40625	1,11061	23
Mandu Saus Gochujang	Corndog Mozzarella	0,01842	0,4375	1,07955	13

Berdasarkan aturan asosiasi yang terbentuk pada Klaster 1 dengan jumlah transaksi sebanyak 760, ditemukan beberapa pola keterkaitan antar produk yang cukup signifikan. Aturan dengan kekuatan asosiasi paling tinggi ditunjukkan oleh kombinasi Odeng Fish Soup Extra Odeng 2 Tusuk terhadap Odeng Fish Soup, dengan nilai *support* 0,01842, *confidence* 0,51852; dan *lift* sebesar 1,41753; yang berarti setiap pembelian menu tambahan tersebut hampir selalu diikuti oleh pembelian Odeng Fish Soup dengan kekuatan asosiasi yang sangat kuat. Pola serupa juga terlihat pada Corndog Tiramisu terhadap Corndog Choco dengan *lift* 1,41369; meskipun nilai *support* lebih kecil yaitu 0,01184. Selain itu, terdapat keterkaitan antara Topokki Extra Odeng

dan Corndog Sosis Mozzarella dengan *support* 0,01842, *confidence* 0,51852, dan *lift* 1,41, yang menunjukkan adanya kecenderungan konsumen membeli corndog setelah memilih menu topokki tersebut. Hubungan lain terlihat pada Corndog Tiramisu dan Corndog Choco dengan *support* 0,01974, *confidence* 0,41667, dan *lift* 1,41, yang menandakan adanya hubungan substitusi atau kecenderungan pembelian bersama antar varian corndog manis. Sementara itu, Topokki Extra Mozzarella memiliki asosiasi yang cukup kuat dengan Corndog Mozzarella, ditunjukkan dengan *support* tertinggi 0,08553 dan *confidence* 0,53279 serta *lift* 1,31. Aturan lain seperti Pakokki terhadap Corndog Sosis Mozzarella (*lift* 1,30), Mandu Saus Gochujang terhadap Corndog Sosis Mozzarella (*lift* 1,11), dan Mandu Saus Gochujang terhadap Corndog Mozzarella (*lift* 1,08) memperlihatkan bahwa produk berbasis mandu atau pakokki cenderung dibeli bersama dengan varian corndog populer, meskipun kekuatan asosiasinya relatif lebih rendah.

Tabel 5. 21 Aturan Lengkap Asosiasi Pada Klaster 2

antecedent	consequent	support	confidence	lift	support_count
Odeng Fish Soup	Odeng Fish Soup	0,0142387	1	12,0131	13
Extra Odeng 2 Tusuk		73		5789	
Odeng Saus Korea	Odeng Saus Korea	0,0142387	1	7,87068	13
Extra Odeng 2 Tusuk		73		9655	
Rabokki Extra	Corndog	0,0197152	0,40909090	1,05508	18
Mozzarella	Mozzarella	25	9	4746	
Topokki Extra	Corndog	0,0503833	0,4	1,03163	46
Mozzarella	Mozzarella	52		8418	

Pada Klaster 2 dengan total 913 transaksi, ditemukan beberapa aturan asosiasi yang menggambarkan keterkaitan antar produk. Aturan paling kuat terlihat pada hubungan Odeng Fish Soup Extra Odeng 2 Tusuk terhadap Odeng Fish Soup dengan nilai *support* 0,014, *confidence* 1,00, dan *lift* 12,01. Artinya, setiap kali konsumen membeli tambahan odeng tersebut, hampir selalu diikuti dengan pembelian Odeng Fish Soup, dengan kekuatan asosiasi yang sangat tinggi. Pola serupa juga terlihat pada Odeng Saus Korea Extra Odeng 2 Tusuk terhadap Odeng Saus Korea dengan *support* 0,014, *confidence* 1,00, dan *lift* 7,87, yang menunjukkan keterkaitan erat antara menu tambahan dan menu utamanya. Sementara itu, Rabokki Extra Mozzarella memiliki hubungan dengan Corndog Mozzarella dengan *support* 0,020, *confidence* 0,41, dan *lift* 1,06, namun asosiasi ini tergolong lemah karena *lift* mendekati 1. Hal yang sama juga terjadi pada Topokki Extra Mozzarella terhadap Corndog Mozzarella, di mana *support* mencapai 0,050 dengan *confidence* 0,40, namun *lift* hanya 1,03 sehingga

mengindikasikan keterkaitan yang rendah. Dengan demikian, aturan asosiasi pada Klaster 2 menunjukkan bahwa pola pembelian paling kuat muncul pada menu tambahan odeng dengan produk odeng utama, sedangkan keterkaitan antara menu berbasis topokki atau rabokki dengan corndog relatif lemah.

Tabel 5. 22 Aturan Lengkap Asosiasi Pada Klaster 3

antecedent	consequent	support	confidence	lift	support_count
Odeng Fish Soup Extra Odeng 2 Tusuk	Odeng Fish Soup	0,01646542 3	1	10,845238 1	Odeng Fish Soup Extra Odeng 2 Tusuk 17
Rabokki Extra Mozzarella	Corndog Choco	0,019	0,459	1,50	13
Chikuwa Saus Korea	Corndog Sosis Mozzarella	0,014	0,481	1,36	67
Topokki Extra Mozzarella	Corndog Mozzarella	0,074	0,519	1,33	11
Corndog Choco, Topokki Extra Mozzarella	Corndog Mozzarella	0,012	0,5	1,28	14
Topokki Extra Odeng	Corndog Sosis Mozzarella	0,015	0,452	1,27	19
Pakokki	Corndog Sosis Mozzarella	0,021	0,432	1,22	16
Rabokki Extra Mozzarella	Corndog Mozzarella	0,018	0,432	1,11	11
Corndog Blueberry	Corndog Mozzarella	0,012	0,423	1,09	

Pada Klaster 3 dengan jumlah 911 transaksi, ditemukan 8 aturan asosiasi yang menunjukkan pola keterkaitan antar produk. Aturan dengan kekuatan paling tinggi terlihat antara kombinasi Odeng Fish Soup Extra Odeng 2 Tusuk terhadap Odeng Fish Soup dengan *support* 0,0267, *confidence* 1,00, dan *lift* 9,63. Lalu ada Rabokki Extra Mozzarella dan Corndog Choco dengan *lift* 1,50, meskipun nilai *confidence* hanya 0,459, sehingga asosiasinya lebih spesifik pada sebagian konsumen. Keterkaitan lain yang cukup menonjol adalah Chikuwa Saus Korea dengan Corndog Sosis Mozzarella (*lift* 1,36) serta Topokki Extra Mozzarella dengan Corndog Mozzarella (*support* 0,074, *confidence* 0,519, *lift* 1,33), yang menjadi aturan dengan dukungan

terbesar di klaster ini. Beberapa kombinasi lain seperti Corndog Choco dengan Topokki Extra Mozzarella terhadap Corndog Mozzarella (*lift* 1,28), Topokki Extra Odeng terhadap Corndog Sosis Mozzarella (*lift* 1,27), dan Pakokki terhadap Corndog Sosis Mozzarella (*lift* 1,22) juga memperlihatkan keterkaitan meski kekuatannya relatif moderat. Sementara itu, Rabokki Extra Mozzarella terhadap Corndog Mozzarella (*lift* 1,11) serta Corndog Blueberry terhadap Corndog Mozzarella (*lift* 1,09) menunjukkan hubungan yang lemah karena nilai *lift* mendekati 1. Dengan demikian, aturan asosiasi pada Klaster 3 memperlihatkan kecenderungan keterkaitan antara varian topokki atau rabokki dengan corndog.

Tabel 5. 23 Aturan Lengkap Asosiasi Pada Klaster 4

antecedent	consequent	support	confidence	lift	support_count
Odeng Fish Soup Extra Odeng 2 Tusuk	Odeng Fish Soup	0,026706 231	1	9,628571 429	36
Odeng Saus Korea Extra Odeng 2 Tusuk	Odeng Saus Korea	0,0111	1	8,43	15
Topokki Extra Odeng Mozzarella	Corndog Mozzarella	0,0208	0,491	1,31	28

Pada Klaster 4 dengan jumlah 1.348 transaksi, ditemukan 3 aturan asosiasi yang menggambarkan pola keterkaitan produk. Aturan paling kuat terlihat pada kombinasi Odeng Fish Soup Extra Odeng 2 Tusuk terhadap Odeng Fish Soup dengan *support* 0,0267, *confidence* 1,00, dan *lift* 9,63, yang menunjukkan bahwa setiap kali menu tambahan tersebut dipilih, selalu diikuti dengan pembelian Odeng Fish Soup. Pola serupa juga terlihat pada Odeng Saus Korea Extra Odeng 2 Tusuk terhadap Odeng Saus Korea, dengan *support* 0,0111, *confidence* 1,00, dan *lift* 8,43, sehingga menegaskan keterkaitan kuat antara menu utama dan tambahan yang sejenis. Sementara itu, aturan ketiga menunjukkan hubungan antara Topokki Extra Odeng Mozzarella dan Corndog Mozzarella, dengan *support* 0,0208, *confidence* 0,491, dan *lift* 1,31, yang mengindikasikan adanya kecenderungan konsumen untuk membeli corndog setelah memilih varian topokki tersebut, meskipun kekuatan asosiasinya tergolong moderat. Dengan demikian, hasil asosiasi pada Klaster 4 menekankan pola pembelian yang sangat kuat pada pasangan menu utama dan tambahan berbasis odeng, serta adanya hubungan tambahan antara varian topokki dan corndog.

Tabel 5. 24 Aturan Lengkap Asosiasi Pada Klaster 5

antecedent	consequent	support	confidence	lift	support_count
Odeng Fish Soup Extra Odeng 2 Tusuk	Odeng Fish Soup	0,035714 286	1	8,842105 263	6
Rabokki	Odeng Fish Soup	0,024	0,4	3,54	4
Corndog Strawberry	Corndog Choco	0,024	0,57	1,88	4
Corndog Matcha	Corndog Choco	0,042	0,5	1,65	7
Bukokki	Corndog Sosis Mozzarella	0,024	0,44	1,22	4
Rabokki	Corndog Sosis Mozzarella	0,024	0,4	1,10	4
Hot Spicy Bibimmyeon	Corndog Sosis Mozzarella	0,024	0,4	1,10	4
Topokki Original	Corndog Mozzarella	0,042	0,41	1,05	7
Rabokki	Corndog Mozzarella	0,024	0,4	1,02	4

Pada Klaster 5 dengan jumlah 168 transaksi, ditemukan sembilan aturan asosiasi yang menunjukkan keterkaitan produk dengan variasi kekuatan hubungan. Aturan paling kuat terlihat pada Odeng Fish Soup Extra Odeng 2 Tusuk terhadap Odeng Fish Soup dengan *support* 0,036, *confidence* 1,00, dan *lift* 8,84, yang menandakan adanya keterkaitan sangat kuat antara menu utama dan tambahan. Selain itu, Rabokki juga memiliki asosiasi dengan Odeng Fish Soup dengan *lift* 3,54, meskipun nilai *confidence* hanya 0,40 sehingga kekuatan asosiasinya terbatas pada sebagian transaksi. Varian corndog juga menunjukkan keterkaitan antar sesama produk, misalnya Corndog Strawberry terhadap Corndog Choco dengan *lift* 1,88 dan *confidence* 0,57, serta Corndog Matcha terhadap Corndog Choco dengan *lift* 1,65, yang menandakan adanya kecenderungan konsumen memilih corndog manis secara bersamaan. Beberapa aturan lain memperlihatkan hubungan moderat hingga lemah, seperti Bukokki terhadap Corndog Sosis Mozzarella (*lift* 1,22), Rabokki terhadap Corndog Sosis Mozzarella (*lift* 1,10), dan Hot Spicy Bibimmyeon terhadap Corndog Sosis Mozzarella (*lift* 1,10). Sementara itu, Topokki Original dan Rabokki yang berasosiasi dengan Corndog Mozzarella hanya memiliki nilai *lift* mendekati 1 (masing-masing 1,05 dan 1,02), sehingga keterkaitan keduanya tergolong sangat lemah. Secara keseluruhan, aturan asosiasi pada Klaster 5 menegaskan adanya hubungan dominan antara menu tambahan dan menu utama berbasis odeng, serta menunjukkan bahwa varian

corndog manis cenderung dibeli secara bersamaan, meskipun keterkaitan dengan produk lain relatif rendah.

Tabel 5. 25 Aturan Lengkap Asosiasi Pada Klaster 6

antecedent	consequent	support	confidence	lift	support_count
Odeng Fish Soup Extra Odeng 2 Tusuk	Odeng Fish Soup	0,019	1	10,29	10
Corndog Tiramisu	Corndog Choco	0,011	0,43	1,44	6
Topokki Extra Odeng Mozzarella	Corndog Mozzarella	0,013	0,54	1,33	7
Topokki Extra Mozzarella	Corndog Mozzarella	0,056	0,42	1,05	30
Chikuwa Saus Korea	Corndog Mozzarella	0,021	0,41	1,01	11

Pada Klaster 6 dengan jumlah 535 transaksi, diperoleh lima aturan asosiasi dengan kekuatan hubungan yang bervariasi. Aturan terkuat terlihat pada kombinasi Odeng Fish Soup Extra Odeng 2 Tusuk terhadap Odeng Fish Soup, dengan *support* 0,019, *confidence* 1,00, dan *lift* 10,29, yang menunjukkan keterkaitan sangat kuat antara menu utama dan tambahan. Hubungan lain yang cukup menonjol adalah Corndog Tiramisu terhadap Corndog Choco dengan *support* 0,011, *confidence* 0,43, dan *lift* 1,44, yang menandakan adanya kecenderungan konsumen membeli kedua varian corndog manis secara bersamaan. Sementara itu, Topokki Extra Odeng Mozzarella juga memiliki asosiasi dengan Corndog Mozzarella (*support* 0,013; *confidence* 0,54; *lift* 1,33), yang memperlihatkan adanya keterkaitan moderat antara menu berbasis topokki dan corndog. Aturan lain seperti Topokki Extra Mozzarella terhadap Corndog Mozzarella (*lift* 1,05) dan Chikuwa Saus Korea terhadap Corndog Mozzarella (*lift* 1,01) memiliki nilai *lift* yang mendekati 1, sehingga kekuatan asosiasinya tergolong lemah. Secara keseluruhan, aturan asosiasi pada Klaster 6 menegaskan dominasi keterkaitan kuat pada menu tambahan odeng dengan produk utamanya, disertai beberapa hubungan moderat antara varian topokki atau corndog manis dengan corndog populer, meskipun sebagian besar aturan lainnya relatif lemah.

Nilai *lift* yang tinggi, seperti pada pasangan *Odeng Fish Soup Extra Odeng 2 Tusuk* → *Odeng Fish Soup* dengan nilai mencapai 8 di klaster 2 dan 13 di klaster 1, terjadi karena adanya hubungan deterministik antara kedua produk tersebut yang dipengaruhi oleh nilai *confidence* sebesar 1 menunjukkan bahwa setiap kali produk *Extra Odeng 2 Tusuk* muncul dalam transaksi, produk *Odeng Fish Soup* selalu ikut dibeli, sehingga probabilitas kemunculan bersama ($P(A \cap$

B)) sama dengan probabilitas produk tambahan tersebut ($P(A)$). Sementara itu, produk *Odeng Fish Soup* secara keseluruhan hanya muncul pada sebagian kecil transaksi, sehingga nilai $P(B)$ relatif kecil. Kondisi ini menyebabkan hasil kali $P(A) \times P(B)$ menjadi sangat kecil, dan ketika dibandingkan dengan $P(A \cap B)$, rasio yang dihasilkan menjadi besar, sehingga nilai *lift* meningkat signifikan. Dengan kata lain, *lift* yang tinggi muncul karena produk *consequent* sepenuhnya bergantung pada produk *antecedent* dan tidak pernah muncul sendiri, sehingga hubungan keduanya sangat kuat dan pasti terjadi.

5.3 Rekomendasi Strategi Pemasaran

Berdasarkan hasil analisis *clustering* dan *association rule mining*, fokus strategi akan dilakukan terhadap Klaster 5 yang diidentifikasi sebagai segmen pelanggan dengan karakteristik premium yang memiliki rata-rata jumlah item terbanyak per transaksi, yaitu sekitar 10 item, serta nilai transaksi tertinggi dengan rata-rata Rp113.716.

Produk-produk yang terbentuk oleh pola asosiasi dalam klaster ini meliputi *Odeng Fish Soup Extra*, *Odeng 2 Tusuk*, *Odeng Fish Soup*, *Rabokki*, *Corndog Strawberry*, *Corndog Choco*, *Corndog Matcha*, *Bukokki*, *Corndog Sosis Mozzarella*, *Hot Spicy Bibimmyeon*, *Topokki Original*, dan *Corndog Mozzarella*. Pola dengan nilai *lift* tertinggi terdapat pada kombinasi *Odeng Fish Soup Extra* *Odeng 2 Tusuk* $\rightarrow$ *Odeng Fish Soup* dengan *confidence* sebesar 1,00 dan *lift* 8,84. Hal ini menunjukkan bahwa pelanggan yang membeli tambahan *Extra Odeng* hampir selalu membeli *Odeng Fish Soup* sebagai menu utama, yang menandakan preferensi pelanggan Klaster 5 terhadap produk berkuah dengan porsi tambahan.

Selain itu, ditemukan pula hubungan kuat antara *Rabokki* $\rightarrow$ *Odeng Fish Soup* dengan *lift* sebesar 3,54 yang mengindikasikan kecenderungan pelanggan untuk mengombinasikan menu pedas dan berkuah dalam satu transaksi. Di sisi lain, keterkaitan antarproduk *snack* juga cukup dominan, seperti pada aturan *Corndog Strawberry* $\rightarrow$ *Corndog Choco* dengan *lift* 1,88 dan *Corndog Matcha* $\rightarrow$ *Corndog Choco* dengan *lift* 1,65. Pola ini menggambarkan perilaku pelanggan yang cenderung membeli dua varian rasa manis sekaligus, khususnya pada kategori *Corndog dessert*. Hubungan lain yang patut diperhatikan terdapat pada kombinasi *Bukokki*, *Rabokki*, dan *Hot Spicy Bibimmyeon* terhadap *Corndog Sosis Mozzarella*, serta hubungan *Topokki Original* dan *Rabokki* terhadap *Corndog Mozzarella*, yang menunjukkan peluang *cross-selling* antara menu utama berbasis tepung beras dengan *snack* keju yang populer.

Berdasarkan pola hubungan antarproduk tersebut, strategi promosi yang disarankan untuk Klaster 5 berfokus pada penerapan *product bundling* dan promosi berbasis waktu (*time-based promotion*):

1. Strategi *bundling* dapat diterapkan melalui pembuatan paket Korean Odeng Set yang berisi Odeng Fish Soup dan Extra Odeng 2 Tusuk dengan harga spesial, yang menargetkan pelanggan pecinta menu berkuah dengan porsi besar. Selanjutnya, paket Cheesy Combo Set dapat ditawarkan dengan kombinasi Topokki Original dan Corndog Mozzarella untuk menarik pelanggan yang menyukai cita rasa keju. Untuk segmen pelanggan dengan preferensi rasa manis, dapat dikembangkan paket Sweet Corndog Pair yang berisi Corndog Choco bersama varian Corndog Strawberry atau Corndog Matcha, yang dipromosikan melalui *platform* daring GoFood karena pemesanan terbanyak ada di *platform* tersebut.
2. Selain promosi berbasis paket, penerapan *time-based promotion* juga relevan untuk Klaster 5, mengingat sebagian besar transaksi pada klaster ini terjadi pada malam hari. Program promosi seperti “*Night Korean Delight*” dengan diskon khusus pada pukul 19.00–23.00 dapat mendorong peningkatan frekuensi pembelian dan memperkuat persepsi nilai pelanggan terhadap merek.

BAB VI

PENUTUP

6.1 Kesimpulan

Berdasarkan hasil penelitian serta pembahasan yang telah dilakuakn, maka kesimpulan penelitian ini dapat dijabarkan sebagai berikut:

1. Berdasarkan hasil pengolahan data dengan algoritma *K-Means*, penelitian ini berhasil mengelompokkan pelanggan ke dalam enam klaster dengan karakteristik yang berbeda. Klaster 1 terdiri dari 1.296 transaksi dengan rata-rata 3,025 item dan nilai transaksi Rp36.234,5, sedangkan Klaster 2 berjumlah 1.451 transaksi dengan rata-rata 6,195 item dan total pembelian Rp63.642,7. Klaster 3 mencatat 1.518 transaksi dengan rata-rata 2,733 item dan nilai Rp21.725,0, diikuti Klaster 4 dengan 2.195 transaksi, rata-rata 3,118 item dan nilai Rp36.771,1. Selanjutnya, Klaster 5 terdiri dari 291 transaksi dengan rata-rata 10,306 item dan total pembelian Rp113.716,0, sedangkan Klaster 6 mencatat 904 transaksi dengan rata-rata 2,822 item dan total Rp38.266,3. Perbedaan angka tersebut menunjukkan bahwa tiap klaster memiliki pola pembelian yang khas, baik dari sisi jumlah produk yang dibeli maupun nilai transaksi yang dihasilkan.
2. Analisis aturan asosiasi dengan algoritma *FP-Growth* menunjukkan adanya keterkaitan produk pada tiap klaster. Pada Klaster 1, aturan yang terbentuk adalah Odeng Fish Soup Extra Odeng 2 Tusuk dan Odeng Fish Soup dengan nilai *support* 0,036, *confidence* 1,00, dan *lift* 8,84, sementara pada Klaster 2 aturan yang sama terbentuk dengan *support* 0,019, *confidence* 1,00, dan *lift* 10,29. Klaster 3 memperlihatkan hubungan antara Corndog Tiramisu dan Corndog Choco dengan *support* 0,011, *confidence* 0,43, dan *lift* 1,44, sedangkan Klaster 4 menghasilkan aturan antara Topokki Extra Odeng Mozzarella dan Corndog Mozzarella dengan *support* 0,013, *confidence* 0,54, dan *lift* 1,33. Pada Klaster 5, aturan yang terbentuk adalah Topokki Extra Mozzarella dan Corndog Mozzarella dengan *support* 0,056, *confidence* 0,42, dan *lift* 1,05, serta Klaster 6 menghasilkan aturan Chikuwa Saus Korea dan Corndog Mozzarella dengan *support* 0,021, *confidence* 0,41, dan *lift* 1,01. Angka-angka tersebut menunjukkan adanya kecenderungan konsumen untuk membeli produk tertentu secara bersamaan.
3. Berdasarkan hasil analisis *clustering* dan *association rule mining*, Klaster 5 diidentifikasi sebagai segmen pelanggan premium dengan rata-rata pembelian tertinggi baik dari jumlah

item maupun nilai transaksi. Strategi yang disarankan meliputi penerapan *product bundling* seperti Korean Odeng Set, Cheesy Combo Set, dan Sweet Corndog Pair, serta *time-based promotion* pada pukul 19.00–23.00 untuk meningkatkan loyalitas dan nilai transaksi pelanggan premium..

6.2 Saran

Berikut merupakan saran dari penelitian ini:

1. Saran untuk peneliti selanjutnya

Bagi peneliti selanjutnya, disarankan untuk memperluas cakupan penelitian dengan menambahkan variabel lain seperti karakteristik demografis konsumen, preferensi rasa, ataupun faktor eksternal seperti tren musiman dan kondisi persaingan, sehingga hasil analisis *clustering* dan *association rules mining* dapat memberikan gambaran yang lebih komprehensif mengenai perilaku konsumen. Secara khusus, untuk data kategorikal sebaiknya dipertimbangkan penggunaan algoritma *k-modes* agar hasil klasterisasi lebih representatif pada data yang bersifat kategorikal, mengingat *k-means* lebih optimal diterapkan pada data numerik.

2. Saran untuk Grobbak Korea Pandeyan

Sementara itu, bagi Grobbak Korea Pandeyan, hasil penelitian ini dapat dijadikan dasar dalam merancang strategi pemasaran yang lebih tepat sasaran dengan tetap menyesuaikan pada dinamika pasar dan kebiasaan konsumen. Penerapan program loyalitas, pengembangan *bundling menu premium*, serta pemanfaatan produk unggulan seperti corndog mozzarella sebagai *magnet product* perlu dijalankan secara berkesinambungan dan dievaluasi secara berkala. Selain itu, Grobbak Korea juga sebaiknya mengoptimalkan strategi dengan memperkuat citra merek dengan inovasi menu dan pelayanan, agar dapat mempertahankan loyalitas pelanggan sekaligus menarik konsumen baru di tengah persaingan industri kuliner yang semakin ketat.

DAFTAR PUSTAKA

- Audiana, W., & Tahyudin, I. (2025). *Penerapan Analisis Keranjang Belanja Untuk Analisis Transaksi Penjualan Menggunakan Algoritma Association FP- Growth*. 17(1), 33–49.
- 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), 17–24. <https://doi.org/10.5120/18355-9116>
- Biswas, S., Yoganarasimhan, H., & Zhang, H. (2024). Channel Choice and Customer Value. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4747756>
- Bozanta, A., Berry, S., Cevik, M., Bulut, B., Yigit, D., Gonen, F. F., & Başar, A. (2022). *Time Series Clustering for Grouping Products Based on Price and Sales Patterns*. 2027. <http://arxiv.org/abs/2204.08334>
- Chandra, S., Reddy, K., & Taliyan, S. K. (2025). *MARKET BASKET ANALYSIS FOR RETAIL OPTIMIZATION : “ IDENTIFYING PRODUCT ASSOCIATIONS TO ENHANCE SALES STRATEGIES .”* 04, 565–570.
- Ernst, & Young. (2023). *Future Consumer Index Indonesia: How to navigate the shifting landscape of consumer behavior*. https://www.ey.com/en_id/consumer-products-retail/how-to-navigate-the-shifting-landscape-of-consumer-behavior
- Fajrul, M., & Saptyana, R. F. (2025). *Strategi Peningkatan Daya Saing UMKM Kuliner di Era Digital: Studi Kasus pada Pelaku Usaha di Kota Semarang*. 4(2), 143–157. <https://doi.org/10.55123/mamen.v4i2.5066>
- García, S., Luengo, J., & Herrera, F. (2015). *Data Preprocessing in Data Mining* (Vol. 72). <https://doi.org/10.1007/978-3-319-10247-4>
- Ghous, H., Malik, M., & Rehman, I. (2023). Deep Learning based Market Basket Analysis using Association Rules. *KIET Journal of Computing and Information Sciences*, 6(2), 14–34. <https://doi.org/10.51153/kjcis.v6i2.166>
- Guthrie, C., Fosso-wamba, S., & Brice, J. (2020). Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ’ s public news and information. *Journal of Retailing and Consumer Services*, January.
- Han, J., Kambe, M., & Pe, J. (2011). Data Mining: Concepts and Techniques. In *Data Mining: Concepts and Techniques*. <https://doi.org/10.1016/C2009-0-61819-5>
- Hunyadi, I. D., Constantinescu, N., & Țicleanu, O. A. (2025). Efficient Discovery of Association Rules in E-Commerce: Comparing Candidate Generation and Pattern Growth Techniques. *Applied Sciences* (Switzerland), 15(10). <https://doi.org/10.3390/app15105498>
- Iswavigra, D. U., Zen, L. E., & Hanim, H. (2023). Marketing Strategy UMKM Dengan CRISP-DM Clustering & Promotion Mix Menggunakan Metode K-Medoids. *Jurnal Informasi Dan Teknologi (JIDT)*, 5, 45–54. <https://doi.org/10.37034/jidt.v5i1.260>
- John, J. M., Shobayo, O., & Ogunleye, B. (2023). An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market. *Analytics*, 2(4), 809–823. <https://doi.org/10.3390/analytics2040042>
- Kajornkasirat, S., Gunglin, P., Puangsuwan, K., & Kaewsuwan, N. (2025). *Customer segmentation using association rule mining on retail transaction data*. 14(3), 1919–1929. <https://doi.org/10.11591/ijai.v14.i3.pp1919-1929>
- Kalijaga, M. A., & Putra, B. S. (2021). Penerapan Analytical Hierarchy Process, Clustering,

- dan AR-MBA sebagai Solusi Strategi Pemasaran Pada Toko Swalayan X. *Jurnal Studi Inovasi*, 1(4), 33–39. <https://doi.org/10.52000/jsi.v1i4.63>
- Kanhere, S., Sahni, A., Stynes, P., & Pathak, P. (2021). Clustering based approach to enhance association rule mining. *Conference of Open Innovation Association, FRUCT, 2021-Janua*. <https://doi.org/10.23919/FRUCT50888.2021.9347577>
- Kantardzic, M. (2020). *Data Mining: Concepts, models, methods, and algorithms*. IEEE Press ; Wiley.
- Kementerian Perindustrian Republik Indonesia. (2024). Industri makanan dan minuman diminati investor. *Antara News*. <https://www.antaraneews.com/berita/4585494/kemenperin-industri-makanan-dan-minuman-diminati-investor>
- Ketchen, D. J., & Shook, C. L. (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal*, 17(6), 441–458. [https://doi.org/10.1002/\(SICI\)1097-0266\(199606\)17:6<441::AID-SMJ819>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G)
- Kotler, P., Armstrong, G., & Opresnik, M. O. (2018). Principle Of Marketing. In *Early Greek Mythography, Vol. 1: Texts* (17e ed.). Pearson Education Limited. <https://doi.org/10.1093/oseo/instance.00295839>
- Kotler, P., & Keller, K. L. (2022). *Marketing Management*. Pearson.
- Larose, D. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. In *Sustainability (Switzerland)* (Vol. 11, Issue 1). John Wiley & Sons, Inc. <https://doi.org/10.1002/0471687545>
- Lu, S. Q., & Pauwels, K. (2024). More transactions but a lower average transaction value: How mobile payment apps influence consumer purchases through time-savings. *International Journal of Research in Marketing*, 41(4), 761–776. <https://doi.org/10.1016/j.ijresmar.2024.08.004>
- M. Alawadh, M., & M. Barnawi, A. (2022). A Survey on Methods and Applications of Intelligent Market Basket Analysis Based on Association Rule. *Journal on Big Data*, 4(1), 1–25. <https://doi.org/10.32604/jbd.2022.021744>
- Maddala, V. B., & Sreedevi, M. (2023). CNAIS: Performance Analysis of the Clustering of Non-Associated Items Set Techniques †. *Engineering Proceedings*, 59(1). <https://doi.org/10.3390/engproc2023059014>
- Mandala, E. P. W., & Putri, D. E. (2023). Data mining technique for grouping products using clustering based on association. *Indonesian Journal of Electrical Engineering and Computer Science*, 31(2), 835–844. <https://doi.org/10.11591/ijeecs.v31.i2.pp835-844>
- Martinez, M., Escobar, B., García-Díaz, M. E., & Pinto-Roa, D. P. (2021). Market basket analysis with association rules in the retail sector using Orange. Case study: Appliances sales company. *CLEI Eletronic Journal (CLEIej)*, 24(2), 1–17. <https://doi.org/10.19153/cleiej.24.2.12>
- Maulani, V. R., Barata, M. A., & Yuwita, P. E. (2025). A Improving House Price Clustering Results with K-means through the Implementation of One-hot Encoding Pre-processing Technique. *Journal of Applied Informatics and Computing*, 9(3), 741–748. <https://doi.org/10.30871/jaic.v9i3.9481>
- McCaffrey, J. (2024). *Clustering categorical data using k-means with one-over-n-hot encoding*.
- Ong, A. K. S., Prasetyo, Y. T., Esteller, A. J. D., Bruno, J. E., Lagorza, K. C. O., Oli, L. E. T., Chuenyindee, T., Thana, K., Persada, S. F., & Nadlifatin, R. (2023). Consumer preference analysis on the attributes of samgyeopsal Korean cuisine and its market segmentation: Integrating conjoint analysis and K-means clustering. *PLoS ONE*, 18(2 February), 1–23.

- <https://doi.org/10.1371/journal.pone.0281948>
- Paskasius, E., & Magdalena, E. (2025). *Penerapan Metode Clustering Pada Penjualan Makanan Dan Minuman Dalam Menentukan Strategi Marketing. 1*, 48–56.
- Putri, V. E., & Purnomo, H. D. (2025). *Integrasi algoritma apriori dan k-means dalam analisis pola pembelian untuk meningkatkan strategi pemasaran. 10*(1), 409–423.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C), 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Silviana. (2020). PENGKLASTERAN RISIKO COVID-19 DI RIAU MENGGUNAKAN TEKNIK ONE HOT ENCODING DAN ALGORITMA K-MEANS CLUSTERING. *Jurnal Ekonomi Volume 18, Nomor 1 Maret 201*, 2(1), 41–49.
- Solomon, M. R. (2018). Consumer Behavior Consumer behavior. In *Pearson Education India* (Vol. 27, Issue 23). Pearson. <https://doi.org/https://doi.org/10.1016/B978-0-12-804291-5.00001-5>
- Tan, P.-N., Steinbach, Michael Karpatne, A., & Kumar, V. (2019). *Introduction to Data Mining* (2nd ed). Pearson. <https://doi.org/10.5860/choice.51-3639>
- Teh, M. A., Lim, T. M., & Chin, W. Y. (2023). Unveiling the Wave: Unraveling Sales Trends and Unlocking Market Basket Insights. *International Journal of Membrane Science and Technology*, 10(3), 1917–1931. <https://doi.org/10.15379/ijmst.v10i3.1851>
- Ufeli, C. P., Sattar, M. U., Hasan, R., & Mahmood, S. (2025). Enhancing Customer Segmentation Through Factor Analysis of Mixed Data (FAMD)-Based Approach Using K-Means and Hierarchical Clustering Algorithms. *Information (Switzerland)*, 16(6), 1–25. <https://doi.org/10.3390/info16060441>
- Witten, I. H., Frank, Eibe, H., A., M., & Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (4th ed.). Morgan Kaufmann. <https://doi.org/https://www.sciencedirect.com/book/9780128042915/data-mining>
- Zhang, S., Wu, X., Wang, Y., & Jiang, P. (2024). Does time of day affect consumers' price sensitivity? *Humanities and Social Sciences Communications*, 11(1). <https://doi.org/10.1057/s41599-024-03647-5>
- Zhu, W., Qiu, R., & Fu, Y. (2024). *Comparative Study on the Performance of Categorical Variable Encoders in Classification and Regression Tasks. 1–19.* <http://arxiv.org/abs/2401.09682>

LAMPIRAN

A1-Data Transaksi dari Point of sales pada Grobbak Korea Pandeyan

ID Struk	Status Pembayaran	Platform	Jam	Nama Produk	Kategori	Jumlah Produk	Total Transaksi
4874LLT585 D9Q9DJJ	Success	Grab	17:50:54	Corndog Chocolate, Corndog Mozzarella, Corndog Sosis Mozzarella	Corndog, SweetCo rndog	3	60200
4B7XLLHB QBPJRRRD 4	Success	Grab	19:32:17	Rabokki Extra Mozzarella	Tteobok ki	2	55000
4D7J44HG7 G866BXBB	Success	Grab	18:17:26	Chikuwa Saus Korea, Topokki Extra Chikuwa	Snack, Tteobok ki	2	39400
4L76JJSNQ9 WPK64M	Success	Grab	11:58:39	Chikuwa Saus Korea, Corndog Sosis	Corndog, Snack	2	31500
4M7XLLT46 BZYZRMPP	Success	Grab	20:15:56	Corndog Mozzarella	Corndog	2	38800
4QLLHP474 K4N9	Success	Gofood	20:43:53	Corndog Choco	SweetCo rndog	2	34850
4QLLHP474 Z6D7	Success	Gofood	20:47:20	Corndog Choco, Hot Spicy Bibimmyeo n, Hot Spicy Bibimmyeo n	Bibimmy eon, SweetCo rndog	3	35100

ID Struk	Status Pembayaran	Platform	Jam	Nama Produk	Kategori	Jumlah Produk	Total Transaksi
4QLLHP474 Z97L	Success	Gofood	20:53:23	Corndog Choco, Odeng Fish Soup, Topokki Extra Odeng, Topokki Extra Odeng	Snack, SweetCorndog, Tteobokki	4	48450
4QLLHP474 Z9DY	Success	Shopee	20:54:04	Corndog Choco, Corndog Mozzarella, Corndog Mozzarella, Topokki Original, Topokki Original	Corndog, SweetCorndog, Tteobokki	5	63200
4QLLHP474 ZNMW	Success	Gofood	20:41:56	Corndog Sosis Mozzarella, Corndog Sosis Mozzarella, Odeng Fish Soup, Odeng Fish Soup	Corndog, Snack	6	33100
...	...	...	...	...	...	...	...
ZZL5KKCV W7ZK46W5 R	Success	Grab	15:48:51	Corndog Mozzarella, Topokki Original	Corndog, Tteobokki	2	39400

B1-Code Pyhton untuk Clustering dengan metode K-means

```
# (Opsional) Instal dependensi ketika dibutuhkan
# Jalankan baris di bawah jika paket belum terpasang di environment
# VSCode Anda.
# %pip install -q numpy pandas scikit-learn matplotlib openpyxl

import os
import math
import json
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

from collections import Counter
from datetime import time

from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score, davies_bouldin_score
from sklearn.pipeline import Pipeline

# === KONFIGURASI Jalur Data ===
# Ubah ke nama/letak file Anda saat menjalankan di VSCode.
preferred_name = "Data terpilih.xlsx"
fallback_path = "/mnt/data/Data_terpilih.xlsx" # berguna jika Anda
# membuka notebook ini di ChatGPT
data_path = preferred_name if os.path.exists(preferred_name) else
fallback_path

print("Memakai data:", data_path)
assert os.path.exists(data_path), f"File tidak ditemukan: {data_path}"
```

```
# Baca sheet pertama secara otomatis
raw = pd.read_excel(data_path, sheet_name=0)
print("Dimensi data mentah:", raw.shape)
raw.head(3)

print("\n=== Informasi Kolom & Tipe Data ===")
print(raw.dtypes)

print("\n=== Jumlah Missing Value per Kolom ===")
mv = raw.isna().sum().sort_values(ascending=False)
print(mv[mv>0])

# Tampilkan beberapa baris kandidat cleaning yang umum di data F&B
kolom_kunci = ["ID Struk", "Status Pembayaran", "Platform", "Jam", "Nama
Produk", "Kategori", "Jumlah Produk", "Total"]
ada = [k for k in kolom_kunci if k in raw.columns]
preview = raw[ada].head(10)
preview
```

1. Data Cleaning

Aturan pembersihan yang diterapkan (dapat Anda sesuaikan):

- Hanya transaksi dengan Status Pembayaran = "Success" yang disertakan.
- Baris dengan Nama Produk mengandung teks "Custom Amount" dihapus.
- Baris dengan Jumlah Produk ≤ 0 atau Total ≤ 0 dihapus.
- Nilai Platform yang kosong diisi "Kios" (atau "Unknown" sesuai kebutuhan).

```
df = raw.copy()

def safe_contains(s, keyword):
    try:
        return s.astype(str).str.contains(keyword, case=False,
na=False)
    except Exception:
        return pd.Series([False]*len(s), index=s.index)

# Tandai kandidat baris terhapus (untuk pelaporan)
mask_not_success = df.get("Status Pembayaran",
pd.Series(index=df.index, dtype=object)).ne("Success")
mask_custom_amount = safe_contains(df.get("Nama Produk",
pd.Series(index=df.index, dtype=object)), "Custom Amount")
mask_qty_le0 = df.get("Jumlah Produk", pd.Series(index=df.index,
dtype="float")).fillna(0) <= 0
mask_total_le0 = df.get("Total", pd.Series(index=df.index,
dtype="float")).fillna(0) <= 0
```

```
mask_drop = mask_not_success | mask_custom_amount | mask_qty_le0 |
mask_total_le0
rows_to_drop = df.loc[mask_drop, ["ID Struk", "Status
Pembayaran", "Platform", "Jam", "Nama Produk", "Kategori", "Jumlah
Produk", "Total"]].copy()

print("Total baris akan dihapus:", rows_to_drop.shape[0])
print("Contoh baris yang dihapus (maks 10):")
display(rows_to_drop.head(10))

# Terapkan penghapusan
df = df.loc[~mask_drop].copy()

# Isi Platform kosong
if "Platform" in df.columns:
    df["Platform"] = df["Platform"].fillna("Kios").replace({"Gofood":
"GoFood", "gofood": "GoFood"})
else:
    df["Platform"] = "Kios" # fallback bila kolom tidak ada

# Pastikan tipe numerik
for c in ["Jumlah Produk", "Total"]:
    if c in df.columns:
        df[c] = pd.to_numeric(df[c], errors="coerce")

print("\nDimensi setelah cleaning:", df.shape)
df.head(3)
```

2. Data Transformation

- Parsing jam transaksi lalu dikelompokkan menjadi 4 kategori waktu (*Pagi, Siang, Sore, Malam*).
- Normalisasi penulisan Platform.

```
# Parse jam
def to_hour(x):
    try:
        if pd.isna(x):
            return np.nan
        # jika sudah berupa waktu/hh:mm:ss string
        s = str(x)
        parts = s.split(":")
        if len(parts) >= 2:
            return int(parts[0])
        return np.nan
    except Exception:
        return np.nan

df["hour"] = df.get("Jam", pd.Series(index=df.index)).apply(to_hour)
```

```
# 4 kategori waktu (umum di ritel Indonesia)
def waktu_kategori(h):
    if pd.isna(h):
        return "Tidak Diketahui"
    h = int(h)
    if 6 <= h < 10: # 06:00-09:59
        return "Pagi"
    elif 10 <= h < 14: # 10:00-13:59
        return "Siang"
    elif 14 <= h < 18: # 14:00-17:59
        return "Sore"
    else: # 18:00-05:59
        return "Malam"

df["Waktu_Kategori"] = df["hour"].apply(waktu_kategori)

# Samakan penamaan platform (contoh normalisasi)
df["Platform"] = df["Platform"].astype(str).str.strip()
df["Platform"] = df["Platform"].replace({"gofood": "GoFood", "GOF00D": "GoFood", "Grabfood": "Grab", "grab": "Grab"})
df[["ID Struk", "Jam", "hour", "Waktu Kategori", "Platform"]].head(10)
```

3. Data Selection dan Konstruksi Fitur (per ID Struk)

Variabel utama yang digunakan sesuai kebutuhan klasterisasi:

- Jumlah Produk per struk (`total_items`),
- Subtotal/Total per struk (`subtotal_sum`),
- Waktu (4 kategori) berdasarkan median hour per struk (`waktu_kategori`),
- Platform (*modus*) per struk (`platform_mode`).

Catatan: Bila `Subtotal` tidak tersedia, digunakan kolom `Total`.

```
# Tentukan kolom nilai uang
nilai_kolom = "Subtotal" if "Subtotal" in df.columns else "Total"

# Agregasi per ID Struk
if "ID Struk" not in df.columns:
    raise ValueError("Kolom 'ID Struk' tidak ditemukan di data.
Pastikan nama kolom sesuai.")

grp = (
    df
    .groupby("ID Struk")
    .apply(lambda g: pd.Series({
        "total_items": g["Jumlah Produk"].sum(skipna=True),
        "subtotal_sum": g[nilai_kolom].sum(skipna=True),
```

```
        "hour_median": np.median(g["hour"].dropna()) if
q["hour"].notna().any() else np.nan,
        "waktu_kategori":
Counter(g["Waktu_Kategori"].dropna()).most_common(1)[0][0] if
g["Waktu_Kategori"].notna().any() else "Tidak Diketahui",
        "platform_mode":
Counter(g["Platform"].dropna()).most_common(1)[0][0] if
g["Platform"].notna().any() else "Kios",
        "unique_categories":
q["Kategori"].astype(str).str.split(",").apply(lambda s: [x.strip()
for x in s if x]).explode().nunique() if "Kategori" in g.columns else
np.nan
    )))
    .reset_index()
)

print("Dimensi matriks struk:", grp.shape)
grp.head(5)
```

4. Data Normalization & Encoding

- Fitur numerik dinormalisasi ke *z-score* menggunakan `StandardScaler`.
- Fitur kategorikal (`waktu_kategori` dan `platform_node`) di-*encode* menggunakan *one-hot encoding* (tanpa normalisasi karena nilainya biner).

```
features_num = ["total_items", "subtotal_sum"]
features_cat = ["waktu_kategori", "platform_node"]

ct = ColumnTransformer(
    transformers=[
        ("num", StandardScaler(), features_num),
        ("cat", OneHotEncoder(handle_unknown="ignore",
sparse_output=False), features_cat),
    ],
    remainder="drop"
)

X = ct.fit_transform(grp[features_num + features_cat])

# Nama kolom hasil transformasi
ohe = ct.named_transformers_["cat"]
cat_names = list(ohe.get_feature_names_out(features_cat))
X_cols = [f"z_{c}" for c in features_num] + cat_names

print("Shape matriks fitur:", X.shape)
print("Fitur:", X_cols)
```

5. Penentuan Jumlah Kluster (*Elbow, Silhouette, Davies–Bouldin*)

Kita evaluasi untuk $k = 2..10$ lalu bandingkan:

- **SSE/Inertia** (*Elbow*),
- **Silhouette Score** (semakin besar semakin baik),
- **Davies–Bouldin Index** (semakin kecil semakin baik).

```
ks = list(range(2, 11))
sse_list, sil_list, dbi_list = [], [], []

for k in ks:
    km = KMeans(n_clusters=k, n_init=30, random_state=42)
    labels = km.fit_predict(X)
    sse_list.append(km.inertia_)
    sil_list.append(silhouette_score(X, labels) if len(set(labels))>1
    else np.nan)
    dbi_list.append(davies_bouldin_score(X, labels) if
    len(set(labels))>1 else np.nan)

metrics = pd.DataFrame({"k": ks, "SSE": sse_list, "Silhouette":
sil_list, "DBI": dbi_list})
display(metrics)

# Visualisasi (satu plot per metrik)
plt.figure()
plt.plot(ks, sse_list, marker="o")
plt.title("Elbow (SSE vs k)")
plt.xlabel("k"); plt.ylabel("SSE")
plt.show()

plt.figure()
plt.plot(ks, sil_list, marker="o")
plt.title("Silhouette vs k (lebih besar lebih baik)")
plt.xlabel("k"); plt.ylabel("Silhouette")
plt.show()

plt.figure()
plt.plot(ks, dbi_list, marker="o")
plt.title("Davies–Bouldin vs k (lebih kecil lebih baik)")
plt.xlabel("k"); plt.ylabel("DBI")
plt.show()

# Heuristik saran k
def pick_k(metrics: pd.DataFrame):
    # ambil k terbaik versi silhouette dan dbi, lalu pilih modulusnya;
    # jika beda, prioritaskan silhouette
    k_sil = metrics.loc[metrics["Silhouette"].idxmax(), "k"] if
```

```
metrics["Silhouette"].notna().any() else None
    k_dbi = metrics.loc[metrics["DBI"].idxmin(), "k"] if
metrics["DBI"].notna().any() else None
    if k_sil is not None and k_dbi is not None and k_sil == k_dbi:
        return int(k_sil),
    ("by": "agree", "k_sil": int(k_sil), "k_dbi": int(k_dbi))
    elif k_sil is not None:
        return int(k_sil),
    ("by": "silhouette", "k_sil": int(k_sil), "k_dbi": int(k_dbi) if k_dbi is
not None else None)
    elif k_dbi is not None:
        return int(k_dbi),
    ("by": "dbi", "k_sil": None, "k_dbi": int(k_dbi))
    else:
        return 3, ("by": "default", "k_sil": None, "k_dbi": None)

k_opt, why = pick_k(metrics)
print("Saran k:", k_opt, "| alasan:", why)
```

6. Proses *Clustering* (K-Means) & Ringkasan Hasil

- Menjalankan K-Means dengan k terpilih.
- Menyajikan pusat klaster dalam *z-score* (sesuai kebutuhan pelaporan).
- Tambahan: pusat klaster pada skala asli untuk interpretasi bisnis.

```
k_final = k_opt
km_final = KMeans(n_clusters=k_final, n_init=50, random_state=42)
labels = km_final.fit_predict(X)

# Pusat klaster pada ruang fitur terstandardisasi (z-score utk fitur numerik)
centroids = pd.DataFrame(km_final.cluster_centers_, columns=X_cols)
centroids.index = [f"Cluster {i}" for i in range(k_final)]

print("=== Pusat Klaster (dalam ruang fitur; z-score utk fitur numerik) ===")
display(centroids)

# Tambahan: pusat klaster dalam skala asli untuk fitur numerik
# Ambil hanya bagian numerik dari centroid dan inverse transform
scaler = ct.named_transformers_["num"]
num_centroids = scaler.inverse_transform(centroids[[c for c in centroids.columns if c.startswith("z ")]])
num_centroids_df = pd.DataFrame(num_centroids, columns=features_num, index=centroids.index)

print("\n=== Pusat Klaster (skala asli; numerik saja) ===")
display(num_centroids_df.round(3))
```

```
# Lampirkan label ke data struk
grp_clu = grp.copy()
grp_clu["cluster"] = labels

print("\nDistribusi anggota per klaster:")
display(grp_clu["cluster"].value_counts().sort_index())

grp_clu.head(10)
```

7. Deskripsi Statistik Tiap Klaster & Rentang Variabel

Menyajikan statistik ringkas (count, mean, std, min, 25%, 50%, 75%, max) per klaster untuk fitur numerik, serta tabulasi komposisi Platform dan Waktu.

```
desc_by_cluster = grp_clu.groupby("cluster")
[features_num].describe().round(3)
display(desc_by_cluster)

print("\nKomposisi Platform per Klaster:")
display(pd.crosstab(grp_clu["cluster"], grp["platform_mode"],
normalize="index").round(3))

print("\nKomposisi Waktu (4 kategori) per Klaster:")
display(pd.crosstab(grp_clu["cluster"], grp["waktu_kategori"],
normalize="index").round(3))

# Rentang persebaran numerik per klaster
rng = grp_clu.groupby("cluster")
[features_num].agg(["min", "max"]).round(3)
print("\nRentang (min-max) per klaster:")
display(rng)
```

8. Export Hasil & Data Selection untuk AR-MBA

- **kmeans_assignments.csv**: anggota klaster per struk dengan fitur ringkas.
- **mba_transactions_by_product.csv**: 1 baris = 1 struk; kolom **items** berisi daftar nama produk (dipisah koma).
- **mba_transactions_by_category.csv**: 1 baris = 1 struk; kolom **items** berisi daftar kategori (dipisah koma; sudah *split* dari teks kategori bertanda koma).

```
out_assign = "kmeans_assignments.csv"
grp_clu.to_csv(out_assign, index=False, encoding="utf-8-sig")
print("Tersimpan:", out_assign, "| shape:", grp_clu.shape)

# Data selection utk AR-MBA (by product & category)
cols_need = ["ID Struk", "Nama Produk", "Kategori"]
missing_cols = [c for c in cols_need if c not in df.columns]
```

```
if missing_cols:
    print("Peringatan: kolom berikut tidak ditemukan dan *export* MBA
    mungkin tidak lengkap:", missing_cols)

# Produk: gabungkan nama produk per struk (gunakan baris cleaned)
if all(c in df.columns for c in ["ID Struk", "Nama Produk"]):
    mba_prod = (df.groupby("ID Struk")["Nama Produk"]
                .apply(lambda s: ", ".join(sorted(set(str(x).strip()
for x in s if pd.notna(x) and str(x).strip()))))
                .reset_index()
                .rename(columns={"Nama Produk": "items"}))
    mba_prod.to_csv("mba_transactions_by_product.csv", index=False,
encoding="utf-8-sig")
    print("Tersimpan: mba_transactions_by_product.csv | shape:",
mba_prod.shape)

# Kategori: pecah kategori bertanda koma dan gabungkan unik per struk
if all(c in df.columns for c in ["ID Struk", "Kategori"]):
    def split_cats(val):
        if pd.isna(val):
            return []
        return [x.strip() for x in str(val).split(",") if x and
str(x).strip()]
    mba_cat = (df.assign(cats=df["Kategori"].apply(split_cats))
               .explode("cats")
               .dropna(subset=["cats"])
               .groupby("ID Struk")["cats"]
               .apply(lambda s: ", ".join(sorted(set(s))))
               .reset_index()
               .rename(columns={"cats": "items"}))
    mba_cat.to_csv("mba_transactions_by_category.csv", index=False,
encoding="utf-8-sig")
    print("Tersimpan: mba_transactions_by_category.csv | shape:",
mba_cat.shape)
```

B2-Code Python untuk Association rules minning dengan fp-growth

```
# == Persiapan
import os, re, math, itertools
from pathlib import Path
import pandas as pd
import numpy as np

# Path file input (ubah jika perlu)
BASE = Path('.')
INPUT_PATH = BASE / "Cluster_Siap-AR-MBA.xlsx" # ganti ke path file
Anda bila beda
OUT_DIR = BASE / "output ar mba"
OUT_DIR.mkdir(parents=True, exist_ok=True)

def split_items(s):
    if pd.isna(s):
        return []
    s = str(s)
    s = re.sub(r'\\[\\]\\[\\]\\[\\]', '', s)
    parts = [p.strip() for p in s.split(',')]
    parts = [p for p in parts if p]
    return parts

df_raw = pd.read_excel(INPUT_PATH)
df_raw["items"] = df_raw["Nama Produk"].apply(split_items)
df_raw["n_items"] = df_raw["items"].apply(len)
display(df_raw.head())
```

1) *Data Cleaning*: buang transaksi dengan 1 produk

```
df_single = df_raw.loc[df_raw["n_items"] <= 1, ["ID Struk", "Nama
Produk", "n_items", "cluster"]].copy()
df_clean = df_raw.loc[df_raw["n_items"] > 1].copy()

df_clean_c = df_clean.dropna(subset=["cluster"]).copy()
df_clean_c["cluster"] = df_clean_c["cluster"].astype(int)
df_clean_c["cluster_label"] = df_clean_c["cluster"] + 1

summary_clean = pd.DataFrame({
    "Total (sebelum cleaning)": [len(df_raw)],
    "Hanya 1 produk (dibuang)": [len(df_single)],
```

```
    "Tersisa (sesudah cleaning)": [len(df_clean)],
    "Baris siap AR-MBA (punya cluster)": [len(df_clean_c)]
})
display(summary_clean)
display(df_single.head(15))
```

2) Data Selection per kluster (ID Transaksi & Nama Barang)

```
selection_table = df_clean_c[["cluster_label", "ID Struk", "Nama
Produk"]].rename(
    columns={"cluster_label": "Kluster", "ID Struk": "ID Transaksi", "Nama
Produk": "Nama Barang"}
).sort_values(["Kluster", "ID Transaksi"]).reset_index(drop=True)
display(selection_table.head(20))
```

3) Transformasi ke Boolean (True/False)

```
def build_boolean_matrix(transactions):
    all_items = sorted({it for b in transactions for it in b})
    mat = pd.DataFrame(False, index=range(len(transactions)),
        columns=all_items, dtype=bool)
    for i, b in enumerate(transactions):
        for it in set(b):
            mat.at[i, it] = True
    return mat

boolean_mats = {}
for c in sorted(df_clean_c["cluster_label"].unique()):
    trans = df_clean_c.loc[df_clean_c["cluster_label"]==c,
        "items"].tolist()
    boolean_mats[c] = build_boolean_matrix(trans)
    display(boolean_mats[c].head(10))
```

4) *FP-Growth* (tanpa paket eksternal)

```
# Implementasi minimal FP-Growth
from collections import Counter

class FPTreeNode:
    slots = ("item", "count", "parent", "children", "node_link")
    def __init__(self, item, count, parent):
        self.item = item
        self.count = count
        self.parent = parent
        self.children = {}
        self.node_link = None
```

```
class FPTree:
    def __init__(self):
        self.root = FPTreeNode(None, 0, None)
        self.header_table = {} # item -> [support, first_node]

    def build_fp_tree(transactions, min_support_abs):
        freq = Counter()
        for trans in transactions:
            freq.update(set(trans))
        freq = {item:cnt for item,cnt in freq.items() if cnt >=
min support abs}
        if not freq:
            return None, {}
        order = sorted(freq.items(), key=lambda x: (-x[1], x[0]))
        ordering = {item:i for i,(item,_) in enumerate(order)}
        tree = FPTree()
        tree.header_table = {item:[cnt, None] for item,cnt in
freq.items()}
        for trans in transactions:
            items = [i for i in set(trans) if i in freq]
            if not items:
                continue
            items.sort(key=lambda i: ordering[i])
            cur = tree.root
            for item in items:
                if item in cur.children:
                    node = cur.children[item]; node.count += 1
                else:
                    node = FPTreeNode(item, 1, cur); cur.children[item] =
node
            head = tree.header_table[item][1]
            if head is None: tree.header_table[item][1] = node
            else:
                last = head
                while last.node_link is not None: last =
last.node_link
                last.node_link = node
            cur = node
        return tree, freq

    def _mine_tree(tree, min_support_abs, prefix=(), out=None):
        if out is None: out = {}
        items = sorted(tree.header_table.items(), key=lambda x: x[1][0])
        for item, (support, first_node) in items:
            new_itemset = tuple(sorted(prefix + (item,)))
            out[new_itemset] = support
            cond_trans = []
            n = first_node
            while n is not None:
```

```

        path = []
        parent = n.parent
        while parent and parent.item is not None:
            path.append(parent.item)
            parent = parent.parent
        if path:
            cond_trans.extend([path]*n.count)
        n = n.node_link
    if cond_trans:
        cond_tree, cond_freq = _build_fp_tree(cond_trans,
min support abs)
        if cond_tree is not None and cond_tree.header_table:
            mine_tree(cond_tree, min_support_abs, prefix=prefix+
(item,), out=out)
    return out

def fpgrowth(transactions, min_support_ratio=0.82):
    n = len(transactions)
    min_support_abs = max(1, math.ceil(min_support_ratio * n))
    tree, freq = _build_fp_tree(transactions, min_support_abs)
    if tree is None: return {}
    out = mine_tree(tree, min_support_abs, prefix={}, out={})
    return {tuple(sorted(k)): v for k,v in out.items()}

def gen_rules(freq_itemsets, n_trans, min_conf=0.3, min_lift=1.0):
    support = {tuple(sorted(k)): v/n_trans for k,v in
freq_itemsets.items()}
    rules = []
    for itemset, sup in support.items():
        if len(itemset)<2: continue
        items = list(itemset)
        for r in range(1, len(items)):
            for antecedent in itertools.combinations(items, r):
                antecedent = tuple(sorted(antecedent))
                consequent = tuple(sorted(set(items)-set(antecedent)))
                sup_item = support.get(itemset, 0.0)
                sup_a = support.get(antecedent, 0.0)
                sup_c = support.get(consequent, 0.0)
                if sup_a==0 or sup_c==0: continue
                conf = sup_item/sup_a
                lift = conf/sup_c if sup_c>0 else np.nan
                leverage = sup_item - sup_a*sup_c
                conviction = (1-sup_c)/(1-conf) if (1-conf)>0 else
np.inf

                if conf>=0.3 and lift>=1.0:
                    rules.append({
                        "antecedent": antecedent,
                        "consequent": consequent,
                        "support": sup_item,

```

```

                        "confidence": conf,
                        "lift": lift,
                        "leverage": leverage,
                        "conviction": conviction
                    })
    rules.sort(key=lambda d: (-d["lift"], -d["confidence"], -
d["support"]))
    return pd.DataFrame(rules)

```

5) Perbandingan Nilai *Minimum Support* dan Aturan per Klaster

```
min_support_grid = [0.01, 0.02, 0.03, 0.05]
summary_rows = []
rules_by_cluster = {}
itemsets_by_cluster = {}

for c in sorted(df_clean_c["cluster_label"].unique()):
    trans = df_clean_c.loc[df_clean_c["cluster_label"]==c,
"items"].tolist()
    rules_by_cluster[c] = {}
    itemsets_by_cluster[c] = {}
    for ms in min_support_grid:
        fi = fpgrowth(trans, min_support_ratio=ms)
        rules = gen_rules(fi, n_trans=len(trans), min_conf=0.3,
min_lift=1.0)
        summary_rows.append({"Cluster": int(c),
                             "MinSupport": ms,
                             "Jumlah Itemset": len(fi),
                             "Jumlah Aturan": len(rules)})
        itemsets_by_cluster[c][ms] = pd.DataFrame(
            [{"itemset": " ".join(k), "panjang": len(k),
"support count": v, "support": v/len(trans)} for k,v in fi.items()])
        itemsets_by_cluster[c][ms].sort_values(["panjang","support count","itemset"],
ascending=[True,False,True]).reset_index(drop=True)
        rules_by_cluster[c][ms] = rules

summary_df = pd.DataFrame(summary_rows)
display(summary_df)
```

6) Simpan ke Excel

```
excel_path = OUT_DIR / "AR MBA FP Growth Output.xlsx"
with pd.ExcelWriter(excel_path, engine="xlsxwriter") as writer:
    df_raw.to_excel(writer, sheet_name="raw_input", index=False)
    df_raw[["ID Struk", "Nama
```

```
Produk", "n items", "cluster"]].to_excel(writer,
sheet_name="before_clean", index=False)
    df_single.to_excel(writer, sheet_name="single_item_tx",
index=False)
    df_clean.to_excel(writer, sheet_name="after_clean", index=False)
    summary_clean.to_excel(writer, sheet_name="cleaning_summary",
index=False)
    selection_table.to_excel(writer,
sheet_name="selection_by_cluster", index=False)
    for c in sorted(df_clean_c["cluster_label"].unique()):
        mat =
build_boolean_matrix(df_clean_c.loc[df_clean_c["cluster_label"]==c,
"items"].tolist())
        # Write a compact preview to keep file size manageable
        mat.iloc[:500, :min(100, mat.shape[1])].to_excel(writer,
sheet_name=f"bool C({c})", index=False)
        summary_df.to_excel(writer, sheet_name="support_compare",
index=False)
        for c in sorted(itemsets_by_cluster.keys()):
            for ms in min_support_grid:
                itemsets_by_cluster[c][ms].to_excel(writer,
sheet_name=f"itemsets_C({c})_sup({int(ms*100)}", index=False)
                rules_by_cluster[c][ms].to_excel(writer,
sheet_name=f"rules_C({c})_sup({int(ms*100)}", index=False)
print("Tersimpan ->", excel_path.resolve())
```